



# CERC 2020

Collaborative European Research Conference

Belfast, UK

10 - 11 September 2020

<https://www.cerc-conf.eu>

# Proceedings

Editors (in alphabetical order) :

Haithem Afli

Udo Bleimann

Dirk Burkhard

Robert Loew

Stefanie Regier

Ingo Stengel

Haiying Wang

Huiru (Jane) Zheng

Version of 17 April 2021 @ 17:25

# Preface

In today's world, which has recently seen fractures and isolation forming among states, international and interdisciplinary collaboration is an increasingly important source of progress. Collaboration is a rich source of innovation and growth. It is the goal of the Collaborative European Research Conference (CERC2020) to foster collaboration among friends and colleagues across disciplines and nations within Europe. CERC emerged from long-standing cooperation between the Cork Institute of Technology, Ireland and Hochschule Darmstadt - University of Applied Sciences, Germany. CERC has grown to include more well-established partners in Germany, the United Kingdom, Greece, Spain, Italy, and many more.

CERC is truly interdisciplinary, bringing together new and experienced researchers from science, engineering, business, humanities, and the arts. At CERC researchers not only present their findings as published in their research papers. They are also challenged to collaboratively work out joint aspects of their research during conference sessions and informal social events and gatherings.

Organizing such an event involves the hard work of many people. COVID-19 pandemic has impacted our daily life and research. It has been a significant change to CERC2020 and this is the first time the conference was held virtually online. The conference has received submissions from worldwide, not just European countries. Thanks go to the international program committee and my fellow program chairs, particularly to Prof Udo Bleimann for invaluable support throughout the conference. Prof Ingo Stengel, Dr. Haiying Wang, Dr. Ali Haithem, and Dr. Stefanie Regier for supporting me in the review process. Dirk Burkhardt and Dr. Robert Loew put a great effort into setting up the website and conference management system and preparing the conference programme and proceedings. Thank my colleagues from Ulster University, Hochschule Karlsruhe and Hochschule Darmstadt, and the Cork Institute of Technology, Ireland for providing invaluable support to the conference. CERC2020 has received supports from Ulster University, VISit Belfast, and Belfast City Council.

Professor Huiru (Jane) Zheng

Conference and Program Chair, CERC 2020  
Belfast, UK, September 2020

## Conference Chairs

Prof. Huiru (Jane) Zheng, Ulster University, UK  
Dr. Haithem Afli, Cork Institute of Technology, Ireland  
Prof. Dr Udo Bleimann, Darmstadt University of Applied Sciences, Germany  
Prof. Dr. Bernhard Humm, Darmstadt University of Applied Sciences, Germany  
Dr. Robert Loew, Darmstadt University of Applied Sciences, Germany & DIPF | Leibniz Institute for Research and Information in Education, Germany  
Prof. Dr. Stefanie Regier, University of Applied Sciences Karlsruhe, Germany  
Prof. Dr. Ingo Stengel, University of Applied Sciences Karlsruhe, Germany  
Dr. Haiying Wang, Ulster University, UK

## International Scientific Committee

Jens-Peter Akelbein, Darmstadt University of Applied Sciences, Germany  
Tatjana Archipov, University of Applied Sciences Karlsruhe, Germany  
Lu Bai, Ulster University, United Kingdom  
Raymond Bond, Ulster University, United Kingdom  
Fiona Brown, Datatics Ltd, United Kingdom  
Dirk Burkhardt, Darmstadt University of Applied Sciences, Germany  
Zoraida Callejas Carrión, Granada University, Spain  
En-Chi Chang, AMG Digital Marketing & Design Company, Taiwan  
Zhibin Chen, Durham University, United Kingdom  
Simone Dogu, DB Systel GmbH, Germany  
Markus Döhring, Darmstadt University of Applied Sciences, Germany  
Lei Duan, Sichuan University, China  
Klaus-Peter Fischer-Hellmann, Digamma GmbH, Germany  
Marius Faßbender, Daimler AG, Germany  
Matthias Feiner, University of Applied Sciences Karlsruhe, Germany  
Orla Flynn, Galway-Mayo Institute of Technology, Ireland  
Steven Furnell, Plymouth University, United Kingdom  
Jianliang Gao, Imperial College, United Kingdom  
Gunter Grieser, Darmstadt University of Applied Sciences, Germany  
Vic Grout, Glyndwr University, United Kingdom  
Markus Haid, Darmstadt University of Applied Sciences, Germany  
Joe Harrington, Cork Institute of Technology, Ireland  
Andreas Heberle, University of Applied Sciences Karlsruhe, Germany  
Andreas Heinemann, Darmstadt University of Applied Sciences, Germany  
Peter Henning, University of Applied Sciences Karlsruhe, Germany  
Melina Hillenbrand, University of Applied Sciences Karlsruhe, Germany  
Reimar Hofmann, University of Applied Sciences Karlsruhe, Germany  
Martin Knahl, University of Applied Sciences Furtwangen, Germany  
Thorsten Leize, University of Applied Sciences Karlsruhe, Germany  
Habin Lee, Brunel University, United Kingdom  
Olive Lennon, University College Dublin, Ireland  
Margot Mieskes, Darmstadt University of Applied Sciences, Germany  
Kawa Nazemi, Darmstadt University of Applied Sciences, Germany  
Rainer Neumann, University of Applied Sciences Karlsruhe, Germany  
Thomas Pleil, Darmstadt University of Applied Sciences, Germany  
Nacim Ramdani, Université of Orléans, France  
Erwin Rauch, Free University of Bozen-Bolzano, Italy  
Sven Rogalski, Darmstadt University of Applied Sciences, Germany  
Pablo Mesejo Santiago, Granada University, Spain  
Christian Schalles, DHBW Mosbach, Germany  
Jan Schikofsky, LaunchMarkets Institute, Germany  
Bryan Scotney, Ulster University, United Kingdom

Helen O' Shea, Cork Institute of Technology, Ireland  
Melanie Siegel, Darmstadt University of Applied Sciences, Germany  
Roy Sleator, Cork Institute of Technology, Ireland  
Lina Stankovic, Strathclyde University, United Kingdom  
Vladimir Stankovic, University of Strathclyde, United Kingdom  
Uta Störl, Darmstadt University of Applied Sciences, Germany  
Jan Stöß, University of Applied Sciences Karlsruhe, Germany  
Bernhard Thull, Darmstadt University of Applied Sciences, Germany  
Ulrich Trick, University of Applied Sciences Frankfurt, Germany  
Hui Wang, Ulster University, United Kingdom  
Christoph Wentzel, Darmstadt University of Applied Sciences, Germany  
Andrea Wirth, University of Applied Sciences Karlsruhe, Germany  
Matthias Wölfel, University of Applied Sciences Karlsruhe, Germany  
Angela Wright, Cork Institute of Technology, Ireland  
Tuba Yilmaz Abdolsaheb, Istanbul Technical University, Turkey



# Table of Content

## Chapter 1

### **Internet of Things, Networks and Robotics**

---

<b>Spatial Mapping for Visually Impaired and Blind using BLE Beacons</b>	15
Alan McGibney, Roman Pospisil, Kevin O'Mahony, Juan Francisco Martinez, Susan Rea	
<b>SenseBot: A Wearable Sensor Enabled Robotic System to Support Health and Well-being</b>	30
Luigi D'Arco, Huiru Zheng, Haiying Wang	
<b>SDR Demonstration System for the Investigation of Cooperative Communication and the Scaling Behaviour of MANETs</b>	46
Mus'ab Yueksel, Milenko Starcik, Florian Hofmann, Raphael Rolny, Marc Kuhn, Michael Kuhn	
<b>A new metric for assessing the performance of 2D Lidar SLAMs</b>	64
Bingxin Zi, Haiying Wang, Jose Santos, Huiru Zheng	
<b>Possibilities for software development for energy-limited constrained devices</b>	78
Mario Hoss, Jens-Peter Akelbein	

## Chapter 2

### **Data Computing and Artificial Intelligence**

---

<b>Development and Validation of an Algorithm for Emissivity-Corrected Pyrometry Independent of Material Properties</b>	89
Denise Reichel	
<b>Detecting Sarcasm in News Headlines</b>	100
Onyinye Chudi-Iwueze, Haithem Afli	
<b>Chatbot based Behaviour Analysis for Obesity Support Platform</b>	112
Yanxin Wu, Ryan Donovan, Binh Vu, Felix Engel, Matthias Hemmje, Haithem Afli	

<b>Online News Analysis on Cloud Computing Platform for Market Prediction</b>	125
Claudia Juarez, Haithem Afli	
<b>Common sense validation and reasoning using Natural Language Processing</b>	141
Shrenik Doshi, Praveen Joshi, Haithem Afli	
<b>ZKPs: Does This Make The Cut?</b>	152
Stavros Kassaras, Leandros Maglaras	
<b>Application of statistical functions for microstructure characterization and determination of elastic properties of ceramic foam</b>	164
Vinit Vijay Deshpande, Romana Piat, Kay Andre Weidenmann	
<b>Automatic Error Detection For Image To Text Tools</b>	183
Suzanne Considine, Haithem Afli, Andy Way	

### Chapter 3

## **Visual Computing and Data Analytics**

---

<b>Determining Elastic Properties of Particle Reinforced Polymer Composites by Numerical Modeling of their Microstructures</b>	193
Pascal Alexander Happ, Timo Bensing, Simon Terwolbeck, Katharina Malek, Martin Moneke, Romana Piat	
<b>Integration of Pattern-Based Learning Management Systems with Immersive Learning Content</b>	211
Michael Winterhagen, Dominic Heutelbeck, Benjamin Wallenborn, Matthias L. Hemmje	
<b>Visual Dashboards in Trend Analytics to Observe Competitors and Leading Domain Experts</b>	222
Lennart Sina, Dirk Burkhardt, Kawa Nazemi	
<b>Emerging Knowledge Extraction and Visualization in Medical Document Corpora</b>	236
Christian Nawroth, Marc Herrmann, Felix Engel, Paul Mc Kevitt, Matthias Hemmje	

<b>A Holistic UX Methodological Framework for Measuring the Dynamic, Adaptive and Intelligent Aspects of a Software Solution</b>	255
Vivien Johnston, Michaela Black, Jonathan Wallace	

<b>A Conceptual Architecture for AI-based Big Data Analysis and Visualization Supporting Metagenomics Research</b>	264
Thoralf Reis, Thomas Krause, Marco X. Bornschlegl, Matthias L. Hemmje	

## Chapter 4

# **COVID-19 Research and Smart Healthcare**

---

<b>Understanding the Impact of COVID-19 on Behavior Changes and Decision Making of Chinese Students and Researchers in the UK</b>	275
Hui Yang, Huiru Zheng, Xi Liu	

<b>A Close Look on the Corona Impact on Surveillance Radar Channel Loads</b>	293
Roman Raekow, Michael Kuhn, Bernd-Ludwig Wenning	

<b>Wearable Gait Analysis — stepping towards the mainstream</b>	306
Philip Morrow, Huiru Zheng, Graham McCalmont, Haiying Wang, Sally McClean	

<b>TeleStroke System (TSS)</b>	325
Rishabh Chandaliya, Praveen Joshi, Haithem Afli	

<b>Exploratory Gait Analysis using Wearable Technology</b>	335
Richard Davies, Paul McCullagh, Huiru Zheng, Joan Condell	

## Chapter 5

## Business and Society

---

<b>Requirements for payment systems in e-commerce from the retailers' perspective</b> Lisa Zuyeva, Ingo Stengel, Stefanie Regier	349
<b>An Innovative IoT Based Financing Model for SMEs</b> Marcel Kehler, Stefanie Regier, Ingo Stengel	361
<b>Operational Management of Data Centers Energy Efficiency by dynamic optimization -Based on a Vector Autoregressive Model- Reinforcement Learning(VAR-RL) approach</b> Yang Hai, Truong Van Nguyen, Youngseok Choi, Mohammed Bahja, Habin Lee	371
<b>Escaping the Everyday Chaos: Assessing the Needs for Internal Knowledge Transfer in SMEs via an Escape Room</b> Irina Ehrlich, Michael Filipenko, Tanja Kranawetleitner, Heike Krebs, Robert Loew, Diana Pistoll, Julia Thurner-Irmler	382
<b>Determinants for the adoption of Regulatory Technology (RegTech) services by the companies in United Arab Emirates: An MCDM Approach</b> Mohamed Muzammil, Nitin Simha Vihari	397
<b>A Future Prospect for European Collaboration on Advanced Analytics in Economy and Society</b> Nicola Below, Dirk Burkhardt, Lukas Kaupp, Kawa Nazemi	414
<b>Millennial's CSR Perception, Social Influences, And Intention to Buy Social Responsible Products: A Conceptual Framework</b> En-Chi Chang, Stefanie Regier, Ingo Stengel	427

# List of Authors

Afli	100, 112, 125, 141, 183, 325	Akelbein	78
Bahja	371	Below	414
Bensing	193	Black	255
Bornschlegl	264	Burkhardt	222, 414
Chandaliya	325	Chang	427
Choi	371	Chudi-Iwueze	100
Condell	335	Considine	183
D'Arco	30	Davies	335
Deshpande	164	Donovan	112
Doshi	141	Ehrlich	382
Engel	112, 236	Filipenko	382
Hai	371	Happ	193
Hemmje	112, 211, 236, 264	Herrmann	236
Heutelbeck	211	Hofmann	46
Hoss	78	Johnston	255
Joshi	141, 325	Juarez	125
Kassaras	152	Kaupp	414
Kehler	361	Kranawetleitner	382
Krause	264	Krebs	382
Kuhn	46, 46, 293	Lee	371
Liu	275	Loew	382
Maglaras	152	Malek	193
Martinez	15	Mc#Kevitt	236
McCalmont	306	McClean	306
McCullagh	335	McGibney	15
Moneke	193	Morrow	306
Muzammil	397	Nawroth	236
Nazemi	222, 414	Nguyen	371
O'Mahony	15	Piat	164, 193
Pistoll	382	Pospisil	15
Raekow	293	Rea	15
Regier	349, 361, 427	Reichel	89
Reis	264	Rolny	46
Santos	64	Sina	222
Starcik	46	Stengel	349, 361, 427
Terwolbeck	193	Thurner-Irmler	382
Vihari	397	Vu	112
Wallace	255	Wallenborn	211
Wang	30, 64, 306	Way	183



Weidenmann	164	Wenning	293
Winterhagen	211	Wu	112
Yang	275	Yueksel	46
Zheng	30, 64, 275, 306, 335	Zi	64
Zuyeva	349		

## Chapter 1

---

# Internet of Things, Networks and Robotics



# Spatial Mapping for Visually Impaired and Blind using BLE Beacons

Alan McGibney<sup>[0000-0002-0665-2005]</sup>, Roman Pospisil, Kevin O'Mahony, Juan Francisco Martinez and Susan Rea<sup>[0000-0002-4388-661X]</sup>

Nimbus Centre, Cork Institute of Technology, Bishopstown, Cork, Ireland  
alan.mcGibney@cit.ie

**Abstract.** This paper describes the development of a set of software services called the Context Awareness Module to support the visually impaired and blind (ViB) to construct a spatial map of their environment through the provision of context information (contextual, directional and positional cues) relating to the surrounding environment. This information is captured through the interaction of the users' smart phone and the deployment of low-cost Bluetooth beacons within the environment to identify objects, landmarks or markers. The solution aims to supplement existing methods that support mobility and navigation through complex spaces by providing an additional layer of information that describes the space, location, object or any entity that a user might come in the vicinity of or interact with. Initial validation of the proposed solution was undertaken with members of the visually impaired community and tested with an example scenario where a visually impaired person is attending a meeting at an unknown building.

**Keywords:** Bluetooth, Location Services, Mapping, Software.

## 1 Introduction

Based on a detailed analysis of existing trends, global projections estimate a continued increase in people with moderate and severe vision impairment from 237.1 million people in 2020 to as high as 587.6 million people by 2050 [1]. According to the definition of visual impairment of the World Health Organisation, currently 1.6 million people suffer from blindness in the EU and only 5% are fully autonomous in their daily mobility. 40% of the visually impaired suffer head level accidents at least once a month, and 30% suffer a fall accident at least once a month. As our cities evolve and population continue to expand mobility is becoming increasingly challenging task facing all citizens, however it is even more significant if a person has a visual impairment or disability. While on one side the Irish Disability Act 2005 states that Government departments and public bodies must work to improve the quality of life for people with disabilities, on the other side public spaces are being designed based on the concept of shared spaces where there is no kerb or level difference to segregate pedestrians and vehicles. This design approach has resulted in unexpected challenges that are adversely affecting vulnerable citizens. Removal of the clear demarcation between paths and roadways

makes mobility significantly more challenging as drivers, cyclists and pedestrians now all occupy the same-shared space with pedestrians relying on the principle of mutual eye contact to navigate safely. For people with sensorial or cognitive disabilities this is not appropriate, and it further marginalises already vulnerable citizens. Similarly, in indoor environments architectural and visually appealing design can often result in the challenges faced by ViB being overlooked, as a result this can limit the level of independence, increase stress and add additional risk for the ViB person when moving in unfamiliar spaces. Technology can play a role in improving how the ViB community can experience the environment around them while also ensuring safety as they navigate through a space.

Several systems termed as Electronic Travel Aid (ETA) have been created to improve the autonomous mobility for ViB people however the adoption rate remains very low. Devices such as wearable solutions (sunglasses, gloves etc) are sometimes considered as extra prosthesis, cumbersome and stigmatising. Inaccuracy of sensor systems that rely on a single sensor technology can diminish the confidence of the user in the benefits of the solutions, for example ultrasound is sensitive to multi-echo and can easily lead to wrong detections. Perception is often limited to range sensing (of the nearest target) and as a result most systems scan the environment without interpreting it, this provides some additional support to the user however it does not provide sufficient detail to allow the visually impaired person to construct a representation and understanding of their specific situation and environment. While existing ETA help a user to navigate and detect obstacles there is a need to provide mechanism that can enhance interaction with the surrounding environment for ViB users. It is proposed that by leveraging low cost Bluetooth beacons and the users smart phone it is possible to add a layer of cognition that will allow the user to build a spatial map of the surrounding environment and ultimately enhance personal autonomy and accessibility rather than just providing directional information for navigation. The solution is distinct from way-finding or navigation and should be considered as a platform that provides additional context about the environment itself through direct or indirect interaction. The remainder of the paper is structured as follows: Section 2 will provide an overview of existing approaches for navigation and interaction with the user. Section 3 will present an overview of the proposed solution. Section 4 will provide an overview of an example use case for the developed technology and Section 5 will conclude the paper.

## **2 Spatial Mapping & Navigation Support**

### **2.1 Spatial Mapping**

An individual generates a spatial map using a number of different sources, the main source of information comes from the visual system, senses such as vision, smell, movement and hearing are all used to infer a person's location within their environment and as they move through it. It also allows a person to create a navigation path through or a vector that represents a person's position and direction, specifically in comparison to an earlier reference point. Directional cues (e.g. signs, arrows, labels) and positional landmarks (entrances, exits, meeting point) all provide valuable input to allow a person

create a spatial map and can be used both when an individual is static and when determining movement paths and also dynamically while a person is moving through the space. Positional landmarks are generally used to compare the relative position of specific objects, whereas directional cues give information about the shape and layout of the environment itself. We rely heavily on our vision to map our environment and move safely. For a ViB person they must rely on their other senses with touch, hearing, and smell becoming the more dominant senses in mapping their environment and they use items such as a long cane as an obstacle detector or a guide dog as an obstacle avoider. Wall edges, and kerbs are used as a navigational tool and support straight line principle. In addition, over 80% of persons registered blind have some residual vision, and as such colour contrast enhances perception and aids way finding. Textured surfaces can act as a warning and indicate particular types of situations including pedestrian crossings and location of stairs or escalators. For a ViB person, to create a mental model or representation of the environment around them they must decode and aggregate information about their relative location and leverage knowledge of attributes of the spatial environment. This is generally built dynamically firstly by creating a bearing map, which represents space through self-movement and gradient cues for example using a cane can create a rough 2D map of the environment, this can be combined with specific positional cues, to sketch a mental map by integrating specific objects or landmarks with their relative locations to create “minds eye” view of the environment. The process of navigating for ViB can be mentally exhausting particularly in unfamiliar environments.

## 2.2 Navigation Support Tools

Navigation and wayfinding GPS applications such as Google maps have been adopted for many years by the mainstream for independent travel when mapping data and satellite transmission is available. Whereas popular outdoor navigation apps such as Ariadane<sup>1</sup>, GetThere and BlindSquare previously have been developed specifically for people with visually impairments. Most smartphones and tablet devices are GPS and Bluetooth enabled, therefore allowing developers to create applications, which take advantages of location-based technologies and services. Taking advantage of mobile devices that are already embedded with GPS and positional sensing technologies (gyroscope, accelerometer, digital compass, IMU etc.) can be cost effective as it eliminates the requirement to procure install and maintain tracking and sensing technologies. However, although GPS is the most widely used real-time location system, it relies on continuous signal transmission from several satellite source therefore it does not work well indoors or within closed environments where there is significant signal interference. In addition, orientation supported by GPS can be inaccurate and disorientating for the user as a result. Within a closed indoor setting and where navigational and contextual audio-based information needs to be triggered at a more precise location and time, alternative location tracking technologies and methods need to be considered. For example, without precisely tracking a mobile device’s location and pose (proximity and

---

<sup>1</sup> <https://www.ariadnegps.eu/>

orientation) relative to a point of interest as the user moves, it would be difficult to provide contextual audio based information relevant to be played when required at the right time and moment. For indoor location tracking most systems are based upon using wireless technologies such as Wi-Fi, Bluetooth, ultra-wideband (UWB), and Radio-frequency identification (RFID). Most indoor location and positional tracking systems use wireless sensor nodes such as tags that emit signals (beacons), typically points of interest or optimal communication areas are embedded or attached with tags or badges (iBeacons, RFID tags) that broadcast signals to receivers (mobile device). There are more accurate indoor tracking system such as the Decawave DW-1000 UWB chip which can achieve high precision tracking of between 10-30cm, however this technology has not become widespread as the hardware is not yet positioned as low cost for mainstream consumers and most smartphones are not UWB enabled. The selection of technology is dependent on several factors: accuracy required for application specific needs, battery lifetime, cost of installation and maintenance and ease of integration with other processes or systems.

### 2.3 BLE Beacons

Bluetooth Low Energy (BLE) beacons have been widely used for indoor tracking, where once a receiver (mobile device) is in proximity of a beacon, content can be triggered where its position can be tracked if within range of 2 or more beacons by processing the distance data. With BLE, location-tracking accuracy can vary but tracking accuracy can be  $<1.5\text{m}$ , they are easy to install and maintain and affordable. Real-time indoor location services (RTLS) have begun to gain wider attraction from many industry domains, where there are many examples from airports and hospitals taking advantage of BLE beacons to help users navigate large indoor spaces, to retailers providing directed, personalized marketing content to shoppers entering their stores. A number of studies have focused on detailed analysis of BLE accuracy in indoor environments and have demonstrated sub-meter accuracy can be achieved [5], however this can vary significantly across different environments and other aspects such as positioning and orientation of the phone on a persons body can reduce the ability to achieve fine grained positioning information. For the application under consideration providing inaccurate positioning information has a much greater adverse effect on a user that is ViB (from a safety perspective). As such the focus of the work presented is not to improve the accuracy of BLE localisation but rather to investigate how solutions can leverage existing proximity data to trigger the provision of key information relating to the surrounding environment for the ViB user. For spatial mapping BLE beacons provide sufficient accuracy for satisfying the criteria to trigger contextual information when the ViB person is within defined proximities of indoor areas (reception, halls, stairwells, room and toilets) and points of interest (doors, signage as potential collision risks). Proximity detection conditions can be determined by adjusting the beacons antenna power, therefore beacons could be set to varying proximity ranges (2m, 10m, 70m), however it has to be noted that if the beacon antenna is powered up for a longer proximity range the lifetime of the device is reduced to only several month, whereas

environmental factors (temperature, beacon placement) will also effect power consumption and reliability of the beacons.

## 2.4 Other Tools

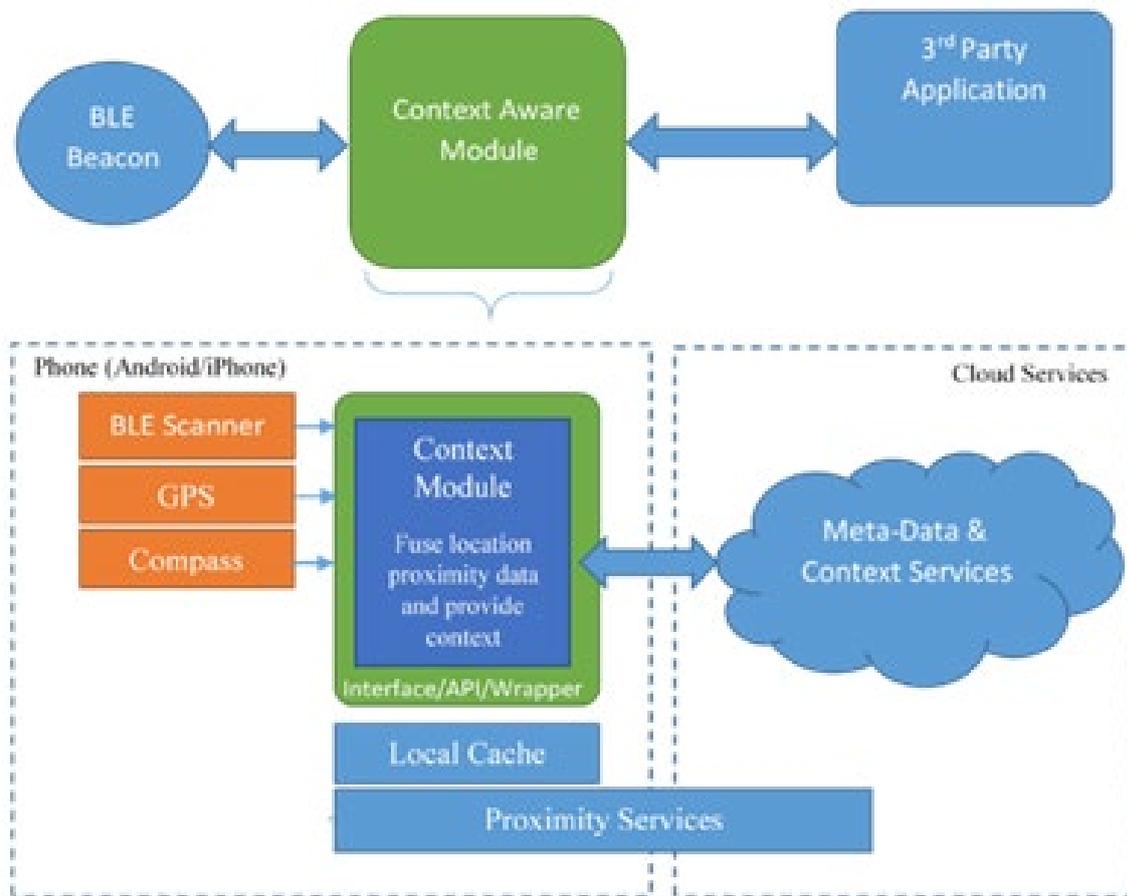
Markers and fiducials can be used to provide additional information, QR codes have become widespread on products and adverts where a person can use their mobile device camera (QR reader) to access further information such as triggering information exchange or even an interactive experience using mobile applications. Essentially, marker-based applications use a devices camera to estimate the position of the device (center point, orientation, range) based upon what it is “seeing”, such as the visual information attained from the fiducial marker. Markers such as QR codes, have a unique predefined shape and pattern that can be easily detected in low lighting conditions and easily printed to be attached to a point of interest. Markers can be an inexpensive and technically simple method for gathering the devices position and therefore provides a very accurate positional cue. For example, BlindSquare, has a QR reader built-in to their app, where they have developed a super-set of the QR barcode matrix purpose built to be more accessible for VIB people when acquiring (scanning) a QR code. For example, the app provides audible and haptic feedback to the user while they are searching and acquiring a QR code. In use cases presented [2] where BlindSquare QR reader is demonstrated, QR code are printed and attached to doors, whereas the user has to find and scan the QR code on the door, where information associated with the room (room name, purpose, member of staff who work there) is read aloud (Voice Over TTS) to the user. The QR codes are placed at optimal locations above the door handle on each door as an early required skill for cane-travel is to trail walls, discern doors and locate door handles so placing illuminating information nearby is helpful. Whereas Blind-Square also aim to aid VIB people in finding and scanning QR code through audio and haptics cues, this still requires manual effort, and explicit interaction that is not so intuitive for the user. Natural feature tracking (NFT) is an image-based tracking method that recognizes and tracks natural features (edges, corners, patterns etc) within a scene or object (building, ornament etc.). Therefore, to the user this is a maker-less tracking method as there is no identifiable marker such as an identifiable fiducial marker (QR code, ID Marker) to scan. NFT extract key point descriptors that are associated with an image captured from a camera, where these key points then query a database to identify matching images and those interpret potential position. Using 3D object recognition and augmented reality visioning systems the physical world and contextual information can be rendered more visible to people with vision impairments, e.g. objects and signage could be enhanced though rendering the increase colour contrast, tone, dimensions or brightness of images based upon a particular persons visual impairment condition type. Augmented Reality glasses such as OxSight and AceSight have been developed specifically for people with vision impairments. Simultaneous Localisation and Mapping (SLAM) is a more complex and progressive computer vision method that is currently a very popular topic within the computer vision community. Through a SLAM system and process a device can create a map of its surroundings whilst at the same time have the capability to localize (position and orientation) itself within the map.

### 3 Context Awareness Module

The context awareness module is a set of software services that enables interaction between Bluetooth Low Energy (BLE) devices deployed in the environment, the user via a mobile application and the provision of audio feedback. The objective is to provide positional and directional cues in a format that is easily configured, interpreted, and used to build a spatial map of the surrounding space.

#### 3.1 System Architecture

**Fig 1.** provides a high-level representation of the context aware module. The module provides common functionality for the interaction of existing BLE beacons and devices while also provide an extension point for integration with other applications and services. The context awareness modules are available across multiple platforms including Android and iOS.



**Fig. 1.** Context Awareness Module Components

The base context services and libraries were developed using the Xamarin framework which supports cross-platform compatibility. This included the development of a front

end to support testing and evaluation of the services. In addition, a separate set of libraries were developed using SWIFT and Objective C specifically for the iOS platform, this was to support the integration of the modules with 3<sup>rd</sup> party iOS applications. The module consists of four main components, firstly all interaction is location driven, as such libraries to estimate the location of the devices were developed, once location is established the next component is to map this to specific context data. The last two components are to support management of the system and user interaction.

### 3.2 Location Services

Location services were developed to leverage existing location capabilities available on smart phone platforms (iOS and Android), these include extracting sensor data such as GPS, accelerometer, compass and other location services that may be available on the mobile platforms. This data is fused with the scanning of BLE advertisement packets using existing protocols (iBeacon and Eddystone) that are generated from devices deployed in the surrounding environment and registered with the system. Leveraging this raw data sets, a number of localisation algorithms were investigated and developed to fuse various source of data and to provide an estimate of the users location (i.e. proximity to the beacon). Localisation approaches generally incorporate prior knowledge of the environment, sensor location, coverage fingerprinting and utilise techniques such as map filtering to improve positioning accuracy. BLE provides less precision however offers a sufficient level of accuracy in terms of proximity to the device (far, near, immediate) utilising received signal strength indicator and other metrics. If there are multiple beacons present in the space techniques such as triangulation can be used to provide more accurate estimate of position. While running initial tests with potential end users, privacy was highlighted as a key requirement, to ensure user privacy is maintained the context-aware modules were developed with the following requirements: the system does not record or maintain any historical data on location information, the location estimation is calculated in real-time based on the live information extracted from the environment. The services do not record any identifiable information relating to the user or their personal devices to protect user identity. Only pre-defined beacons are used in processing the user's location, i.e. only "trusted" beacons that have been registered with the system are used for estimating the user's proximity/location. The module only operates in beacon mode so as they do not create any persistent connections to external device or services. From a data processing point of view the processed data i.e. location information/history is not stored locally or on a cloud server once used to provide context data it is purged from memory.

### 3.3 Context Services

The context services use the estimated location information the provision of context information by a combination of predefined meta-data capturing beacon locations, environmental layout and relevant environmental/object descriptors. From a performance perspective the application manages data by a combination of locally caching context information and context services running in a cloud environment. The context services

essentially contain meta-data and information on the locations, e.g. buildings, floors or areas and the beacons, their position and mapping of the context data or action (i.e. user notification) to these devices. When defining the content of context descriptors, it is important to consider how a person can build an image of the environment. The special map can be characterised based the following features of the environment, paths that provide “straight lines” through a city or environment, edges such as walls, kerbs, building boundaries that provide edges that can be followed and guide a person, nodes which represent focal points for people such as crossing points, door entrance, exit or lift. And zones can be large areas where people can congregate (meeting rooms, reception areas, park). While a cane can be used to detect and object, touch is the main source of information and provides insight to the height, size, type of object that is in proximity. People often rely on others to provide a description of a room or space to help construct a representation of the zone, this can be static information about layout of room, position of tables, where sockets are located, things to avoid etc. Any potential risk that may reside in a space needs to be highlighted to the ViB person e.g. steps down, circulation route what to avoid. Generally, there is a need to provide information that enables the user to feel safer and confident and this has to be driven by easier interaction with an emphasis on simplicity. The context awareness module focuses on delivering spatial contextual information to enhance wayfinding information. This is provided as the person’s location is gathered, their proximity to points of interest and objects (potential collision risks) and description about their physical surroundings (space, layout, location of furniture etc). Spatial contextual awareness has been defined as information such as an individual's location, activity, the time of day, and proximity to other people or objects and devices [3], our approach supplements this to also include a description of the functionality of objects in the environment (e.g. opening configuration of doors, width, height of objects) also. As such it aligns with the definition provide by [4] that specifies any information that can be used to characterize the situation of an entity, where entity means a person, place, or object, which is relevant to the interaction between a user and an application. Presenting contextual information to the person must be relevant to the user’s current task and situation. Therefore, for a VIB person visiting an unfamiliar environment for the first time it is necessary to provide spatial contextual based information to enable them to build a mental representation of their surrounding environmental features, while also providing usability information in order to complete tasks (opening doors, lifts, using furniture etc).

### **3.4 User Interaction**

Once context information is constructed driven by a user’s location it must be provided back to the user, the focus was on providing audio-based feedback via the user’s smart phone. As such an application was developed that used text to speech which automatically converted the context data to audio relayed to the user via headphones or speaker. Through engagement with the ViB community it was highlighted that audio feedback should not mask other sounds from the environment that are currently used for mobility (e.g. listening for cars, signals at traffic lights), being aware of your surroundings during outdoor environments specifically is a necessity for safe navigation. To address this

concern the use of bone conducting headphones to relay audio back to the user was investigated. These headphones are positioned on your cheekbone and do not create a seal the ear canal, this allows a wearer to hear other sounds, or potential hazards coming from the environment while also receiving the audio cues from the context awareness services. It is envisaged the further modes of feedback will also be used such as haptic to provide specific cues to the end user driven by the location information. To support validation the mobile application incorporated a map of the environment where beacons are deployed, and the estimation of the users location is placed on the map while also a list of beacons within the proximity was included to show the id and quality of the signal received as well as the estimated proximity to that beacon. To simplify the specification and collection of context data a context information model was defined, this allows for a common representation of the data is captured, prioritised, and relayed to the end user. The model enables more flexibility in how context information is defined by the deployer and delivered to the end user, e.g. prioritise information based on distance to an object. The model can be linked to different layers of the environment, building, floor, regions, objects, or beacon proximities.

### 3.5 Content Management System

To support the management of BLE infrastructure a web-based content management system was developed, this allows the user (e.g. deployer of BLE beacons) to map the real position of BLE beacons to locations mapped out in the environment the context awareness module will operate in.

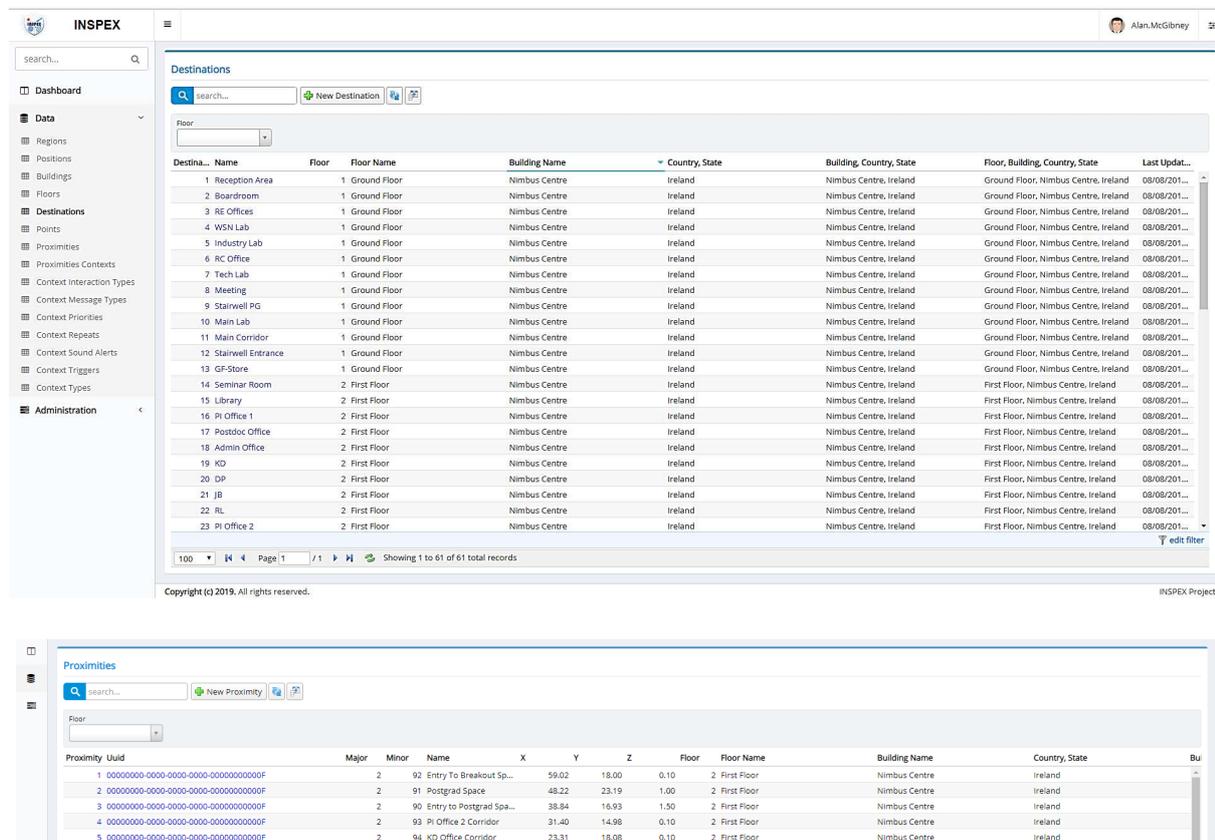


Fig. 2. UI to allow the definition of Buildings, Floors and Destinations

For example, for an indoor environment a user can define a set of beacons along typically used paths and specify the type of interaction expected by the end user. **Fig. 2** (top screen) presents the user interface to define indoor destinations that are linked to a particular building and floor. This information allows the interaction algorithms to not only estimate the location from a coordinates perspective but to link the users position to a more descriptive representation of where in the environment they are, such as room number, name or area description. **Fig. 2** (bottom screen) provide a view of the interface listing the proximities or beacon identifiers, this captures the unique identifier of the beacon and positions it within the environment that can be used to infer the users location when an advertisement packet is received identifying a particular beacon. It also provides the list of beacons that are considered by the application so as not all beacons that may be deployed in the environment are scanned by the application and it is limited to specific devices only.

**Fig. 3** shows how the user via the content management system captures the context information model. The configuration is linked to a parent attribute (proximity, floor, building etc) and stored as part of the context aware services. They can be updated at any time and adjusted as needed by the user, the context services will update its cached data to refresh this data automatically meaning the ViB person will always have the most up to date and relevant information regarding the environment. This flexibility is essential particularly in scenarios where dynamic obstacles can be moved to new locations or new configurations of spaces might be common (e.g. event or meeting room). For indoor environments it is possible to define proximities within the structure using local coordinates, this requires a geometric representation of the building or environment where the beacons will be deployed. In addition where the environment description is not available the beacon positions can be defined using GPS coordinates, the position can then be converted to local coordinates if a representation of the building becomes available, these positions provide a visual context for the deployer to support the planning and setting up of the context path in a site specific scenario.

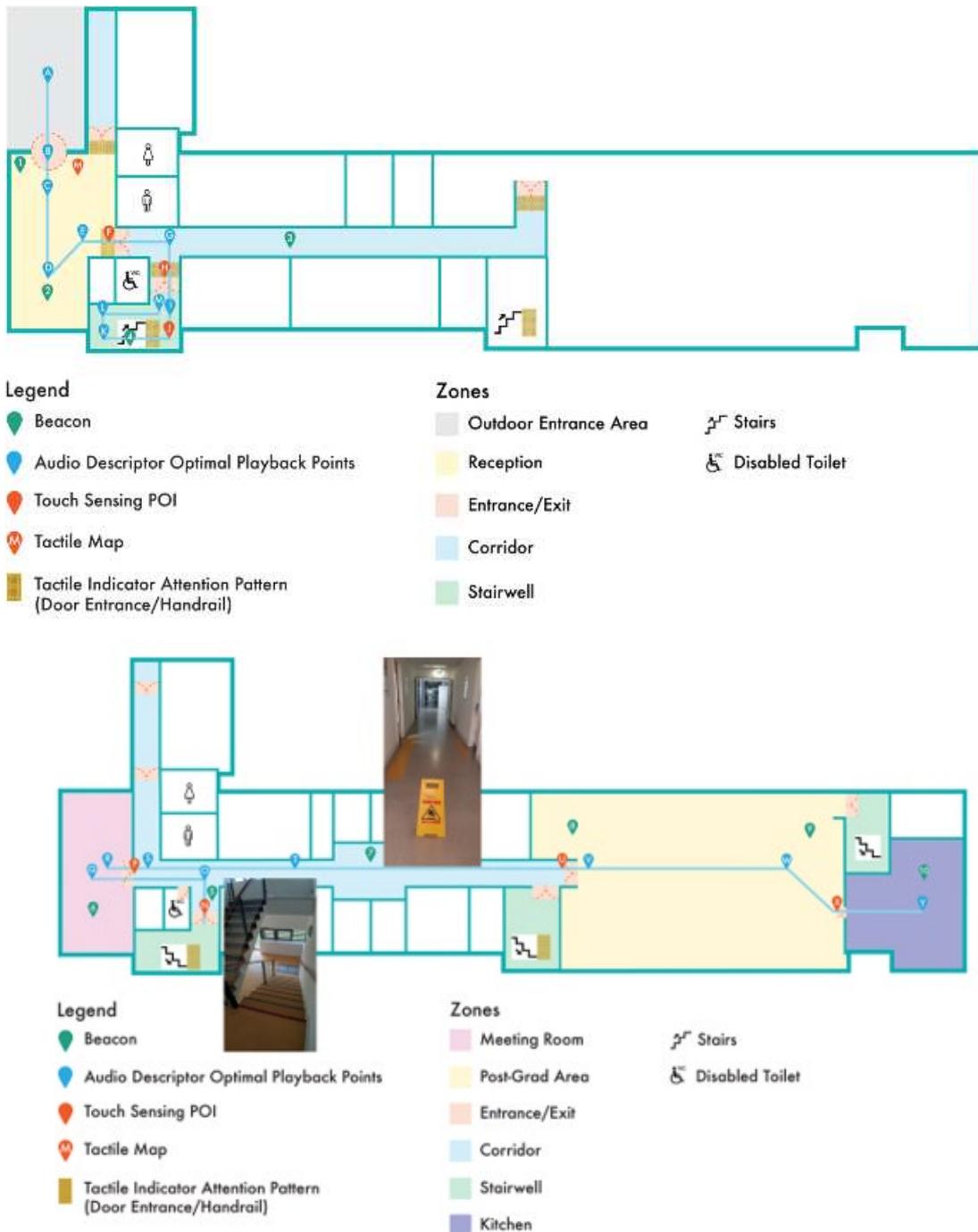
**Fig. 3.** Context meta-data definition

## 4 Use Case Example

The following scenario was considered as an example of how the context-awareness module in an indoor context. A person who has sight loss has confirmed that they will be attending a meeting at a facility they have never been to before. They have contacted the meeting coordinator who has scheduled the meeting and has also gathered any requirements they may have to aid their appointment prior to their visit. Prior to the meeting the building administrator will use the content management system and application to specify where beacons are deployed and provide the configuration needed to facilitate the provision of audio messages (wayfinding instructions, meeting room contextual information, collision risk alerts) to aid the ViB user's visit within the unfamiliar indoor environment. The objectives are as follows:

- Provide a mechanism that offers ViB users a customised, intuitive, and independent way of getting around an indoor facility.
- Provide meaningful audio descriptors that inform the user about the environment characteristics and context (space/room function, size, layout, objects therein etc...)
- Alert users to potential collision risks within the environment (head collision, slip hazards).

Firstly, a spatial map is specified within the context of the target building (**Fig. 4**), this outlines how a user may move through the space to understand the level of graduality for context information and the types of interaction that may be required. Directional and positional cues are captured based on a review of the building, this includes pre-existing cues such as tactile mats, definition of entrances, doors, potential risks and hazards. Any mobile objects deployed in the environment are tagged with a specific beacon. Potential navigation paths are outlined and generated based on point to point trajectory between nodes, zones and landmarks. Contextual description transcripts for the various indoor space (reception area, corridors, meeting room, toilets etc.) were specified, information relating to navigation followed the open standard ITU-F.921 (03/2017) Audio-based network navigation system for persons with vision impairment, that provides recommendations relating to how audio-based navigation systems can be designed to ensure that they are inclusive and meet the needs of persons with visual impairments. The placement of beacons and proximity range need to be carefully considered and optimized in order to ensure appropriate contextual audio-based information can be triggered at the right time and location. For example, it would not be advisable to set the distance range of a beacon to 20m for triggering contextual audio information related to a specific room door in a large space consisting of many other doors, as it would be difficult for the ViB visitor to determine which door the information relates to. As for approaching larger outdoor buildings it could be more optimal to set the range at larger distance such as when the visitor is in proximity to a building or site. When selecting a building, it would be recommended to review the topology of the buildings space and for each areas and point of interest determine the proximity ranges and conditions for triggering playback of wayfinding and contextual audio-based information.



**Fig. 4.** Spatial Layout of Target Building over two floors

**Fig. 5** provides an example of a zone in the target building that had beacons deployed to provide additional context information. The reception area is unmanned and there are several obstacles are present including low level furniture, plants, chairs and display cases that need to be highlighted to the user.



**Fig. 5.** Example area with many obstacles BLE was deployed

The context descriptors aligned to **Fig. 5** are defined as follows: the audio is generated based on proximity to the entrance and follows the flow of messages as the user enters the main door of the building: 1. “You have arrived at the entrance of the [Building] reception area heading towards the reception desk” This provides both positional cue in terms of location and also directional cue. 2. “Please beware of the carpet mat and furniture just ahead of you located in the centre of the reception area” this provides information of a collision risk to the user. 3. “The reception desk is located straight ahead. Located to the left of the entrance are accessible toilets” this message provides information relating to the surroundings. 4. “Located directly left of the reception desk is a secure double door leading to the corridor on the ground floor.” The final message delivers information about next possible course of action and highlights an intersection point between zones/spaces that must be considered. Beacons are then strategically placed at other points in the building, on the entrance to new spaces. The emphasis was placed on providing context information relating to high risk objects such as stairwells and dynamic obstacles that are introduced within the environment. Beacons are deployed in these zones and attached to obstacles (e.g. floor sign as depicted in bottom section of Fig 4.) and mapped to specific context descriptors such as “[Collision Risk] Caution, wet floor sign directly ahead, proceed with caution.” The following flow of events are enabled through the use of the context-awareness module:

- The building is already equipped with BLE beacons, these beacons are already mapped to specific contextual data as described above.
- The ViB person downloads and installs the mobile application to their smart phone prior to arrival. The context services download the meta-data and context information based on regional location.

- When the user arrives at the building they come into range of a beacon and beacon signal is received, an estimated location is calculated and associated contextual information is generated and provided to the ViB person (via headset or phone speaker).
- When the user comes into proximity of specific objects (doors, posters, tactile indicators) they are provided with an audio descriptor, further interaction is supported via user touch.
- The user can find the meeting room location. Furthermore, the user is provided with contextual audio descriptors (where am I, describe surroundings) to allow them to build a spatial map of their surroundings
- The user can navigate and explore their environment confidently and independently.

As part of a user centred design process, an initial qualitative evaluation of the proposed solution was undertaken with a number of representative users (ViB individuals) as part of an observational study. The users operated the system under real conditions allowing us to understand the benefit of the solution from a technology and usability perspective. This provided valuable feedback that was used to inform the subsequent technology design iterations. Initial tests demonstrated the need to reduce the amount of information being delivered to the user, initially the information was very descriptive however due to mobility patterns of users the amount of time required to deliver this level of detail was too short and the user had already moved to another part of the space, resulting in them receiving data that was not relevant to their current position. This also had an impact on the cognitive load of the user. This was somewhat addressed by the modification of the triggers within the context model, i.e. the administrator was able to provide short bursts of information at different proximities (far, near and immediate) to the beacons as well as prioritise critical messages such as collision risks. It was also found that the responsiveness of the users action and the provision of context data was influenced based on the type of device the user had and its location on their person (e.g. in pocket, in hand etc), as such it is not possible to use BLE beacons alone to provide precision navigation steps however they offer sufficient accuracy for the provision of additional descriptive information that allows the user to understand how they could move through and interact with the space they are in. This deployment provides a testbed environment to evaluate the capabilities of the context awareness module and further tests will be carried out in collaboration with ViB people to ensure the solution is useful and reliable for the end user.

Additional tests are required with a wider cohort of users from the ViB community to ensure that a broader performance assessment can be conducted with individuals that have different capabilities, expectations and usage requirements to ensure the solution can adapt to their specific needs. Therefore, personalisation is an important criterion, every individual has different capabilities and needs, however this emphasises another critical consideration, protecting the privacy of the user. While personalisation is required it must be delivered in a privacy preserving manner (e.g. leveraging edge processing, anonymisation etc) that will impact the system architecture.

## 5 Conclusion and Future Work

The context-awareness module leverages low-cost BLE devices and existing infrastructure to provide additional cues and information to a ViB person that can support them in building a spatial map of the environment they are moving through. This has the potential to provide the user with more confidence when moving through and interacting with environments that are unfamiliar to them and offer a better level of experience in these spaces including being more aware of their surroundings and safer mobility. Future work includes the integration of the context-awareness module with other modes of interaction and sensors for example touch that can generate events and automate interaction with other smart connected systems (e.g. seamless access control). In addition, the use of BLE has gained significant attention due to the COVID-19 pandemic, it has obvious applications to support contact tracing and as such a number of protocols have emerged extending existing BLE and localisation approaches to be utilised for this purpose in a privacy preserving manner. The solution proposed here can be extended to this application, in addition it provides a mechanism to support spatial analysis and utilisation management for indoor environments, i.e. it can be used to understand patterns of use within buildings, provide information to users on how to navigate and interact with the environment considering constraints such as social distancing rules etc, and also has the potential to support organisation to digitise space management, workflows and site access tracability etc.

## Acknowledgement

The work presented relates to EU project INSPEX which received funding from the EU's Horizon 2020 research and innovation programme under grant agreement No 73095, and from the Swiss Secretariat for Education, Research and Innovation (SERI) under Grant 16.0136 730953

## References

1. RRA. Bourne, SR . Flaxman, T. Braithwaite, MV. Cicinelli, A .Das, JB. Jonas, et al.; Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob Health*. 2017 Sep;5(9):e888–97.
2. BlindSquare. 2019. Pioneering accessible navigation – indoors and outdoors. [ONLINE] Available at: <https://www.blindsquare.com>. [Accessed 25 January 2020].
3. Chen, Guanling, and David Kotz. 2000. A Survey of Context-Aware Mobile Computing Research. Dartmouth Computer Science Technical Report TR2000-381.
4. Dey, Anind K. 2001. Understanding and Using Context. *Personal and Ubiquitous Computing*, Volume 5, 4-7. Springer London.
5. K. Phutcharoen, M. Chamchoy and P. Supanakoon, "Accuracy Study of Indoor Positioning with Bluetooth Low Energy Beacons," 2020 Joint International Conference on Digital Arts, Media and Technology, Pattaya, Thailand, 2020, pp. 24-27, doi: 10.1109/ECTIDAMTNCON48261.2020.9090691.

# SenseBot: A Wearable Sensor Enabled Robotic System to Support Health and Well-being

Luigi D'Arco<sup>1,2</sup>, Huiru Zheng<sup>2</sup>, Haiying Wang<sup>2</sup>

<sup>1</sup> Department of Computer Science DI, University of Salerno, Salerno, Italy  
l.darco12@studenti.unisa.it

<sup>2</sup> School of Computing, Ulster University, Newtownabbey, Antrim, UK  
{darco-1, h.zheng, hy.wang}@ulster.ac.uk

**Abstract.** The exponential growth of the technology industry over the past years has led to the development of a new paradigm. Pervasive Computing, which seeks to integrate computational capabilities into everyday objects so that they can communicate efficiently and perform useful tasks in a way that minimises the need for the end-users to interact with computers. Along with this paradigm, new technologies, such as wireless body area sensor network (WBASN) and robotics, have been rapidly growing. Such innovations can be used as health and wellness enablers. This paper introduces a system to support the health and well-being of people in both in-door and out-door environments. The system consists of an app, a robot and a remote server. The app and the robot can be connected to a wristband to monitor movements, physical activities, and heart rates. Furthermore, the app provides functions for users to record and monitor their calories intakes, and completed workouts. The robot is equipped with speech capabilities which, integrated with the emotion recognition algorithm, provide supportive feedback to the user. The proposed system represents an early step towards automated care in everyday life which opens the doors to many new scenarios, such as for elderly people who need help but live independently or also for people who would like to improve their lifestyles.

**Keywords:** Wearable Sensor · Robotic System · Emotion Recognition · Speech Recognition

## 1 Introduction

In 1991 Mark Weiser introduced his vision about the evolution of computer in the next century and the possibility of using computers and other devices in our lives without the user being aware of it [1]. At the time, this vision was too futurist, because the progress in computing was not so profound, however, the progress made in the last twenty years in both hardware and software has opened the doors to new concepts and approaches. The early big “computing machines” have been reduced to minimum levels enabling to embed computers in many parts of our environments. Now the term ”pervasive computing” has replaced the old ”ubiquitous computing” and that vision has become real.

Pervasive computing is typically associated with the race to miniaturise hardware with some degree of connectivity, intelligence and mobility [2][3]. With the growth of this paradigm a lot of applications which were initially available at a fixed location only have been transformed into ubiquitous applications, to be used wirelessly and flexibly at anytime and anywhere, and new technologies have been increased, such as the Wireless Body Area Sensor Networks (WBASNs) which defines an autonomous system consisting of several intelligent sensor nodes which monitor without hindering the daily life activities of users [4]. WBASNs have become one of the most promising technology for enabling health monitoring at home, especially for supporting older people's health and well-being, thanks to the low power consumption devices and the possibility to monitor activities, movements and vital human body signals of users continuously and remotely [5] [6]. The world's population is ageing, as provided by the World Health Organization [7] the proportion of the world's population over 60 years will nearly double from 12% to 22% between 2015 and 2050, and different governments are trying to promote home-based elder-care to reduce long-term hospital stays, home care services for elderly people. In several cases, however, to provide healthcare at home is not enough using a WBASN because patients can have physical problems or require additional help, so, new alternatives forms of healthcare are needed. One potential solution is the use of robotics which can provide tangible help for needy people [8] [9].

This paper proposes a solution to support the health and well-being of people in daily living, integrating data from wearable sensors with a robot. The system is composed of three heterogeneous sub-systems: a mobile application, a robotics system and a remote server. The mobile application allows keeping track of the user's everyday activities, such as food consumed and workouts performed, along with his extrapolated health details from a wristband. The mobile application allows also, to check the user trends by the user himself or by a designated parent. The robotic system is designed as an assistance agent capable of speaking, recognising voice commands and collecting fitness data from the user. The robotic system aims to recognise the user's emotion and consequently start a conversation. The remote server is the main source of storage for the system. It stores data from both the mobile application and the robotic system, subsequently, it performs data processing and reasoning. The collaboration between these three sub-systems enables the development of a system that is capable of collecting information from the user via multiple sources of information, offering more accuracy and reliability. The multiple sources of knowledge allow the system to operate in both indoor and outdoor environment, since they are not mandatory together, allowing a greater degree of independence for the user. The objective is to design a system that can be used by most people without any hindrance, keeping costs to a minimum and making the system as simple as possible.

This paper is structured as follow. Related works are described in Section 2. The architecture and implementation of the system are presented in Section 3. The tests are described and evaluated in Section 4, followed by the findings

discussed in Section 5. The paper is concluded by a summary and future work in Section 6.

## 2 Related Work

Thanks to the various improvements in the area of robotics and the miniaturisation of wearable devices, different works can be found in the literature, that combine these two technologies to improve healthcare. Huang et al. [10] proposed an omnidirectional walking-aid robot to assist the elderly in the daily living movement. The robot is controlled during normal walking using a conventional admittance control scheme. When a fall inclination is detected the robot will immediately respond to prevent the user from falling. The fall detection is calculated using the assumption of the human Center of Pressure (COP) and through a wireless sensor, the Center of Gravity (COG) of the user can be approximated.

Goršič et al. [11] introduced a gait phase detection algorithm for providing feedback in walking with a robotic prosthesis. The algorithm is developed as a state machine with transformation rules based on thresholds. The algorithm is finally evaluated with three amputees, walking with the robotic prosthesis and wearable sensors. The studies in which wearable sensors and robotics meshed together are multiple and cover a large area but if we focus only on home healthcare, there are few studies in which it is treated.

Novel robotics and cloud-assisted healthcare system (ROCHAS) was developed by Chen et al. [12]. This study takes as its target users the empty-nesters. The system incorporates three technologies in terms of body area networks, robotics, and cloud computing. In particular, it consists of a robot with the speaking skills that allow the empty-nester to communicate with his/her children, several body sensors that can be deployed in or around the empty-nester, and a cloud-assistant healthcare system that stores and analyses the data provided by the robot. The system helps the empty-nester to be in touch with his/her children and at the same time allowing the children to be mindful of their elderly people's conditions.

Ma et al. [13] developed a healthcare system based on cloud computing and robotics, which consists of wireless body area networks, robots, software system and cloud platform. This system is expected to accurately measure a user's physiological information for analysis and feedback, which is assisted by the robot integrated with various sensors. To boost the viability of multimedia delivery in the healthcare system, this paper proposes a new scheme for transmitting video content in real-time via an enhanced User Datagram Protocol (UDP)-based protocol.

## 3 System Design and Implementation

This section will describe in detail the system architecture, dealing with the communication between the different subsystems.

The proposed system is composed of three main components: remote server (henceforth called PyServer), mobile application (henceforth called *MyFit*), and robot infrastructure (henceforth called PyBot). These three components work together to gather information from a person, store the information and take actions to help the person to be healthy. As can be seen in Fig. 1, the communication between these components is led by the HTTP protocol, and the MyFit and PyBot exploit a Bluetooth Low Energy (BLE) protocol to interface with the wearable sensor. The choice of the HTTP protocol is driven by the need for a protocol that can be suitable on multiple and heterogeneous devices in different locations. On top of the HTTP protocol are created different Representational State Transfer (REST) API architectures, as described below, to encourage the use of an understandable language for all the components, JSON (JavaScript Object Notation), which is an open standard file format, and data interchange format, that uses human-readable text to store and transmit data objects consisting of attribute–value pairs and array data types (or any other serializable value) [14].

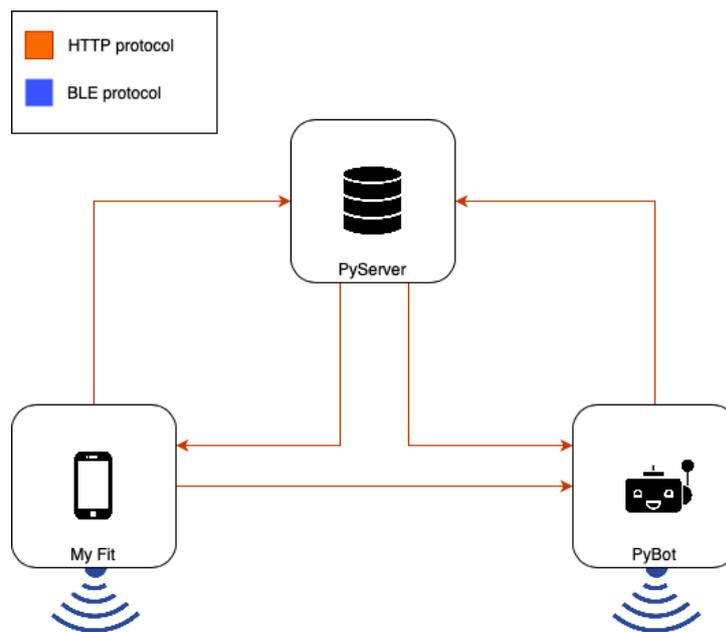


Fig. 1: System behaviour

### 3.1 MyFit

MyFit is a mobile application that can help the overall system to acquire new data from the user during the day. The app is divided into two versions: the *fit version* and the *robot version*.

The main purpose of the *fit version* is to gather information from wristband (steps done, distances travelled, calories burned, heart rates), instead, the main purpose of the *robot version* is to control remotely the PyBot. Leaving aside the

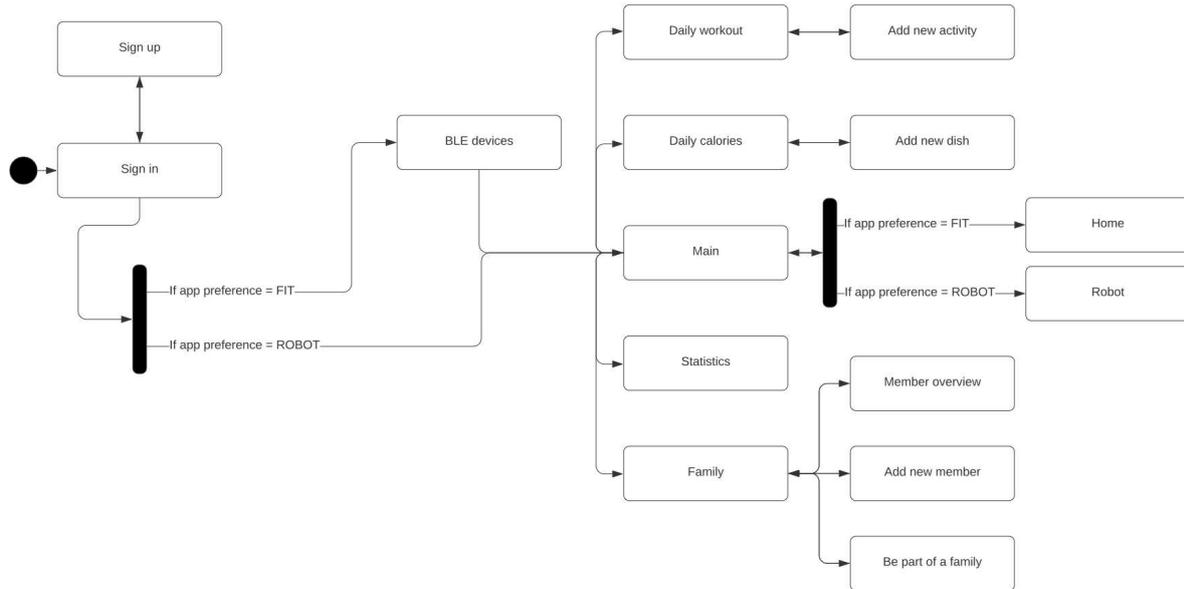
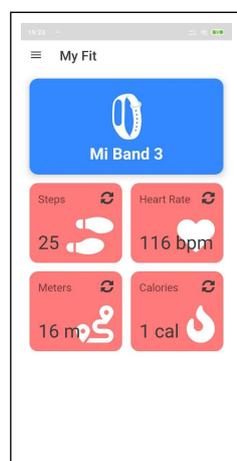


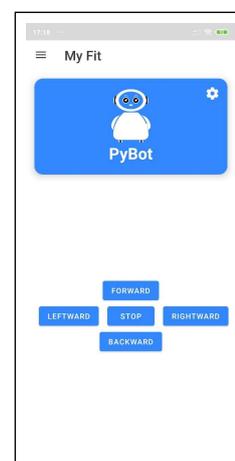
Fig. 2: MyFit: navigational path

differences in the main purpose, the other functionalities of the app are shared for both the versions. The entire navigational path of MyFit is shown in Fig. 2

The app can be accessed only after authentication, then the user, when starts the app, has to sign-in, if already register, otherwise he can sign-up and automatically he is sign-in. After the authentication step if the app run is the *fit version* the user has to choose the wristband from the list of available Bluetooth devices and then it is redirected to the main view where are shown all his fitness data, otherwise, if the app run is the *robot version* it is redirected to the main view where it can control the PyBot. Fig. 3 shows the different views that the user sees according to the version.



(a) fit version



(b) robot version

Fig. 3: MyFit: main views according to the versions

Once the user is in the main view, he can move among other views through the left navigation drawer, activated by the button on the top left. The possible views are:

- *daily calories*: this view allows the user to see and to add the food eaten with the numbers of calorie, in the current day or the past days according to the meal.
- *daily workouts*: this view allows the user to see and to add the activities done with the amount of time spent, in the current day or the past days.
- *statistics*: this view allows the user to see the trend of its fitness data in the last 10 days.
- *family*: this view allows the user to see its family members<sup>1</sup>, add a new one scanning his QR-code, be a family member of another user generating the QR-code and letting him scan it.

MyFit is developed for Android Devices, the devices supported are all the devices that run a version of Android between *Android 7.0* and *Android 10*. It is developed in JAVA and XML.

### 3.2 PyBot

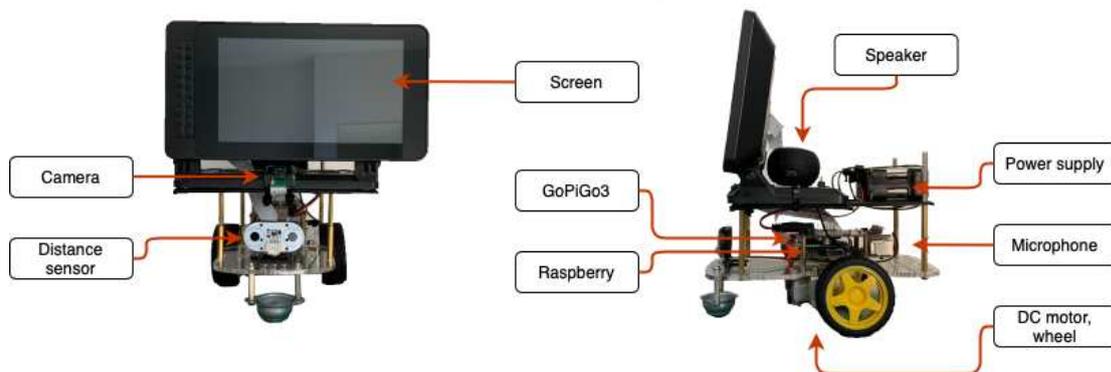


Fig. 4: PyBot: design

The physical design of the robot platform allows to interact in a friendly way with the user, to recognise the user's emotions, and to recognise the fitness activity made by the user. The robot is based on a design made by [17] with different updates. As shown in Fig. 4 the robot structure is composed of layers that can allow modularity so that in future improvements new components can be added easily. The main component of the robot, as well as the core of the robot, is a Raspberry Pi board, this choice is due to a need of maintaining the overall cost of the robot low to reach as many as possible people. Two DC motors

<sup>1</sup> family member: a user who you can check his fitness data trends

with two wheels are combined to allow the robot to move, allowing flexibility on which type of surfaces the robot can travel, thus restricting it to flat surfaces. Since the robot can move, it has incorporated a distance sensor to prevent it from colliding with objects on the front. The DC motors and the distance sensor are managed by the GoPiGo3 board. The robot has to communicate with the user in a friendly way, to reduce the gap between human and machine, for this reason, are integrated a screen that visualises information, a speaker, and a microphone. The robot has to, also, recognise facial emotion of the user, to allow this a Raspberry Pi Camera module is mounted on the front of it.

The robot platform can interact with the user and take decisions thanks to a python program. The program is built of four processes that run parallel on the machine: digital assistant, API manager, fit manager, emotion manager.

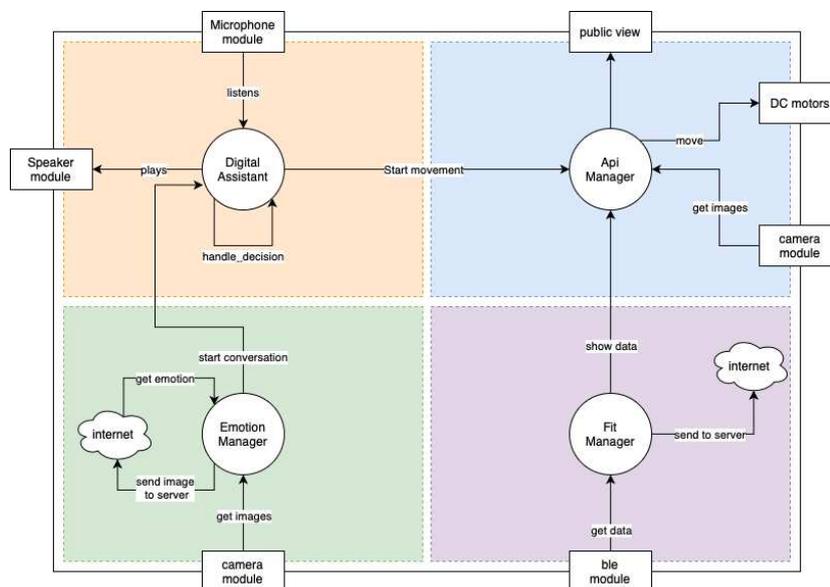


Fig. 5: PyBot processes

**Digital Assistant** This process is mainly responsible for the conversation with the user. The digital assistant can recognise and speak four different languages (English, Italian, Chinese, Spanish) but only one at a time, this can be set at the start-up of PyBot.

When the PyBot starts up the digital assistant begins recording sound. Through the SpeechRecognition library [18] the recording is cleared by the ambient noises, and after thanks to its internal engine, GoogleSpeechRecognition, it tries to detect speech within the recording, if so the speech is converted to text, otherwise, the flow is stopped and the digital assistant starts to record again. When the text recognised is available a matching function between the preset voice commands and the text is applied. If the match has a positive result, the linked action is triggered, otherwise, the flow is interrupted and the

digital assistant starts recording again. The triggered action is followed by a vocal response. The vocal response is performed by converting the English textual response connected to the action into the current language of the PyBot using the googletrans library [19], then the translated response is converted to voice thanks to the gTTS (Google Text-To-Speech) library [20] which creates an mp3 file with the spoken data from text. Finally, with pygame library [21] it is possible to play the created file. This flow is repeated continuously providing effective support to the user.

The examples of recognised commands and the respective responses are:

- *hello*: it says "hello"
- *how are you*: it says "I'm fine and you?"
- *follow me*: it says "Let's go" and starts walking forward
- *turn right*: it says "turn right" and turns right
- *turn left*: it says "turn left" and turns left
- *go back*: it says "ok I'll go backward" and go backward
- *play*: it plays random music

**Api Manager** This process is mainly responsible for providing external APIs. The APIs allows the user to access the PyBot resources from a different location without the need to have the PyBot nearby. The APIs handle different resources:

- *movements*: movements performed by the PyBot
- *stream*: camera stream of the PyBot

These APIs are developed using the Flask framework,

**Fit Manager** This process is mainly responsible for the gathering of data from the wristband. The process can acquire from the user: steps done, distances travelled, calorie burned, heart rates.

The process is based on the library provided by [22] with some updates, that exploits the BLE protocol to connect to the wristband.

**Emotion Manager** This process is mainly responsible for the recognition and handling of user emotions. The process captures every 10 seconds a frame from the camera module and then sends the image captured to the PyServer, which aims to recognise the emotion and send back the result. The way of emotion recognition by PyServer is described in the next subsection. The emotion manager when receives the result from the PyServer, analysis the result, and according to the emotion obtained it starts a conversation with the user.

### 3.3 PyServer

The PyServer is designed to act as the remote server of the system proposed. It is responsible for the storing of the information generated by both the MyFit

and the PyBot, as well as for the operations of emotion recognition.

As mentioned at the beginning of this section the architecture used is the REST API which can use an interchangeable language to manipulate the HTTP protocol and to provide resources.

The APIs handle different resources:

- *users*: users registered in the system
- *foods*: foods eaten by users
- *activities*: training activities done by users
- *steps*: steps done by users
- *distances*: distances travelled by users
- *calories*: calories burned by users
- *heart rates*: heart rates of the users

The PyServer is designed to be secure, provided an authentication strategy, mandatory to access resources, based on a token, JWT. The token allows identifying users and their role, to restrict the resources according to which actor is logged in.

To allow emotion recognition a pre-trained CNN, developed by Serengil [15], is used. The CNN is trained on the FER 2013 dataset [16]. Seven emotions can be recognised: anger, disgust, fear, happiness, sadness, surprise, neutral.

The flow of the activities to predict emotions is shown in Fig. 6.

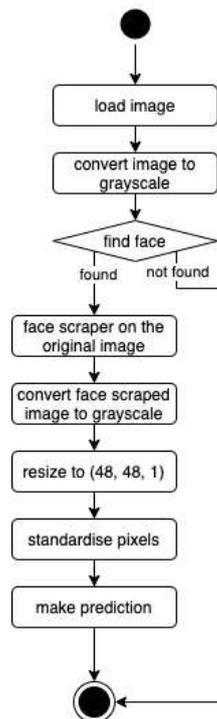


Fig. 6: PyServer: emotion recognition flow

When the emotion recognition resource is called the server expects to receive an image as input. When the image is loaded, the conversion of its colour scheme

is applied, which is changed in grayscale. The segment where the face is located is obtained, if present, by using an OpenCV function [23]. If a face is found, the section where the face is located is cropped to create a new image with only the face. The new image is converted to grayscale, resized to a size of (48, 48, 1) to fit well in the machine learning model, and the pixels in the image are standardised, this means that all pixels that were initially in a range from 0 to 255 are now converted into a range from 0 to 1 to improve prediction performance. Now the image can be used as an input for the machine learning model to obtain the prediction.

The PyServer is developed in Python, the APIs are developed using the Flask framework. For the storage is used a relational database to provide a well-defined structure with multiple relationships.

## 4 Tests and Evaluation

To evaluate the functionality offered by the system, several usability tests had been defined to be submitted to different people to evaluate the ease of use of the system and the correct functioning, but due to the current critical world situation, caused by the spread of a disease on a large scale, the tests could not be completed.

Three macro groups of tests have been identified, one in which the MyFit app is tested, one in which the PyBot is tested, and one in which the facial detection features are tested. For each test, the user had to be asked to complete some activities to test the usability and correctness of the system. The experiments were conducted in the laboratory. Participants were given a set of tasks to test MyFit and PyBot. The following results were recorded:

- Time to complete a task per user
- Number and type of error per task
- Number of errors per unit time per user
- Number of users completing a task successfully

After carrying out the tests a questionnaire on how the users feel about using the product, by asking them to rate it along with a number of scales, after interacting with it.

The facial detection test was carried by one person only, due to the covid19. The person involved is a male of 23 years old. The test was set up in a room of around 10 sqm with no artificial light with few background noises. The person was asked to stand still in front of the PyBot for 4 minutes, two minutes looking at it and two minutes with their backs facing it. The test is repeated four times at different distances: 50 cm, 100 cm, 150 cm, 200 cm. The aims are to evaluate the performance of the underline face detection infrastructure when the user is in front of the robot in both looking and not looking scenarios.

The data obtained are classified into two classes, face recognised (1) and face not recognised (0), so it will be:

- True Positive (TP): the person is looking at the PyBot and the face is recognised
- False Positive (FP): the person is not looking at the PyBot and the face is recognised
- True Negative (TN): the person is not looking at the PyBot and the face is not recognised
- False Negative (FN): the person is looking at the PyBot and the face is not recognised

The confusion matrices from the result of each experiment are shown in table 1.

<i>50 cm</i>		
	<b>not recognised</b>	<b>recognised</b>
<b>not present</b>	14	1
<b>present</b>	0	15

<i>100 cm</i>		
	<b>not recognised</b>	<b>recognised</b>
<b>not present</b>	15	0
<b>present</b>	1	14

<i>150 cm</i>		
	<b>not recognised</b>	<b>recognised</b>
<b>not present</b>	15	0
<b>present</b>	0	15

<i>200 cm</i>		
	<b>not recognised</b>	<b>recognised</b>
<b>not present</b>	15	0
<b>present</b>	2	13

Table 1: Confusion matrices of face detection tests, obtained at different distances among the person and the PyBot: 50cm, 100cm, 150cm, 200cm

Distance	Average execution time (s)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)
50 cm	3.07	96.67	100	93.75	100
100 cm	3.05	96.67	93.34	100	93.75
150 cm	3.02	100	100	100	100
200 cm	3.04	93.33	86.67	100	88.24

Table 2: Performance of face detection

Table 2 summarises the classification performance of face detection in different distances.

The evaluation metrics used are calculated as following:

Accuracy:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision:

$$PREC = \frac{TP}{TP + FP}$$

Sensitivity:

$$SN = \frac{TP}{TP + FN}$$

Specificity:

$$SP = \frac{TN}{TN + FP}$$

Results show that the underlying infrastructure is robust and can also be used for further emotion detection, which is not covered in this paper. However, this must be considered only as a preliminary study as the experiments were conducted on one person and the images collected are fairly few.

## 5 Discussion

The system aims at gathering the user's health data without destroying what is his daily life. Using different heterogeneous systems has allowed reaching a good compromise and it has allowed monitoring the user's activities even when the user decides to leave home. In addition, the use of a robotic system to communicate with the user allows the distance between the user and the system to be minimised making the user feel more comfortable.

To better understand what the advantages and disadvantages of this system are, it is good to analyse the three proposed sub-systems differently.

### 5.1 MyFit

The MyFit is responsible for gathering the major number of information from the user. After considerable and prolonged use, several advantages have emerged. The most important is that the user as soon as he accesses the application starts a background service that collects information from the wristband without the user having to interact with it. The only limitation of this service is that the app must be run in the background even if closed, otherwise, the service will not work until the app is reopened. Another advantage is that the app provides the

user an all in one place to record his daily habits, usually a user has to install different applications. In addition, the user can check his trends, or even more useful can check the trends of a family member. The main issue raised up is the network reliability for the data storage, because there is an algorithm that sends the data to the server and stores in local the information added by the user, but in case of network connection leak, there is no way to understand if the data was been sent to the server or not and if so retry. MyFit can be connected to a wristband but is now only compatible with Xiaomi Mi Band 3, so it can think of expanding compatibility to many more wristbands.

## 5.2 PyBot

The PyBot is responsible for acquiring information from the user, but user interaction is even more useful. PyBot can offer several advantages to the user. The most important thing is that if the user is not happy to wear a wristband, the data can be acquired by the PyBot, leaving the user free of wearable devices, but at the same time under the control of the robot. It is also important that PyBot can be used to encourage the user, for example on a bad day, by providing support through conversation or playing music. During the development different design issues are raised:

- *Energy issue*: the PyBot has several sensors connected to itself, so the average energy consumption is significant. Due to its mobility, it cannot be charged for long and often, otherwise, it becomes useless. Optimising energy consumption is, therefore, the main question that must be solved
- *Network reliability*: the PyBot uses different services over the internet, the emotion recognition provided by the remote server, the text-to-speech and the speech-to-text provided by a third part. Therefore, is necessary to ensure a reliable connection or provide an algorithm that can compute in local.
- *Quality of the modules*: the PyBot integrates different modules, fig 4. The use and the reliability of these modules is a key concept to allow the PyBot to work well. Problems relating to modules are identified during tests. The camera used is not capable of operating in a low light environment and the resolution is poor, so it can lead to user emotion confusion due to low image quality in everyday use. The microphone used takes on a lot of ambient noise, so it can lead to wrong speech recognition. The distance sensor is not able to identify all the obstacles in front of it. The distance sensor utilises ultra-sonic waves that are useful because they are not influenced by object light, colour or transparency, but they are not reflected by soft materials so that the robot sometimes fails to identify the person in front of it.
- *Emotion recognition*: the emotion recognition is one of the most complex machine learning problem because the emotions of people can affect differently the face of each person, so it is very difficult to identify a pattern. To improve the reliability of emotion recognition is possible to integrate with face images also audio recording and analysis of body movements [24].

To solve the design-related issues, just assume the components are easily replaceable thanks to the system's modularity, which can lead to great improvements according to needs.

### 5.3 PyServer

The PyServer is the main source of storage. Since a huge amount of extensive data of the user is collected the privacy invasion will be very serious. After considerable and prolonged usage, it can be established that the system used for authentication, token, enables to avoid possible holes in the system while protecting user data. Furthermore, all data collected are stored following the GDPR guidelines. The other function of PyServer besides storage is to recognise emotions, the server-side emotion recognition can be modified to take advantage of new sources of knowledge according to the previous subsection on PyBot.

In summary, a good level of data collection has been achieved which can lead to different new scenarios. One possible scenario is to create a dataset containing the fitness activities, the foods eaten, the training activities, the emotions belong to the users, related to their health, in order to create a machine learning algorithm that can predict the health status of a user collecting only this information.

## 6 Conclusion and Future Works

A heterogeneous framework for controlling and improving health and well-being is proposed in this paper. A smartphone app, a remote server, a robot and a wristband make up the whole system. The mobile app facilitates the processing of information from the wristband, allowing, also, the user to record the foods consumed and the activities performed. The robot is capable of gathering information from the wristband without user intervention and is capable of understanding the user's emotional state to assist it, additionally the robot's abilities to talk and listen lead to reduce the gap between robot and human. A limitation of the study is the lack of a usability study, due to the closure of university campuses caused by the spread of Covid-19. In fact, to improve the performance obtained one of the most important future works will be to organise test sessions in a controlled environment to consolidate the work done.

In conclusion, it can be noticed that the mutual exclusivity between the app and the robot let the user a greater degree of freedom maintaining a good level of information collection and thanks to the use of the PyBot it is possible to make the user feel safer and more peaceful. Furthermore, the possibility of having a remote view of family members can be used in an easy way to monitor their habits without being too invasive. The proposed system represents an early step towards automated care in everyday life which opens the doors to many new scenarios. The system could be used as an encouragement for people who are reluctant to play sports or other physical activity, motivating them to increase

their participation on days when they are more sedentary, to maintain their good health and to prevent obesity. The system could be used as a help for elderly people, who wish to maintain their autonomy but are required to seek third-party assistance.

Future work will be carried out to incorporate other wristbands with the MyFit app, to improve the facial detection and emotion detection algorithms, and to undertake a large scale of data collection and evaluate the system.

## 7 Acknowledgement

This research is partially supported by the Beitto-Ulster collaboration programme.

## References

1. M. Weiser, "The computers for the 21st century," *Sci. Amer.*, September 1991.
2. D. Saha, and A. Mukherjee, "Pervasive computing: a paradigm for the 21st century," in *Computer*, vol. 36, no. 3, pp. 25-31, March 2003.
3. M. Satyanarayanan, "Pervasive computing: vision and challenges," in *IEEE Personal Communications*, vol. 8, no. 4, pp. 10-17, Aug. 2001.
4. A. Sangwan, and P. Bhattacharya, "Wireless Body Sensor Networks: A Review," *International Journal of Hybrid Information Technology*, vol 8, no. 9, pp. 105-120, 2015
5. R. Dobrescu, D. Popescu, M. Dobrescu, and M. Nicolae, "Integration of WSN-based platform in a homecare monitoring system," *International Conference on Communications & Information Technology (CIT'10)*, pp. 165-170, July 2010.
6. H. Pei-Cheng, and W. Chung, "A comprehensive ubiquitous healthcare solution on an Android<sup>TM</sup> mobile device," *Sensors (Basel, Switzerland)*, vol. 11, June 2011.
7. World Health Organization, "Ageing and health," February 2018. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>.
8. G. Wilson, C. Pereyda, N. Raghunath, G. De La Cruz, S. Goel, S. Nesaei, B. Minor, M. Schmitter-Edgecombe, M. E. Taylor, D. J. Cook, "Robot-enabled support of daily activities in smart home environments," *Cognitive Systems Research*, 54, pp. 258-272, October 2018.
9. M.J. Mataric, "Socially assistive robotics: Human augmentation versus automation", *Science Robotics*, pp. 1-3, 2017.
10. J. Huang, W. Xu, S. Mohammed, Z. Shu, "Posture estimation and human support using wearable sensors and walking-aid robot," *Robotics and Autonomous Systems*, vol. 73, pp. 24-43, 2015
11. M. Goršič, R. Kamnik, L. Ambrožič, N. Vitiello, D. Lefeber, G. Pasquini, and M. Munih, "Online Phase Detection Using Wearable Sensors for Walking with a Robotic Prosthesis," *Sensors*, vol 14, pp. 2776-2794, 2014
12. M. Chen, Y. Ma, S. Ullah, W. Cai, E. Song, "ROCHAS: Robotics and Cloud-assisted Healthcare System for Empty Nester," *8th International Conference on Body Area Networks, BodyNets*, pp. 217-220, September 2013
13. Y. Ma, Y. Zhang, J. Wan, D. Zhang, N. Pan, "Robot and cloud-assisted multi modal healthcare system", *Cluster Comput* 18, 1295–1306, May 2015

14. Wikipedia contributors, "Json," April 2020, [Online]. Available: <https://en.wikipedia.org/wiki/JSON>
15. S. I. Serengil, "TensorFlow 101: Introduction to Deep Learning for Python Within TensorFlow". [Online]. Available: <https://github.com/serengil/tensorflow-101>
16. "Challenges in Representation Learning: Facial Expression Recognition Challenge," 2013. [Online]. Available: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/overview>
17. R. Sridhar, H. Wang, P. McAllister, H. Zheng, "E-Bot: A Facial Recognition Based Human Robot Emotion Detection System," Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI), July 2018
18. "SpeechRecognition 3.8.1," [Online]. Available: <https://pypi.org/project/SpeechRecognition/>
19. "googletrans 2.4.0," [Online]. Available: <https://pypi.org/project/googletrans/>
20. "gTTS (Google Text-to-Speech)," [Online]. Available: <https://pypi.org/project/gTTS/>
21. "pygame 1.9.6," [Online]. Available: <https://pypi.org/project/pygame/>
22. Y. Ojha, "MiBand3," [Online]. Available: <https://github.com/yogeshojha/MiBand3>
23. "opencv-python 4.2.0.34," [Online]. Available: <https://pypi.org/project/opencv-python/>
24. M. El Ayadi, M. S. Kamel, F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, Volume 44, Issue 3, pp. 572-587, 2011,

# SDR Demonstration System for the Investigation of Cooperative Communication and the Scaling Behaviour of MANETs

Mus'ab Yüksel<sup>1</sup>, Milenko Starcik<sup>1</sup>, Florian Hofmann<sup>1</sup>,  
Raphael Rolny<sup>2</sup>, Marc Kuhn<sup>3</sup>, Michael Kuhn<sup>1</sup>

<sup>1</sup> Hochschule Darmstadt (h-da), Darmstadt, Germany

<sup>2</sup> armasuisse Science & Technology, Thun, Switzerland

<sup>3</sup> ETH Zurich, Zurich, Switzerland

**Abstract.** Efficient broadcasting is very important for currently poorly scaling mobile ad-hoc networks (MANETs). Cooperative transmission protocols can significantly improve the scaling behaviour. This paper focuses on the implementation of a scalable MANET demonstration system with multiple software-defined radios (SDRs) which allows to thoroughly investigate different cooperative transmission and broadcast techniques and the improvements which can be achieved in practical systems. Measurement results show the advantages of cooperative broadcasting, distributed transmit diversity, and signal-to-noise ratio (SNR) aggregation.

**Keywords:** Demonstration-System · Cooperative Communication · Distributed Transmit Diversity · Space-Time Block Codes · Alamouti · Software Defined Radio (SDR) · Signal-To-Noise-Ratio (SNR) Aggregation · Cooperative Broadcast · Mobile Ad-hoc Network (MANET) Scalability · Quadrature Amplitude Modulation (QAM), Orthogonal Frequency Division Multiplexing (OFDM), Filter Bank Multi-Carrier (FBMC)

## 1 Introduction

Mobile ad-hoc networks (MANETs) play a vital role in mission critical communication systems that should deliver reliable communication means in all situations, e.g. if conventional networks (LTE, Wi-Fi, wired broadband systems, etc.) are not available or collapse. As such systems cannot rely on fixed infrastructure such as wired routers or managed access points, they have to be designed in a decentralized manner, should be self-organizing, and avoid single points of failure. Advantages of MANETs over managed networks include flexibility (they can be built everywhere), low administration costs (no need to build infrastructure), self-healing ability (new paths can be established if links break or nodes disappear), and scalability (networks can easily be extended by more nodes) [1]. Due to their dynamic and flexible nature, MANETs are thus widely used in secured and robust networks for public authorities, disaster rescue, transportation, unmanned vehicles, sensor networks, or military communications [1, 2].

As shown in [3], the accumulated throughput in a MANET with  $N$  nodes can however not grow larger than  $O(\sqrt{N})$  when the messages are forwarded in a unicast multi-hop fashion. This means, such networks scale poorly with an increasing number of nodes. Besides this, also the necessary network organization introduces a growing overhead. As each node can move independently and in any direction, the links between nodes may change frequently. The multi-hop forwarding thus requires sophisticated routing algorithms. Especially in large dynamic networks where the topology is constantly changing, routing information has to be shared among all nodes in the network to establish and maintain correct and up-to-date routing tables [2, 4, 5]. The overhead introduced by sharing routing information can thus lead to a significant decrease in network performance (e.g. long delays, throughput degradation, connectivity problems, and others). In case of a proactive routing strategy, typically *Hello* and *Topology-Control (TC)* messages are distributed. The TC messages are broadcast (or multicast) messages sent from each node to all (or a subset of) other nodes in the MANET. Hence, effective and efficient broadcasting is very important to implement reliable and efficient routing strategies.

With cooperative transmission protocols, the scaling behavior can significantly be improved: Information theoretical work has shown that a scaling law of  $O(N \log N)$  is achievable [6]. The proposed cooperative transmission method of *distributed hierarchical multiple-input multiple-output (MIMO) transmission* is, however, of high complexity and difficult to implement in practice. In order to improve the scaling behavior of practical MANETs, some other cooperative approaches have been developed. Examples are *multistage cooperative broadcast* [7] and *barrage relaying* [8]. These cooperative approaches focus on broadcast (or multicast) instead of unicast communication. Thereby, all nodes which were able to correctly decode the message transmitted by a source node in a first time slot support the transmission in the next time slot by re-transmitting the same message simultaneously. Nodes that could decode the message in the second time slot start to re-transmit it in the following time slot and so on until all intended nodes receive the message successfully. In order to achieve a diversity gain, [7] proposes to use a distributed transmit diversity scheme such that the different signal contributions add up in power at the receiving nodes and the messages spread through the network quickly. Barrage relaying [8] follows a similar approach but applies a specific phase dithering scheme in addition to turbo-like error correction. Each transmitting node pseudo-randomly dithers its carrier phase, such that the superposition of these signals induces a time-varying channel characteristic at the receiving nodes. An error correction code is then used to extract time diversity provided by the time-varying fading channel.

Compared to classical broadcasting, such cooperative communication methods can achieve large gains by reducing the required number of time slots to spread messages to all (intended) nodes. Improving the efficiency of broadcasting in a MANET can also enhance the scaling behavior of the network when unicast traffic is considered. This is because the overhead introduced by routing is one main reason for the poor scalability. Establishing and updating routing

tables is often based on broadcasting. Having a more efficient spreading of the routing information thus results in more resources available for transmitting unicast messages. Therefore, cooperative transmission schemes can improve the performance of MANETs in various ways:

- To quickly spread messages to multiple (multicast) or all (broadcast) other nodes in the network.
- To efficiently establish and refresh proactive routing tables in a MANET, specifically to distribute Hello and TC messages.
- To distribute the message of a source node to its neighborhood to form a virtual MIMO (VMIMO) cluster that can further improve the performance by spatial multiplexing (cf. distributed hierarchical MIMO transmission [6]).

**Contribution:** In this paper, we focus on the implementation of a demonstration system to study different cooperative transmission and broadcast techniques and the gains that can be achieved with them in practice. The developed system is based on multiple software-defined radios (SDRs), is very flexible, implements single-carrier quadrature amplitude modulation (QAM), orthogonal frequency division multiplexing (OFDM) as well as filter bank multi-carrier (FBMC) modulation, and allows for comparison of all combinations of the implemented modulation and cooperation schemes. Using this demonstration system, we develop, implement, and study practical and efficient MANET broadcasting schemes based on cooperative transmission protocols (e.g., distributed transmit diversity schemes) for multi-carrier systems.

In a first step, we study a distributed Alamouti scheme [9] and extend it to a combination with signal-to-noise ratio (SNR) aggregation for cooperative broadcasts in a single-carrier setup. With SNR aggregation, each receiver node stores two or more observations of the broadcast message and, eventually, combines them to enhance the probability of successful decoding.

In a second step, we show that the system is very flexible and scalable regarding both the number of participating nodes and number of multi-carrier tones. Furthermore, we outline how these schemes can drastically improve the broadcasting in MANETs.

## 2 Design and Setup of the Demonstration System

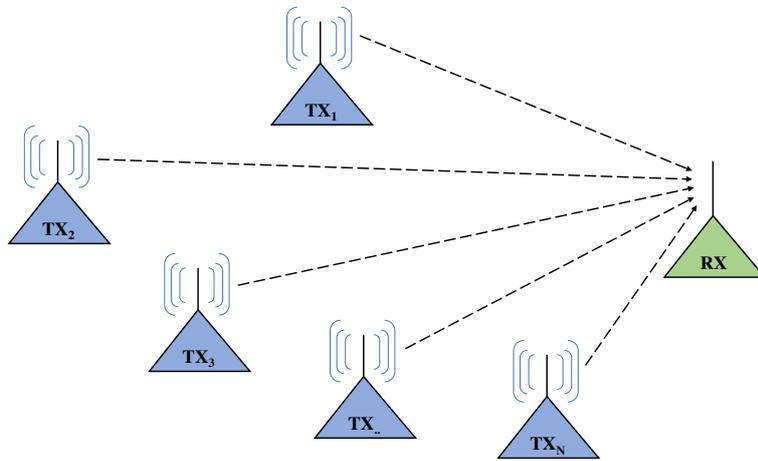
To build a scalable and flexible MANET demonstration system, different hardware platforms and development environments have been investigated to find the most suitable selection. Thereafter, first a simulation environment has been developed to prove that the signal processing and principal signal chains are working as expected. Second, the software has been adapted on real hardware using SDRs. Last, some measurements have been conducted to prove the concept and show that the demonstration system can be used to verify and demonstrate various cooperation schemes.

Within this section, the system model used for the demonstration system and a brief overview of the state of the art will be provided. Thereafter, the

hardware and software setup will be explained in detail. The section will end by discussing the transmitter and receiver implementation.

## 2.1 System Model

For the demonstration system we consider a distributed multiple-input single-output (MISO) system consisting of multiple source nodes  $\text{TX}_n, n = 1, 2, \dots, N$ , and a single destination node  $\text{RX}$ , all equipped with a single antenna as shown in Figure 1. Distributed MISO systems are often considered to be a promising approach in MANETs as several nodes can assist each other in transmitting data and thus increasing the range and reducing the outage probability.



**Fig. 1.** System model for the demonstration system.

The transmit symbol vector  $\mathbf{x} \in \mathbb{C}^{1 \times N}$  comprises the symbols of all cooperating nodes. The structure of  $\mathbf{x}$  depends on the cooperation scheme used by the nodes (e.g. distributed transmit diversity or distributed beamforming). The single antenna receiver  $\text{RX}$  observes the received symbol  $y \in \mathbb{C}^{1 \times 1}$  with

$$y = \mathbf{h}^T \cdot \mathbf{x} + n, \quad (1)$$

where  $\mathbf{h} \in \mathbb{C}^{1 \times N}$  is the channel vector assuming a narrowband channel model in the equivalent baseband, comprising the complex channel taps between each transmitter and the receiver. The scalar  $n \in \mathbb{C}^{1 \times 1}$  models additive white Gaussian noise.

The demonstration system is designed such that it allows to study different forms of transmit cooperation combined with different modulation schemes, as e.g., single-carrier modulation, OFDM, and FBMC. Thereby, we are particularly interested in cooperation techniques that apply transmit diversity.

## 2.2 State of the Art

In literature, different transmit diversity techniques are proposed to mitigate the effect of short term fading. Examples of such techniques include delay diversity schemes [10], space-time codes, space-frequency codes, phase roll diversity schemes [11]. In the following, we focus on space-time codes. We distinguish roughly between orthogonal space-time block codes (OSTBCs) [9, 12], non-orthogonal space time block codes (NOSTBCs) [13, 14], and trellis codes [15]. In a first step, we are going to study the use of distributed OSTBCs with the help of our demonstration system.

The first orthogonal space-time block code was the *Alamouti code*, a transmit diversity scheme for two transmit antennas [9]. Tarokh et al. generalized the principle of this scheme in [12], where they proposed space-time block codes from orthogonal designs. These codes can be designed for any number of transmit antennas and they achieve the full transmit diversity gain. The maximum-likelihood (ML) decoding is very simple and of very low complexity. The most important disadvantage is, that OSTBCs for more than two antennas do not achieve the full rate, i.e. they lead to a rate loss. For more than four antennas the maximum rate is only 0.5 (for  $n = 3$  and  $n = 4$  transmit antennas, there are OSTBCs achieving a rate of  $R = \frac{3}{4}$ ).

In this paper, we apply a distributed transmit diversity scheme based on the Alamouti code. The goal is to implement an efficient cooperative broadcasting scheme.

A similar multi-antenna diversity demonstration system which utilizes the Alamouti scheme on USRP boards is described in [16]. The authors have successfully implemented a complete communication system using SDRs, MATLAB and GNU Radio. However, the focus of their work was on the implementation aspects using two SDRs and not on the scaling behaviour of MANETs. Hence, only a single SDR equipped with two antennas was used on the transmit side and another SDR on the receiving side.

One very important aspect to achieve transmit diversity through the use of space-time block codes (STBC) like the Alamouti scheme is knowledge about the channel impulse response (CIR) at the receiver. The channel estimation performance of a MIMO-OFDM system based on the Alamouti scheme is studied in [17]. However, no measurements on hardware have been performed.

MATLAB in combination with SDRs seems to be a popular approach to build up a demonstration system with low costs. Measel et. al. also implemented an OFDM-MIMO demonstration system utilizing the Alamouti scheme [18]. Unlike the focus of this paper, their demonstration system was implemented to characterize already existing communication systems but not to investigate cooperative broadcast techniques with respect to the scalability of MANETs.

Horváth and Bakki implemented a prototype transmission link for FBMC [19]. They also used the USRP SDR platform from Ettus Research for the transmitter and receiver. Moreover, they performed some measurements regarding the power spectral density, the peak-to-average power ratio, and the bit error rate (BER) to validate their system design. In contrast to this paper, they focused on

a SISO system and not on a MISO system. Their main aim was the realization of an FBMC transmission testbed based on one link.

Dziri et al. implemented a comparable real-time FBMC transmission link [20]. In addition to [19], they also implemented a channel, hence built up a more sophisticated transmission model. Comparable to the demonstration system presented in this paper, they used MATLAB to develop the software and the USRP SDR platform from Ettus Research. But in contrast, they stuck to a SISO model and did not further consider any diversity scheme. Their focus was to solve some practical problems, mainly time and frequency synchronization as well as channel estimation and equalization.

### 2.3 Space-Time Coding

Space-time coding means to code data across both space and time to achieve transmit diversity. A simple scheme is to use a repetition scheme with two time slots. In time slot 1, only transmit antenna 1 is transmitting symbol  $\alpha$ , in time slot 2, the same symbol  $\alpha$  will be transmitted by antenna 2. This can be described in matrix notation:

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}. \end{aligned} \quad (2)$$

Note, that due to the time-slotted approach, the data rate is only 1/2. The receiver will combine the two signals using the maximum-ratio combining approach to maximize the SNR of the received signal:

$$\begin{aligned} y &= y_1 \cdot h_1^* + y_2 \cdot h_2^* \\ &= (|h_1|^2 + |h_2|^2) \alpha + h_1^* n_1 + h_2^* n_2. \end{aligned} \quad (3)$$

### 2.4 Alamouti Scheme

To overcome the drawback of rate reduction, two symbols have to be transmitted in two time slots. This is achieved by a transmit symbol vector  $\mathbf{x}_1 = (\alpha_1, \alpha_2)^T$  in a first time slot and  $\mathbf{x}_2 = (-\alpha_2^*, \alpha_1^*)^T$  in a second. The receiver observes  $y_1 \in \mathbb{C}^{1 \times 1}$  in the first time slot, and  $y_2$  in the second; both are stacked in a vector  $(y_1, y_2)^T$ .

The resulting matrix notation for this is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2^* & \alpha_1^* \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}, \quad (4)$$

which can be rewritten as

$$\begin{pmatrix} y_1 \\ y_2^* \end{pmatrix} = \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2^* \end{pmatrix} \quad (5)$$

$$\mathbf{y} = \sqrt{\frac{E_s}{2}} \cdot \mathbf{H} \cdot \boldsymbol{\alpha} + \mathbf{n}. \quad (6)$$

Decoding will be performed by multiplying with  $\mathbf{H}^H$ :

$$\mathbf{r} = \mathbf{H}^H \cdot \mathbf{y} \quad (7)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \mathbf{H}^H \mathbf{H} \cdot \boldsymbol{\alpha} + \mathbf{H}^H \mathbf{n} \quad (8)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{pmatrix} \cdot \begin{pmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{pmatrix} \cdot \boldsymbol{\alpha} + \mathbf{H}^H \mathbf{n} \quad (9)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} |h_1|^2 + |h_2|^2 & 0 \\ 0 & |h_1|^2 + |h_2|^2 \end{pmatrix} \cdot \boldsymbol{\alpha} + \mathbf{H}^H \mathbf{n}. \quad (10)$$

Thus, two independent channels are established which gives about 4 dB improvement compared to the repetition scheme in order to achieve the same data rate [21].

Considering a frequency offset of both transmitters with respect to the local oscillator of the receiver, (4) can be rewritten as

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} x_{11} \cdot e^{j\omega_1 t} & x_{21} \cdot e^{j\omega_2 t} \\ x_{12} \cdot e^{j\omega_1(t+T_s)} & x_{22} \cdot e^{j\omega_2(t+T_s)} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \quad (11)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} \alpha_1 \cdot e^{j\omega_1 t} & \alpha_2 \cdot e^{j\omega_2 t} \\ -\alpha_2^* \cdot e^{j\omega_1(t+T_s)} & \alpha_1^* \cdot e^{j\omega_2(t+T_s)} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}, \quad (12)$$

where  $T_s$  is the slot duration. Assuming that the frequency offset is so small that the phase shift that occurs due to the frequency offset between time slot 1 and time slot 2 can be neglected, i.e.

$$e^{j\omega_1 t} \approx e^{j\omega_1(t+T_s)} \quad (13)$$

and

$$e^{j\omega_2 t} \approx e^{j\omega_2(t+T_s)}, \quad (14)$$

equation (11) can be rewritten as

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} h_1 \cdot e^{j\omega_1 t} & h_2 \cdot e^{j\omega_2 t} \\ h_2^* \cdot e^{-j\omega_2 t} & -h_1^* \cdot e^{-j\omega_1 t} \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}. \quad (15)$$

As described above, decoding can thus be performed by multiplying with  $\mathbf{H}^H$ :

$$\mathbf{r} = \mathbf{H}^H \cdot \mathbf{y} \quad (16)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \mathbf{H}^H \mathbf{H} \cdot \boldsymbol{\alpha} + \mathbf{H}^H \mathbf{n} \quad (17)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} h_1^* \cdot e^{-j\omega_1 t} & h_2 \cdot e^{j\omega_2 t} \\ h_2^* \cdot e^{-j\omega_2 t} & -h_1 \cdot e^{j\omega_1 t} \end{pmatrix} \cdot \begin{pmatrix} h_1 \cdot e^{j\omega_1 t} & h_2 \cdot e^{j\omega_2 t} \\ h_2^* \cdot e^{-j\omega_2 t} & -h_1^* \cdot e^{-j\omega_1 t} \end{pmatrix} \cdot \boldsymbol{\alpha} + \mathbf{H}^H \mathbf{n} \quad (18)$$

$$= \sqrt{\frac{E_s}{2}} \cdot \begin{pmatrix} |h_1|^2 + |h_2|^2 & 0 \\ 0 & |h_1|^2 + |h_2|^2 \end{pmatrix} \cdot \boldsymbol{\alpha} + \mathbf{H}^H \mathbf{n}. \quad (19)$$

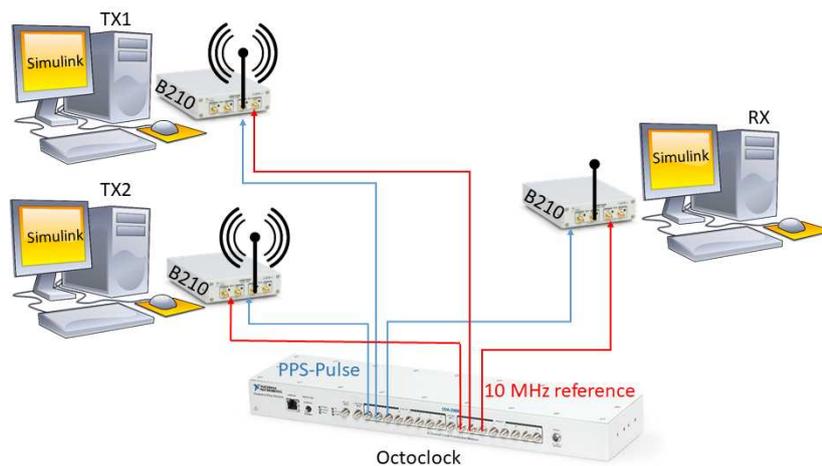
Hence, small frequency offsets between the different local oscillators can be corrected at the receiving side. Estimation of the time-variant channel impulse response will be performed in two steps:

1. estimation of the complex channel coefficient and
2. estimation of the frequency offset.

Combining both estimates can be considered as a time-variant CIR.

## 2.5 Hardware Setup

For the measurements, three different SDRs of type Ettus Research USRP B210 are used, each connected to a computer. Two SDRs are configured as transmitters, one as a receiver. For the initial measurements, the SDRs are connected to the Ettus Research 8-channel clock distribution system CDA-2990 (called OctoClock). This system generates a 10 MHz reference signal which is used to derive the carrier frequency and the sampling timing. Additionally, a one-pulse-per-second (PPS) signal is generated and distributed to the SDRs for timing alignments (cf. Figure 2). The two SDRs are configured to start transmitting at a certain instant of time. Based on the PPS-pulses, the receiver is configured to start 0.01 s prior to the transmission. This ensures that the receiver will not miss the start of a burst, also in case of jitter. Due to the connecting cables between OctoClock and SDRs, the range of the setup is limited to the lengths of the cables.



**Fig. 2.** Demo System Architecture.

## 2.6 Software Setup

**Development Environment** In order to have a user-friendly interface, a high-performance interface to the SDR, and a powerful programming language, we have chosen a combination of three different programming tools:

*Simulink:* The Simulink environment is part of MATLAB and offers a graphical user interface which allows to monitor and change signal parameters and settings during runtime of the measurement. It is therefore the preferred solution for the user interface. Moreover, it offers the possibility to integrate MATLAB and C++ code, which both will be needed, too.

*MATLAB:* MATLAB is very well suited to implement algorithms and signal processing. A lot of functions are provided by various toolboxes which simplifies the development. In addition, figures can easily be generated during runtime of the measurement which eases the development and debugging process. Therefore, all signal processing functionality is implemented using MATLAB blocks in Simulink.

*C++:* As a C++ API is provided for the SDRs, an interface between MATLAB and the SDR is programmed using C++. This allows access to sophisticated functions of the SDRs like synchronization to external pulses and reference frequencies. The C++ interface can be included in Simulink and MATLAB using MEX-functions.

**Configuration** Relevant settings of the SDRs (such as frequency offsets, gains, etc.) can be controlled from the Simulink user interface. All other settings are defined in appropriate initialization-files which are called at the software's startup. In these files carrier frequency, length of training sequences and data, interpolation and decimation factors, etc. are set.

## 2.7 Modulation

In the current implementation, different modulation schemes are implemented. The setup allows to select between a single-carrier 4-QAM modulation scheme, a multi-carrier OFDM and an FBMC scheme using 4-QAM and Offset-4-QAM on the subcarriers.

## 2.8 Transmitter Implementation

The transmitter implementation consists of the burst generation, i.e. combining training data and payload, coding according to the Alamouti scheme, and pulse shaping using rectangular pulses. The complex baseband sequence will then be passed to the SDR interface where the signal is mixed to carrier frequency and transmitted. For the demonstration system, the payload data is always the same which allows for simple BER and packet error rate (PER) calculations at the receiving side.

## 2.9 Receiver Implementation

PPS pulses are used for coarse synchronization and the start of sampling is aligned on the rising edge of the PPS pulse. This ensures that full bursts are sampled and no data will be missed. The SDR mixes data to a low intermediate frequency. Mixing to baseband and applying a matched filter is then performed in MATLAB.

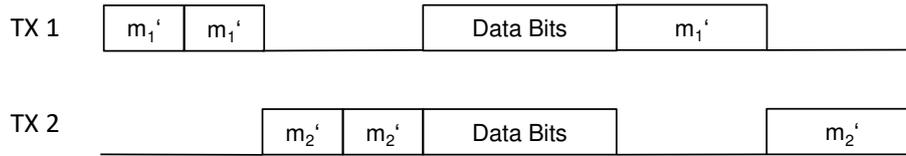
Since for the Alamouti scheme two transmitters are active and therefore CIRs and carrier frequency offsets have to be estimated for both of them, all further signal processing blocks are implemented twice. The explanations will concentrate on one chain as both chains are working similarly. The main difference is that they use different training sequences and, for the staggered burst structure, the offset values are different (i.e. the time difference between the training sequence and the start of the data block has to be adjusted).

**Parameter Estimation** The parameter estimation, such as carrier frequency offset estimation, timing estimation, and estimation of the CIR has to be done differently for single-carrier and multi-carrier modulation schemes. For single-carrier schemes, the estimation is mostly done in time domain, whereas in multi-carrier schemes, the estimation is typically done in frequency domain, resulting in a complex channel coefficient per subcarrier.

For the single-carrier system, which we will consider in the following sections, maximum length sequences (m-sequences) are used to perform the estimation of carrier frequency offset, timing, and CIR. In order to facilitate the estimation, three sequences are used, two at the beginning of the burst (used for timing and CIR estimation as well as coarse frequency offset estimation) and one at the end of the burst (used for fine frequency offset estimation). As multiple transmitters are used, orthogonality between the training sequences of the different transmitters has to be ensured. This can be done using orthogonal sequences, transmitted simultaneously, or using a time-division multiplexing approach to transmit the training sequences of the different receivers in a staggered manner (see Figure 3). Since the used m-sequences are not perfectly orthogonal, i.e. the cross-correlation between the two training sequences is not zero, there will be some interference in the simultaneous mode which degrades the estimation, especially in cases of large carrier frequency offsets.

For the multi-carrier schemes, pilot tones are often used to estimate the channel transfer function and the carrier frequency offset. The orthogonality between the pilots of the different transmitters can be achieved by using a frequency-division multiplexing approach, e.g. by assigning every  $N$ -th carrier to one transmitter and interpolating the channel transfer function. This approach works fine in case the coherence bandwidth of the channel is larger than the spacing between the pilots.

**Decoding** Once frequency offset and complex CIR are estimated, the Alamouti decoding can be performed. This scheme can also be extended to more than two



**Fig. 3.** Burst Structure: "Staggered".

nodes. The requirement for this is that an orthogonal transmit diversity scheme is applied. The Alamouti decoding block will use the estimates from both chains and output symbols according to the time-slotted Alamouti scheme where two symbols are transmitted in two adjacent time slots. To ensure that both signal chains are working properly, BER and PER will be calculated per chain and in total. Finally, demapping is performed and the transmitted and received values are compared in order to calculate BER and PER.

### 3 Measurement Results

In order to prove the aforementioned concepts, several measurements have been performed. In this section, the test scenarios are presented which shall be the base for further measurements. Thereafter, some details are provided for the measurement results for the single-carrier QAM configuration of the demonstration system. For the measurements, the SDRs are positioned in a shape of a right-angled triangle where the receiver is positioned in the corner of the right angle. The path between TX1 and RX is a grazing line of sight, as measurement equipment is positioned in between. The path from TX2 to RX can be considered as a line-of-sight channel. The distance between both transmitters and the receiver is about 1.32 m and 1.1 m, respectively.

#### 3.1 Test Scenarios

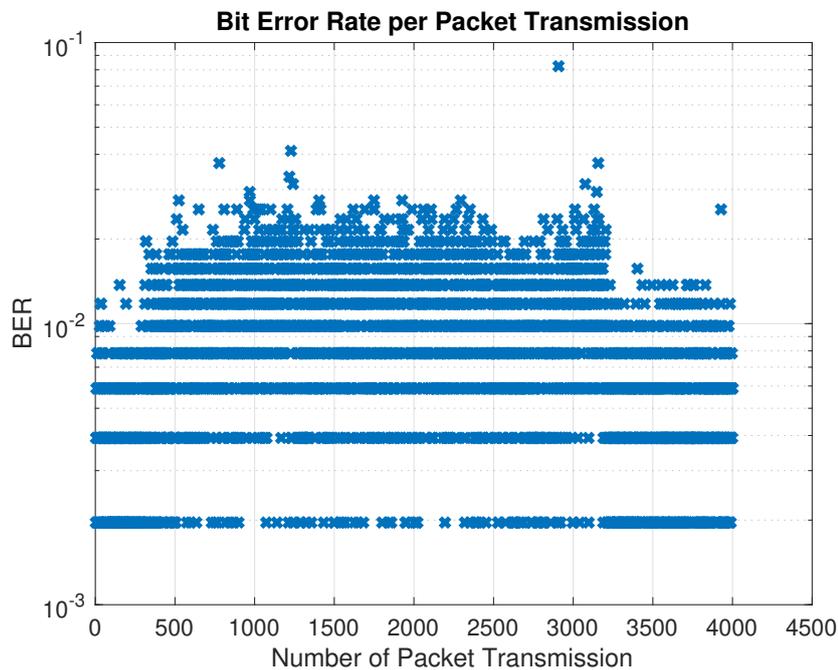
We investigate the broadcast (BRC) of MANETs in four different scenarios:

- Scenario 1 (reference, 1 transmit node and single burst decoding): One transmitter sends a packet in a short burst. The receiver decodes based on the single burst.
- Scenario 2 (BRC, SNR aggregation, one transmit node): One transmitter repeatedly sends the same burst. The receiver stores two or more observations of this burst and combines them to enhance the probability of successful decoding.
- Scenario 3 (cooperative BRC, distributed transmit diversity, single burst decoding): Two transmitters send the same broadcast data, either using the distributed Alamouti transmit diversity scheme, or each one alone.
- Scenario 4 (cooperative BRC, distributed transmit diversity, SNR aggregation): Two transmitters send the same broadcast data repeatedly, either using distributed transmit diversity or each one alone. The receiver stores several observations and combines them for successful decoding.

### 3.2 Results

In the following sections, the main results will be discussed.

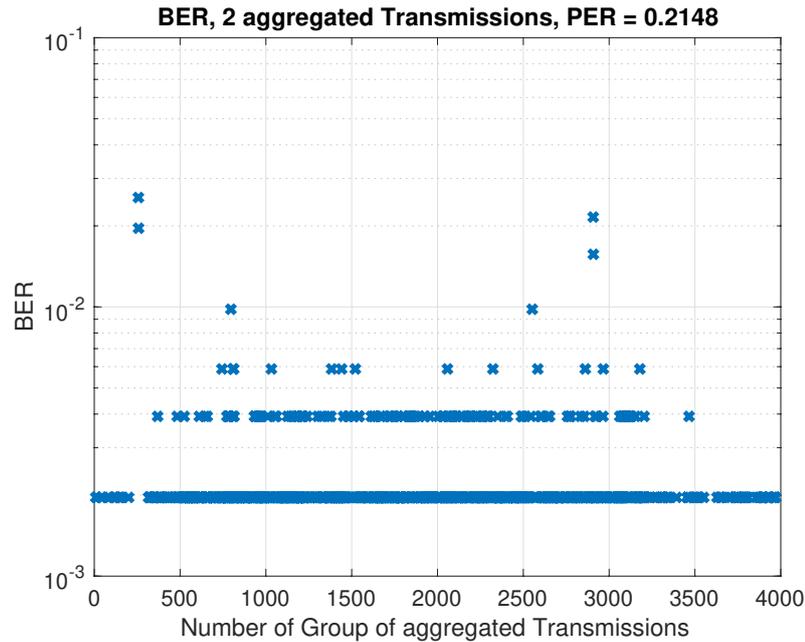
**Scenario 1** Several measurements for different SNR levels have been performed. An example of the resulting BER per packet is shown in Figure 4 for an average SNR of approximately 8.1 dB. Thereby, each one of 4001 packets is decoded separately and no channel coding is used; the channel is estimated using a training sequence. The PER, averaged over all 4001 packets, is about 96%, i.e. only 4% of the received packets are successfully decoded, i.e. without a bit error.



**Fig. 4.** BER for each of 4001 transmitted packets, Scenario 1, mean SNR  $\approx$  8.1dB.

**Scenario 2** One representative result is given in Figure 5 for the same measurements as used in Scenario 1. It shows the significant improvement compared to the experiment in Figure 4 for the SNR aggregation based on two coherently combined packets, i.e. packet 1 is combined with packet 2 before decoding, packet 2 is combined with packet 3 before decoding, then packet 3 with packet 4 and so on. For the coherent combining, the estimated CIR is used. The average PER is decreased to about 21%. Taking all results into account, it can be concluded that the SNR aggregation used for this scenario is an efficient technique to enhance the BER and PER performance in a broadcast significantly.

**Scenario 3** For Scenario 3 we consider a very low SNR case measured with the demonstration system. Node 1 and Node 2 are using the distributed Alamouti



**Fig. 5.** BER for coherently aggregated packets (consisting of two coherently combined observations), Scenario 2, mean SNR  $\approx 8.1$ dB.

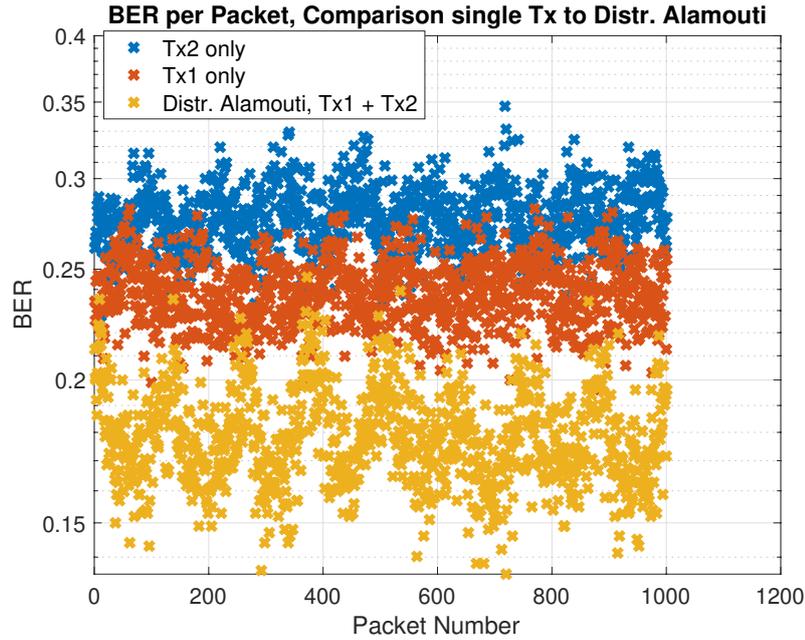
scheme, or each node transmits alone. For Node 1 transmitting, the receive SNR is about -4 dB, for Node 2 it is about -2 dB and for the distributed Alamouti transmission of both nodes it is about 0 dB.

As it can be seen in Figure 6, the BER can be significantly reduced by the use of the distributed transmit diversity. For all three cases the BERs are high and the PERs are 100% (i.e. no error free packet could be detected) because of the very low SNR. In comparison, the distributed Alamouti approach shows by far the lowest BER, due to a diversity as well as a power gain (two nodes transmit jointly and no sum power constraint is limiting the sum transmit power).

The fact, that the Alamouti scheme outperforms both SISO schemes is true as long as good estimates of the channel are available. This can be seen from the following two measurements. In a first case (Case 1) the channel estimation is performed in a low SNR-regime whereas in a second case (Case 2) the channel estimation is performed in a high(er) SNR-regime.

**Table 1.** PER for SISO and Alamouti scheme.

	PER	
	Case 1	Case 2
<b>SISO: TX1 only</b>	81.41%	79.10%
<b>SISO: TX2 only</b>	93.03%	51.74%
<b>Alamouti: TX1 and TX2</b>	94.03%	18.52%



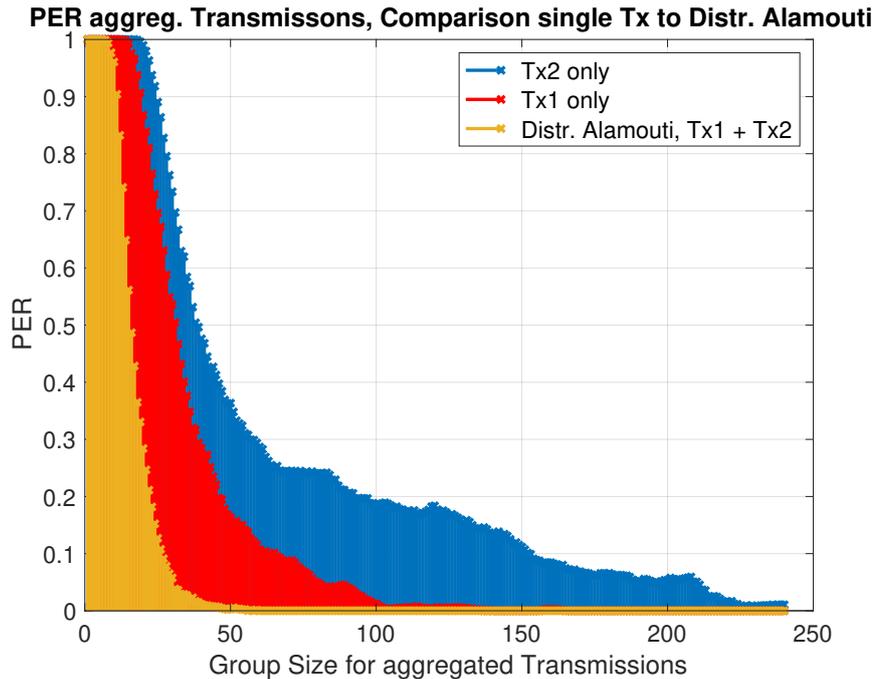
**Fig. 6.** BER for each of 1001 transmitted packets, Scenario 3, low SNR.

As it can be seen in Table 1, the Alamouti scheme outperforms both SISO schemes in case good estimates of the complex CIR can be achieved, i.e. the SNR of the training sequences is sufficiently large. This has been achieved by transmitting the training sequences with a higher transmit power than the payload of the message.

**Scenario 4** In Scenario 3, the resulting BER is very high and the PER equals 100% even for the distributed Alamouti transmission - due to the very low receive SNR. To further reduce the PER, we combine SNR aggregation with the distributed Alamouti scheme. Figure 7 shows the corresponding PER versus the group size  $N_g$  for aggregated transmissions (i.e., the number of combined packet observations). For group size  $N_g = 1$ , each packet is decoded separately, for  $N_g = 2$ , two packet observations are coherently combined, etc. The same measurements as in Scenario 3 are used, i.e. the receive SNR values are as stated there.

Although the packet error rates drop strongly in all three cases, the distributed Alamouti approach shows by far the best performance. However, due to the low SNR a high number of observations have to be combined to achieve low packet error rates. For instance, in order to achieve a PER below 10%, for the distributed Alamouti case 27 packet observations have to be combined. In comparison, in case only Node 1 is transmitting, already 63 packets, and in case of Node 2 even 153 packets have to be combined to achieve  $\text{PER} < 10\%$ .

**Summary of the Results** To conclude, the transmission can be significantly improved in terms of a BER and PER reduction by using the SNR aggregation



**Fig. 7.** PER vs. number of coherently combined packets; low SNR case.

and the Alamouti diversity scheme. The benefits of SNR aggregation become clearly visible in the first two scenarios. The benefits of the Alamouti diversity scheme become obvious in the third measurement scenario. The last scenario shows, that the SNR aggregation in combination with the Alamouti diversity scheme is even more beneficial.

### 3.3 Scalability of the Demonstration System

Currently, the demonstration system consists of three SDRs. One of the major drawbacks of the B210 SDRs is that only USB-connectivity is supported. This means that the SDR devices have to be connected to a PC, which requires a large number of PCs to set up a larger demonstration system, or much more expensive SDRs providing an Ethernet interface.

In order to overcome this issue, a Raspberry Pi, which is a single board computer, can be used. Data to be transmitted can be sent to the Raspberry Pi via Ethernet which is then forwarded to the SDR over the USB interface. Therefore, one PC can be used to control a large number of transmitters. A TCP/IP server is running on the Raspberry Pi and a TCP/IP client in MATLAB on the PC. Thus, the user interface only requires minimal change, i.e. the SDR interface has to be replaced by the TCP/IP client interface. A proof of concept has been performed and the results have shown that this is a suitable approach for further phases of the project.

## 4 Future Work

### 4.1 Scaling Behaviour of MANETs

As sketched in Section 1, one major drawback of MANETs is their scaling behaviour with increasing number of nodes and, correspondingly, routes. Currently, no sophisticated solution exists on how to overcome this issue. One of the major goals of the implemented demonstration system and its planned extensions is to investigate different methods and techniques to improve exactly this. Therefore, it is foreseen to further investigate the use of OSTBCs as well as NOSTBCs like Linear Scalable Dispersion Codes (LSDCs). Moreover, it is nearby to consider Space-Frequency Block Codes, too. In our future work we will also consider synchronization issues in order to implement the system without the OctoClock synchronizer.

### 4.2 Adaptability of the Demonstration System

The implemented Simulink environment has been designed in a modular way for simple exchange of modulation schemes, parameter estimation algorithms, and cooperative diversity schemes. With this model it has been shown that exchanging the modulation with only slight adaptations is possible. The distributed MISO scheme still allows for a diversity gain.

The simulation environments are implemented with a user interface in such a way that simulations can be run on a PC and the same user interface can be used to perform real transmissions using SDRs.

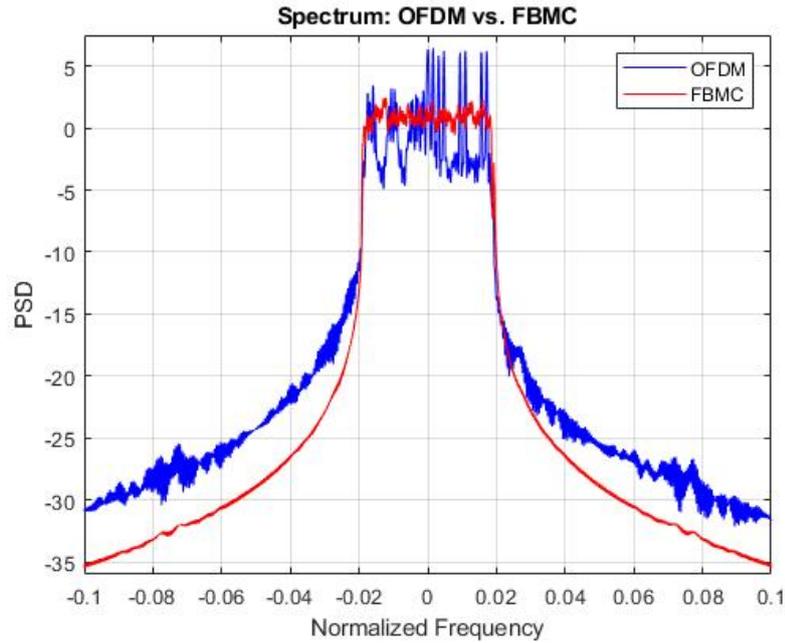
### 4.3 Further Measurements

All measurements shown above have been performed for the single-carrier scheme. The measurements shall be repeated for OFDM and FBMC modulation schemes. We expect that with OFDM and FBMC an efficient use of larger bandwidth and higher data rates can be achieved. A comparison between the spectra of the generated signals shows that the edges for the FBMC spectrum are much steeper than for the OFDM spectrum which corresponds to the theoretical expectations proving that the signal processing in our implementation is working correctly.

## 5 Conclusion

This paper describes how to set up and implement a MANET demonstration system which allows to perform measurements and to gain hands-on experience on cooperative transmit diversity schemes in combination with further signal-processing like the SNR aggregation approach for broadcasting.

Furthermore, the scalability and adaptability of this demonstration system have been proven. For the latter, Simulink environments have been implemented utilizing single-carrier as well as multi-carrier modulation schemes (OFDM and



**Fig. 8.** Comparison between the OFDM and FBMC Spectrum.

FBMC). These environments can be directly deployed to the SDRs in order to perform real-time measurements.

First measurements using the demonstration system in its single-carrier configuration have been performed investigating the benefits of the described burst structure and SNR aggregation. One crucial result is that the Alamouti diversity schemes significantly outperform SISO schemes as long as the estimation of the complex CIR is possible with the necessary quality. Additionally, SNR aggregation can be used to further reduce the PER. Again, the Alamouti diversity scheme shows by far the best results.

## References

1. C. Siva Ram Murthy and B. S. Manoj, *Ad hoc Wireless Networks: Architectures and Protocols*, Prentice Hall PTR, May 2004.
2. J. Loo, J. Mauri, and J. Ortiz, *Mobile Ad Hoc Networks*, Auerbach Publications, 2012.
3. P. Gupta and P. R. Kumar, "The capacity of wireless networks," in *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, March 2000.
4. C. Zhang, X. Zhu and Y. Fang, "On the improvement of scaling laws for large-scale MANETs with network coding," in *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 662-672, June 2009.
5. M. F. Carey, J. M. Chapin and V. W. S. Chan, "Connectivity and Scaling behavior of power-limited directional infrastructureless wireless networks," 2011 - MILCOM 2011 Military Communications Conf., Baltimore, MD, 2011, pp. 985-990.
6. A. Ozgur, O. Leveque and D. N. C. Tse, "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks," in *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549-3572, Oct. 2007.

7. B. Sirkeci-Mergen, A. Scaglione, and G. Mergen, "Asymptotic Analysis of Multi-Stage Cooperative Broadcast in wireless networks," in *IEEE Trans. Inf. Theory*, Vol. 52, No. 6, pp. 2531–2550, June 2006.
8. T. R. Halford, and K.M. Chugg, "Barrage Relay Networks," in 2010 Information Theory and Applications Workshop (ITA), April 2010.
9. Alamouti, S. M., "A Simple Transmit Diversity Technique for Wireless Communications," in *IEEE J.Sel. A. Commun.* 16, Nr. 8, S. 1451-1458, September 2006.
10. A. Wittneben, "Basestation modulation diversity for digital simulcast," [1991 Proc.] 41st IEEE Vehicular Technology Conf., St. Louis, MO, USA, 1991.
11. D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone and W. Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," in *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380-1408, December 2010.
12. V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-time block codes from orthogonal designs," in *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456-1467, July 1999.
13. A. Wittneben and M. Kuhn, "A new concatenated linear high rate space-time block code," *Vehicular Technology Conf.. IEEE 55th Vehicular Technology Conf.. VTC Spring 2002 (Cat. No.02CH37367)*, Birmingham, AL, USA, 2002.
14. B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," in *IEEE Transactions on Information Theory*, vol. 48, no. 7, July 2002.
15. V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," in *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744-765, March 1998.
16. M. Ullah, H. Khan, U. Sharif and M. D. Nisar, "Multi-antenna diversity demonstration via Alamouti scheme on USRP boards," 2017 Int. Symp. on Wireless Systems and Networks (ISWSN), Lahore, 2017, pp. 1-5.
17. S. Manzoor, A. S. Bamuhaisoon and A. N. Alifa, "Channel estimation for MIMO-OFDM systems," 2015 5th Nat. Symp. on Information Technology: Towards New Smart World (NSITNSW), Riyadh, 2015, pp. 1-7.
18. R. Measel et al., "A MATLAB platform for characterizing MIMO-OFDM communications with software-defined radios," 2014 IEEE Int. Workshop Technical Committee on Communications Quality and Reliability (CQR), Tucson, AZ, 2014.
19. B. Horváth and P. Bakki, "Implementation of FBMC transmission link using SDR," 2013 23rd Int. Conf. Radioelektronika (RADIOELEKTRONIKA), Pardubice, 2013, pp. 320-323.
20. A. Dziri, C. Alexandre, R. Zakaria and D. Le Ruyet, "SDR based prototype for filter bank based multi-carrier transmission," 2014 11th Int. Symp. on Wireless Communications Systems (ISWCS), Barcelona, 2014, pp. 878-882.
21. David Tse and Pramod Viswanath, *Fundamentals of Wireless Communication*. 5th printing. Cambridge: Cambridge University Press, 2005.

# A new metric for assessing the performance of 2D Lidar SLAMs

Bingxin Zi, Haiying Wang, Jose Santos and Huiru Zheng

School of Computing, Ulster University, Belfast, BT37 0QB, UK  
hy.wang@ulster.ac.uk

**Abstract.** Simultaneous Localisation and Mapping (SLAM) is a widely studied topic in recent years and has a wide potential in the field of unmanned driving and robotics. Over the past decade, a number of SLAM algorithms have been developed, each exhibiting unique performance in their applications. This paper presents a comparative study of the performance of three well-known Light Detection and Ranging (LiDAR-based SLAM algorithms, i.e. Gmapping, Cartographer and Hector, with an emphasis on the 2D maps constructed by each algorithm. In order to deal with incomplete maps constructed, a new evaluation metric was introduced. To reduce the human error during scene construction and equipment calibration, all experiments were carried out in the 2D simulation available within the Robot Operating system (ROS). Three well-designed maps with different sizes and complexities were introduced to investigate the features of three SLAM algorithms. Besides, to reduce the impact of randomness, each dataset was assessed 10 times to obtain the mean value and the standard deviation. The results show that, in comparison to traditional metrics such as a metric of average distance to nearest neighbour (ADNN), the proposed new metric can clearly reflect both the quality and completeness of maps built by SLAM algorithms.

**Keywords:** SLAM, assessment metric, 2D LiDAR.

## 1 Introduction

Simultaneous Localisation and Mapping (SLAM) is one of the most widely studied topics across multiple subjects including robotics, computer vision and machine learning. In the field of robotics, it gives robots the ability to explore an unknown environment by constructing a map [1] and understand the environment by processing the information from visual sensors [2]. SLAM has been applied in unmanned vehicles, autonomous driving, augmented reality (AR) among other [2]. According to different usage environment, SLAM may apply different algorithms and sensors. With the development of technology, machine learning techniques can also be used to help SLAM process the sensor information [3][4].

SLAM techniques can be divided into two main types, filter-based and graph-based. The filter-based SLAM has become the mainstream over the past decades to

achieve localisation and mapping. The algorithm examples used in SLAM applications include MonoSLAM [5] and Gmapping [6]. MonoSLAM uses an Extended Kalman Filter (EKF) framework to realise the robot's status estimation and mapping. The EKF is a nonlinear extension of the Kalman filter which contains a status equation and an observation equation. The status equation is used for predicting the robot's current status based on the status at the last timestamp, while the observation equation is used for correcting the prediction by taking the sensor observation into account. The Gmapping algorithm applies a particle filter, in which each particle represents a potential trajectory of the robot, to handle the localisation and mapping. Each particle is associated with a weight and stands for a possible status of the robot. The final estimation is determined by the weighted mean of all particles. In recent years, graph-based SLAM has received much attention [7-10]. In this system, there are two main processes: frontend and backend. The frontend processes sensor data and calculates the dynamics of the robot while the backend receives the dynamics and generates a fusion result. At the same time, the backend is also responsible for the optimisation of the whole system, which is a procedure that the filter-based SLAM does not have. The Cartographer algorithm [10] is the classic implementation of graph-based SLAM approaches.

Each of these algorithms has its advantages and limitations. The main problem arising from their use in robotics is how to evaluate their performance. Attentions have been traditionally focused on the assessment of the accuracy of the generated maps and trajectories only. Examples include the use of a metric like average distance to the nearest neighbour (ADNN)[11-14], or evaluating the trajectory in a probabilistic approach[15]. At the same time, a ground truth independent evaluation was introduced in [16], which counts the features of estimated map to independently assess the quality of map.

This study proposes a new metric which would evaluate the SLAM's map results in terms of both the accuracy and the structural completion of the maps derived. Three well-known SLAM systems (Gmapping, Cartographer and Hector) were studied with an emphasis on the analysis of the 2D maps constructed by each algorithm.

The remainder of this paper is structured as follows: the related work is presented in Section 2, followed by a description of SLAM algorithms in Section 3. In Section 4, the experiment environment is introduced. Evaluation metrics used in the study are described in Section 5 and experiment results are analysed in Section 6. The paper concludes with a summary of conclusion and future work discussion.

## 2 Related works

It could be challenging to analyse and evaluate the performance of various SLAM systems because they have different sensor sources, algorithm basements and code frames. For example, Gmapping utilises light detecting and ranging (LiDAR) and an

odometer to calculate the journey of the robot and to construct the map. To apply SLAM to aircraft navigation, instead of using an odometer, Hector [17] combines 2D LiDAR with an inertial measurement unit (IMU). The Cartographer algorithm released in 2016 by Google [10] combines data from various sensors e.g. LiDAR, IMU, and odometer. In order to provide a uniform platform for implementing the multiple types of robot algorithms, the Robot Operating System (ROS) framework [18] was introduced in 2007. It defines a uniform standard of interactions among different robot subsystems. Under the uniformed framework, the different SLAM systems employ the common data inputs, data flows and outputs standard. This makes it possible to compare the performance of different SLAM algorithms under the same pattern. This paper focuses on evaluating the performance of the LiDAR-based 2D SLAMs including Hector, Gmapping and Cartographer implemented in ROS.

Normally, the evaluation of SLAM results performance can be carried on two aspects: the trajectory and the map quality. The trajectory is often evaluated in the vision SLAM (VSLAM) because the map exists in multiple forms (sparse map, dense map, semi-dense map). In a VSLAM system, it is hard to directly compare the quality of a map derived from different systems, for example, Buyval et al. [19] compared the ability of 4 different VSLAM systems to obtain features and the coverage of point clouds but did not perform any quantitative analysis. On the aspect of map comparison, Xu et al.[20] projected the feature points detected by the camera onto the ground to fake the LiDAR laser scans. Then they generated the occupied grid map by using the pseudo laser scans. Therefore, they were able to compare different type of VSLAM maps by projecting the points onto ground. However, the noises points in the space may also be projected onto the ground, thereby forming errors. Some extra procedures are required to provide ground truth data. Yagfarov et al.[11] introduced the high-precision laser tracker FARO to manually construct a ground truth map. They used FARO to get the 3D scans of the experiment room. Then they extracted the intersection lines between the floor plane and the vertical plane. Those lines were viewed as the 2D projections of the experiment room. Due to the lack of wall width information, they used some OpenCV functions to thin the SLAM estimated map's wall width to one grid. However, lacking wall width information may introduce errors. Sturm et al.[12] introduced an extra high-frequency motion capture system to provide ground truth. Filipenko and Afanasyev[13] used a much simpler way to get the ground-truth. They laid some threads on the floor. The robot was manually driven to follow the treads to obtain the estimations. However, there is no guarantee that the actual robot trajectory will perfectly fit these threads, especially in the turning areas. So they only used the straight sections of the threads for comparison. Since they found the Hector's result was closest to the ground truth, they used Hector's result as reference for the other systems. It should be mentioned that they did not get a valid Gmapping result. Santos et al. [14]implemented the evaluation in both the simulation environment STAGE and the real physical world. They evaluated Hector, Gmapping, KartoSLAM, CoreSLAM and the LagoSLAM. Their simulation experiments show that "Gmapping algorithm presents exceptional results" while the "KartoSLAM was the best performing technique in real word". The Cartographer algorithm was not

included in their study. Bayer et al.[21] pointed out that the SLAM ground truth is hard to construct. Anton et al. [16] pointed out ground truth data are not always fetchable even for a lot of open datasets. Instead of comparing with ground truth, they proposed some novel metrics: the proportion of occupied and free cell, the number of corners and the number of enclosed areas for independently evaluating the map quality without acquiring the ground truth. Besides, Le et al.[22] introduced a structural metric of Structure Similarity Index (SSIM) [23] to evaluate the map quality of different SLAM systems in the indoor environment. But the SSIM result cannot provide an intuitive feeling of the structure completion.

Since the accuracy of a trajectory relies on the accuracy of maps, this study will focus on the comparison of map quality. Moreover, because it is hard to obtain ground truth information without introducing some extra interferences, such as the calibration of a reference equipment and setting artificial markers, all the experiments in this study have been conducted in 2D simulation available in ROS.

### **3 2D LiDAR SLAM Algorithms**

In this research, three SLAM algorithms: Gmapping, Cartographer and Hector were investigated. They utilise different algorithm frames and different sensors, as summarised below:

#### **3.1 Gmapping**

Gmapping [6] is one of the most typical filter-based SLAM algorithms. It is based on the Rao-Blackwellized Particle Filter (RBPF). In RBPF, the problem of estimating a robot pose and map is divided into two steps. In detail, RBPF takes the robot pose estimating problem as an incremental estimation problem. Based on the robot pose on the last frame and the robot dynamics on the current frame, the current robot pose can be predicted. Then the mapping can be solved by providing the current pose and observations. In naïve particle filter system [24], each particle stands for one possible status of the robot and one possible representation of the map. All particles contribute to a weighted mean estimation. To reach a closer representation of the true possibility distribution, the particle filter requires a large number of particles to expand the sampling range. But after several propagation iterations, the particles' weights concentrate on a few specific particles, which makes the other majority of particles barely contribute to the mean result. In Gmapping, a novel way was used to establish a more accurate proposal distribution by taking the laser observation into account. Furthermore, an adaptive metric was applied to guarantee that the resampling process was executed only when the weights' concentration rate was higher than a threshold. These measures allow the Gmapping algorithm to use fewer particles than the naïve system which in turn makes it works in real-time.

### 3.2 Hector

Hector [17] is a very straightforward algorithm that estimates the robot status by aligning the laser scan and map. Every new scan is transformed into the discrete occupied grid cell by applying the Bresenham algorithm [25]. After that, the approximate optimal transformation between the current scan and existing map is solved by applying a Gaussian-Newton optimization. Hector accumulates each scan to form a map and to avoid falling into local minima, a multiple-resolution map principle is applied. Besides, Hector is designed to be applied in 3D spaces, so an IMU is equipped instead of the odometer.

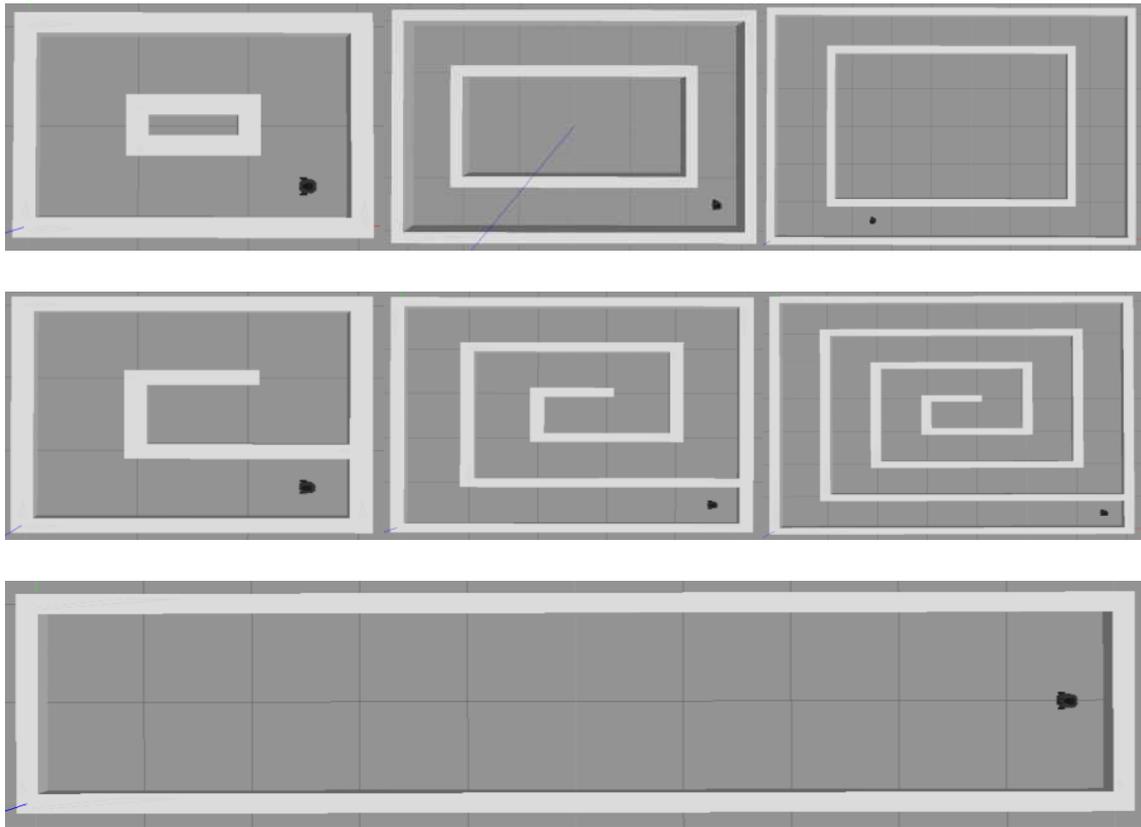
### 3.3 Cartographer

Cartographer [10] is one of the most cited open-source graph-based SLAM algorithms in recent years. The main contribution of the Cartographer algorithm is the realization of the real-time loop-closure in a LiDAR SLAM system. It proposes the principle of submap that consists of recent scans. Once a new scan is acquired, the system will look up recent sub-maps to find a reasonable matching between the scan and a sub-map. The scan will be inserted in the submap without considering the drift during that short time. Presently, a loop closure detection will be carried out along all submaps to minimise the long-term drift. Besides, Cartographer introduces 3 types of the branch and bound algorithms to speed up the loop closure searching process.

## 4 Experiment environment set up

The experiments were carried out on the Ubuntu 16.04 distribution with ROS kinetic. The ROS platform provides related ROS packages to apply all the three algorithms mentioned above. In ROS, data transmission is achieved by subscribing to and publishing to certain topics. The topic defines the data structure of interactions. All the three algorithms are edited to subscribe to the same topic of “/laser\_scan” to receive the laser scans. Besides, all three algorithms subscribe to the same topic of “/tf”. The “/tf” is an essential ROS topic which provides the transformations among each coordinate reference of the whole system. It should also be pointed out that the Gmapping subscribes the “/tf” topic to get the odometer information. While the Hector and Cartographer subscribe to the “/imu” topic to utilize IMU information instead.

Different from Santos et al.’s method [14], the Gazebo\_ROS software, instead of Stage\_ROS software, was used for creating virtual environments because the Stage\_ROS software does not provide the IMU simulation [26]. The algorithms were tested in three types of maps – a rectangle map, a swirl map and a corridor map. For the former two types, each type has three sizes respectively: 3m x 2m, 6m x 4m, 9m x 6m and 3m x 3m, 5m x 5m, 7m x 7m (in meter). The size of the corridor map is 2m x 10 m. Fig. 1 shows the image of the virtual environment and the initial position of the robot.



**Fig. 1.** The 7 simulation environments used in the study

The comparison analysis was carried out using Matlab and for mathematical convenience, each map was transformed to a dot map. The dot map keeps the same resolution – 5cm/pixel as the SLAM algorithms’ setting. Therefore, in our experiment, an error of one pixel is equal to an error of 5 cm.

The robot used for simulation is the Turtlebot3 Burger [27] which has the maximum linear speed of 0.22m/s, and the maximum rotation speed of 2.84 rad/s. In the simulation, the robot’s forward speed is gradually increased to the peak, but no rotating speed exceeds 0.7 rad/s. The ROS bag command is used for recording all topics’ data during the walking. Each algorithm processes these records separately. The map parameters are 0.65 for occupied, 0.2 for free.

It should be noted that, due to the stochastic behaviour of the algorithms, the output from Gmapping and Cartographer could be different even with the same inputs. Therefore, unlike other studies [11][13][14], to mitigate the randomness impact, each algorithm was tested with the same dataset 10 times in this study. The standard deviation of each metric was then calculated along with their mean values.

## 5 Metric for evaluating

In this paper, two metrics were applied to evaluating the SLAM map quality. The ADNN metric focus on the accuracy of map while the grid portion metric we proposed assesses both the completion and quality of the estimated map.

### 5.1 Average distance to the nearest neighbour (ADNN)

ADNN is estimated based on the Iterative Close Point (ICP) algorithm which was used to align a SLAM map with the ground truth [11][14] and the K-Nearest Neighbours (KNN) search used for finding the correspondence between the ground truth and the estimated map. The ICP and the KNN iteratively run until the result converges. The metric is computed as the average distance from each SLAM map point to the nearest ground truth map point as defined below:

$$ADNN = \frac{1}{n} \sum_{i=1}^n (P_{est}^i - P_{grd}^i) \quad (1)$$

Where  $n$  is the total point number of estimated grids,  $P_{est}^i$  is the position of the  $i^{th}$  estimated grid and  $P_{grd}^i$  is the position of the  $i^{th}$  ground truth grid. To avoid falling into local minima, we manually selected the 4 corner points and their average coordinates as 5 initial parameters for the ICP algorithm.

However, the ADNN metric does not perform well when dealing with incomplete estimated maps. For example, if a SLAM algorithm outputs an incomplete map but every single grid of the map is perfectly located, the ADNN is set to zero which may provide a misleading assessment. Therefore, a new metric is proposed by taking into account the completion of estimated map.

### 5.2 Grid Portion Metric

In Anton et al. [16] there are three metrics, i.e. “the portion of occupied and free cell”, “the number of corners” and “the number enclosed”, were proposed to evaluate a SLAM map from the perspective of the grid. Their method counts the number of grids/free cells/corners/enclosed areas which, intuitively, is associated to the accuracy of estimated map. For example, the number of corners is negatively related to the accuracy of estimated map since a good map should be smooth. But this association is not conducive to a quantitative evaluation on the map since a larger map with high-precision may contain more corners than a smaller map with less-precision. In order to address this we proposed a new metric to evaluate the map quality in terms of grid portion. We believe that the completeness of an estimated map can be reflected by the proportion of occupied grids. Specifically, for each ground truth map, the number of occupied grids is constant. For any occupied grids, there are three possible results when aligning a SLAM map with the ground truth: a ground truth occupied grid was correctly identified as occupied in the SLAM map (True Positive (TP)); a ground truth occupied grid was wrongly regarded as free in the SLAM map (False Negative

(FN)); a ground truth free grid was wrongly regarded as occupied in the SLAM map (False Positive (FP)). With this definition, the map quality can be evaluated in terms of sensitivity and precisions as defined below:

$$\text{Precision} = \frac{TP}{(TP+FP)}, \text{Sensitivity} = \frac{TP}{(TP+FN)} \quad (2)$$

The range of these values goes from 0 to 1. The precision value describes the quality of the estimated map. If every occupied grid in the estimated map is correct, the precision value is set to 1. While 0 means every occupied grid in the estimated is wrong. Besides, this metric can also indicate the completeness of the map, where sensitivity value 1 means the estimated map's structural matches the ground truth map in 100%. And 0 means their structures are totally different. At the same time, this metric can provide an intuitive feeling of how correct the estimated map is by drawing each TP/FP/FN grids.

The metric of precision and sensitivity were widely applied to evaluating the performance in other study areas such as supervised machine learning algorithms. However, for the best of our knowledge, we are the first to utilise these metrics to evaluate the completion and the quality of the estimated map of SLAM.

## 6 Experiment Results and Discussion

We first used ADNN to assess the quality of maps constructed by the three algorithms as depicted in Table 1. As expected, the deviations between estimated maps and ground truth represented by ADNN tend to increase as the map's size increases for all three SLAM algorithms. As shown in Table 1, 5/7 of Cartographer's mean errors are the lowest among the three algorithms. Especially when applied to a large map, Cartographer provides fairly low errors of 0.57 and 1.02 in comparison to the errors of using the Hector and Gmapping algorithms. Hector can produce competitive results in two medium scenarios. It also has low errors which are close to Cartographer's result in small scenarios. However, the performance of the Hector algorithm deteriorates significantly when it is applied to large area scenarios. This is because Hector does not utilise the odometer which can provide precise linear dynamics. The error would accumulate inevitably over time. Unexpectedly, Gmapping always has larger mean errors than other two algorithms. It only achieves a better result than the Hector algorithm when applied to large scenarios of Rectangle\_L, Swirl\_L and Corridor.

**Table 1.** The ADNN mean and standard deviation. the unit is pixel/grid

		Rectangle_S	Rectangle_M	Rectangle_L	Swirl_S	Swirl_M	Swirl_L	Corridor
Gmapping	mean	0.44	0.77	3.82	1.28	1.83	2.81	1.57
	std	0.11	0.25	0.94	0.57	0.45	0.54	0.17
Cartographer	mean	0.07	0.28	0.57	0.11	0.50	1.02	0.83
	std	0.03	0.20	0.15	0.01	0.07	0.05	0.12
Hector	mean	0.14	0.16	8.91	0.13	0.16	5.16	2.09
	std	0.00	0.00	0.00	0.00	0.00	0.00	0.00



(a) Gmapping map



(b) Cartographer map



(c) Hector map

**Fig. 2.** Visualisation of SLAM results on the corridor scenario based on the proposed metric. Green grids represent TP estimation, red grids denote FP estimation and yellow grids represent FN estimation.

The similar observation of “Hector’s performance declines in large scenario” can be found when comparing the results obtained with those presented in [14], where Hector and Gmapping’s error were 0.46 and 0.42 (0.01meter/pixel) respectively, in a about 4m\*4m map. But Hector’s error increased to 7.46 in a 12m\*11m map while Gmapping’s error only increased to 5.37. In the study [11], Cartographer is also reported having the lowest error. The error rate of the Gmapping algorithm reached the 2nd place. And Hector was the worst one. Their experiment scenario size was unclear. They tested four motion patterns (fast/slow move, sharp/smooth rotation). Hector’s errors go from 20 cm to 50 cm, while all Cartographer’s errors and 3/4 Gmapping’s errors were less than 10 cm. But it should be noted that in [13], a totally different result was obtained. In that study, Hector produced the trajectory that closest to the ground truth. Then, they used Hector trajectory as reference but did not provide the error of it. Their Cartographer result has an error of 2.4 cm compared with Hector trajectory. Besides, they did not acquire a valid Gmapping map.

In terms of the standard deviation, Hector achieves the lowest standard deviation. When applying hector multiple times, the same map was obtained. This is due to the nature of the Hector algorithm, which relies on scan matching. It stitches together the

**Table 2.** The mean and standard deviation of grid portion metric

			Rectangle_S	Rectangle_M	Rectagnle_L	Swirl_S	Swirl_M	Swirl_L	Corridor
Gmapping	Precision	Mean	0.67	0.44	0.09	0.39	0.20	0.13	0.61
		std	0.09	0.07	0.08	0.15	0.07	0.04	0.04
	Sensitivity	Mean	0.89	0.61	0.12	0.46	0.22	0.12	0.50
		std	0.09	0.07	0.08	0.15	0.07	0.04	0.04
Cartographer	Precision	Mean	0.96	0.78	0.52	0.92	0.63	0.44	0.92
		std	0.02	0.17	0.15	0.01	0.06	0.04	0.00
	Sensitivity	Mean	0.99	0.79	0.61	0.98	0.69	0.46	0.90
		std	0.02	0.17	0.15	0.01	0.06	0.04	0.00
Hector	Precision	Mean	0.88	0.86	0.23	0.89	0.86	0.17	0.72
		std	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sensitivity	Mean	1.00	1.00	0.34	1.00	1.00	0.19	0.62
		std	0.00	0.00	0.00	0.00	0.00	0.00	0.00

aligned frames between consecutive scans. It uses Gaussian-Newton optimization[17] which will always give the constant result as long as the input is the same. Gmapping, on the other hand, produces the largest variance in most of the cases, which may be attributed to the randomness feature of the particle filter employed by the algorithm.

Evaluation based on ADNN does not well describe the structure completion of a map. For example, when applied to the corridor scenario. The ADNN errors of Gmapping, Cartographer and Hector are 1.57, 0.83 and 2.09 respectively. These numbers only tell that the Gmapping and the Hector have less accuracy than Cartographer but it does not reflect the fact that the estimated map's size is only half of the ground truth's size as illustrated in Fig. 2, which was based on the visualisation of the proposed grid portion metric. Together with precision and sensitivity estimation shown in Table 2, the assessment based on grid portion has a great potential to provide a complete picture of the quality of a SLAM map. On the corridor scenario, the mean precision, mean sensitivity of the Gmapping are 0.61, 0.50. The 0.5 sensitivity tells that the Gmapping map only represents half of the ground truth map. And the 0.61 precision tells that about 61% of the estimated grids are correct. These data on the Hector map are 0.72, 0.62 respectively, as the Hector map is slightly longer than the Gmapping map and it has less FP grids. Cartographer reaches 0.92 and 0.90, when using this metric highlighted the Cartographer map is the most complete one among those three maps (Fig. 2). It is clear that the precision value intuitively reflects the correction of the map and the sensitivity value reflects the completion of the structure. Therefore, the grid portion metric along with the visualisation as illustrated in Fig. 2 is more effective in representing the quality of a map. In this view, as the Table 2. Shows, the Hector provides second high precision in the small scenarios i.e Rectangle\_S, Swirl\_S and highest precision in Rectangle\_M and Swirl\_M. But its precision and sensitivity dramatically decrease in the large scenarios. Similar to the conclusion derived from ADNN, the Cartographer can always have relatively high precision and sensitivity in all scenarios though it has a larger variance when applied to the medium and large scenarios, which deserves further investigation. For Gmapping, the variance of new metric no longer changes as drastically as the variance of ADNN. This might mean that the new grid portion metric is more universal than ADNN.

The last column of Table 2. shows that Cartographer (with highest precision and sensitivity) is robust to the geometry of corridor environment. Both Gmapping and Hector do not perform well in the corridor environment which has fewer features. For Gmapping, the proposed particles are distributed along the direction of the corridor. However, the observations obtained in the direction of the corridor are still highly similar, which makes it difficult to estimate the length of a trajectory. Hector uses the Gauss-Newton method to solve scan matching between two frames. However, similar structural information obtained in a featureless environment clearly has a big impact on its performance.

Overall, Hector SLAM algorithm is the 2<sup>nd</sup> best on small scenarios and the best in medium scenarios with both ADNN metric and grid portion metric. However, its performance will significantly drop in larger scenarios. While Cartographer reached the 1<sup>st</sup> place in small scenarios and the 2<sup>nd</sup> place in medium scenarios with both ADNN metric and grid portion metric. The important thing is, it has the lowest ADNN error and highest precision and sensitivity in every large scenario. Finally, Gmapping is worse than the other two algorithms in small and medium scenes and reaches second place in large scenarios also it has the greatest instability in most cases.

## 7 Conclusion and future work

In this work, a new grid portion metric was proposed and introduced for the assessment of the performance of SLAM algorithms. Three representative 2D LiDAR SLAMs were studied. Results show that, in comparison to traditional metrics such as ADNN, the proposed grid portion-based assessment has great potential to provide a more complete picture of the quality of SLAM maps. By visualising the SLAM results on a map, a better understanding of the map can be achieved. The numerical value of sensitivity and precision representing the proportion of FP and FN results respectively can be used to indicate the correctness and completeness of a SLAM map. In order to deal with the potential instability of the SLAM results, the mean and standard deviation of each metric were calculated which provides additional insight to the nature of each SLAM algorithm.

It is worth noting that the evaluation was based on simulation environment provided by ROS. Assessment to be carried out in a real physical world would be part of our future work. In addition, the research would be expanded to cover 3D LiDAR SLAM and vision SLAM. Besides, because each SLAM algorithm, under different scenarios, can be optimised by adjusting the certain parameters. Dynamic determination of optimal parameters represents another direction of our research.

## Acknowledgement

This work is supported by the VCRS Research Student Award from Ulster University.

## References

1. Yuan, W., Li, Z., Su, C.-Y.: RGB-D sensor-based visual SLAM for localization and navigation of indoor mobile robot. 2016 International Conference on Advanced Robotics and Mechatronics (ICARM). pp. 82–87 (2016).
2. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*. 32, 1309–1332 (2016).
3. Zhang, F., Guan, C., Fang, J., Bai, S., Yang, R., Torr, P., Prisacariu, V.: Instance segmentation of lidar point clouds. *ICRA*, Cited by. 4, (2020).
4. Wu, B., Wan, A., Yue, X., Keutzer, K.: Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. 2018 IEEE International Conference on Robotics and Automation (ICRA). pp. 1887–1893 (2018).
5. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 1052–1067 (2007).
6. Grisetti, G., Stachniss, C., Burgard, W.: Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. *Proceedings of the 2005 IEEE international conference on robotics and automation*. pp. 2432–2437 (2005).
7. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*. 31, 1147–1163 (2015).
8. Mur-Artal, R., Tardós, J.D.: Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*. 33, 1255–1262 (2017).
9. Konolige, K., Grisetti, G., Kümmerle, R., Burgard, W., Limketkai, B., Vincent, R.: Efficient sparse pose adjustment for 2D mapping. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 22–29 (2010).
10. Hess, W., Kohler, D., Rapp, H., Andor, D.: Real-time loop closure in 2D LIDAR SLAM. 2016 IEEE International Conference on Robotics and Automation (ICRA). pp. 1271–1278 (2016).
11. Yagfarov, R., Ivanou, M., Afanasyev, I.: Map comparison of lidar-based 2d slam algorithms using precise ground truth. 2018 15th International Conference on Control, Automa-

tion, Robotics and Vision (ICARCV). pp. 1979–1983 (2018).

12. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 573–580 (2012).

13. Filipenko, M., Afanasyev, I.: Comparison of various slam systems for mobile robot in an indoor environment. 2018 International Conference on Intelligent Systems (IS). pp. 400–407 (2018).

14. Santos, J.M., Portugal, D., Rocha, R.P.: An evaluation of 2D SLAM techniques available in robot operating system. 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). pp. 1–6 (2013).

15. Zhang, Z., Scaramuzza, D.: Rethinking Trajectory Evaluation for SLAM: a Probabilistic, Continuous-Time Approach. arXiv preprint arXiv:1906.03996. (2019).

16. Filatov, A., Filatov, A., Krinkin, K., Chen, B., Molodan, D.: 2d slam quality evaluation methods. 2017 21st Conference of Open Innovations Association (FRUCT). pp. 120–126 (2017).

17. Kohlbrecher, S., Von Stryk, O., Meyer, J., Klingauf, U.: A flexible and scalable slam system with full 3d motion estimation. 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics. pp. 155–160 (2011).

18. About ROS, <https://www.ros.org/about-ros/>.

19. Buyval, A., Afanasyev, I., Magid, E.: Comparative analysis of ROS-based monocular SLAM methods for indoor navigation. Ninth International Conference on Machine Vision (ICMV 2016). p. 103411K (2017).

20. Xu, L., Feng, C., Kamat, V.R., Menassa, C.C.: An Occupancy Grid Mapping enhanced visual SLAM for real-time locating applications in indoor GPS-denied environments. Automation in Construction. 104, 230–245 (2019).

21. Bayer, J., Cizek, P., Faigl, J.: On construction of a reliable ground truth for evaluation of visual slam algorithms. Acta Polytechnica CTU Proceedings. 6, 1–5 (2016).

22. Le, X.S., Fabresse, L., Bouraqadi, N., Lozenguez, G.: Evaluation of out-of-the-box ros 2d slams for autonomous exploration of unknown indoor environments. International Conference on Intelligent Robotics and Applications. pp. 283–296 (2018).

23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing. 13, 600–612 (2004).

24. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B., others: FastSLAM: A factored solution to the simultaneous localization and mapping problem. Proceedings of the National

conference on Artificial Intelligence. pp. 593–598 (2002).

25. Black, P.E.: Bresenham’s algorithm, <https://www.nist.gov/dads/HTML/bresenham.html>.

26. stage\_ros summary, [http://wiki.ros.org/stage\\_ros](http://wiki.ros.org/stage_ros).

27. TurtleBot3 - Specifications,  
<http://emanual.robotis.com/docs/en/platform/turtlebot3/specifications/>.

# Possibilities for software development for energy-limited constrained devices

Mario Hoss and Jens-Peter Akelbein

Hochschule Darmstadt, University of Applied Sciences  
{mario.hoss, jens-peter.akelbein} @h-da.de

**Abstract.** With the spread and rising complexity of IoT scenarios there are also new challenges emerging in planing and predicting the lifetime of application specific energy-limited resource constrained devices. To allow for a state of the art software development process, energy consumption as well available energy input will need to be considered during early stages of development. Depending on what information are available or predictable during which stage of the product lifecycle, adaptive behavior could also be used to supplemented or compensated for predictions. This paper presents an early stage work into a new research topic. It gives a review of the possibilities and challenges of predicting and determining energy consumption and input information throughout development and deployment of constrained devices. Contrasting existing approaches from related fields, the paper concludes in the outlook on the upcoming research questions.

**Keywords:** Energy Models · Adaptive Systems · IoT

## 1 Introduction

According to forecasts of exponential growth in the Internet of Things market [1], the demand for application-specific connected devices is expected to reach billions. This growth is driven by the emergence of a variety of new application scenarios, like Industrial IoT (IIOT) and Smart Cities. Many of these scenarios rely on masses of energy-limited constrained devices [2] that wirelessly transmit collected information for up to a decade or more without being connected to a permanent power supply. This poses new challenges in the design and development phases as it is critical for both battery operated and energy harvesting powered devices to optimize the power consumption of hard- and software to ensure that they can reach their targeted lifespan.

With more efficient power-saving modes for a System on a Chip (SoC), the potential complexity of the application has increased drastically. Until recently, commonly only class 0 and class 1 devices [2] have been used in energy-limited scenarios with such long lifespans. Both classes are not intended to use standard internet protocols. By now, class 2 devices are also a realistic option. Since common internet protocols are often used, state-of-the-art security is also required. Providing the necessary functionalities further increases software complexity, as

well as resource- and power-consumption. This increased complexity is also a major driver in the increased use of operating systems for the IoT, which should accelerate further.

This shift will likely be accompanied by a change in requirements for the developer. While writing bare-metal applications requires a more in-depth understanding of the underlying hardware, using operating systems providing abstraction layers and reusable libraries is suitable for more general software developers.

In light of these developments, there is a need to provide software developers with tools and methods to accurately predict the life expectancy of a constrained device in the development phase. Optimally, developers should be able to select combinations of hardware, software and energy-source and immediately receive a life expectancy prognosis. Such a forecast supports early development phases and allows for an "energy by design" approach in product development. Developers are able to test and customize their software for different energy-sources, for example different illuminance levels for solar powered devices. Using such forecasts in test and verification allows automated detection of energy bugs so software defects that cause abnormal power consumption.

Whilst mobile device applications face similar problems, solutions in this area are generally not applicable to the field of energy-limited constrained devices due to vastly different hardware as well as energy consumption pattern. Current lifetime prediction research for constrained devices originates in the Wireless Sensor Network (WSN) field with assumptions for less complex software, stronger hardware bindings, and more complex network communication. Like in the are of mobile devices, research for constrained devices also focuses on adaptive in-situ behavior rather than using a predictive approach during development. Existing approaches also utilize simplified battery models or additional hardware for measurements in the field.

Verifying the functionality and life expectancy by measuring power consumption during HIL testing is also not a realistic option. In addition to the to be discussed practical problems, achieving a sufficient test coverage is problematic. With a life expectancy of a decade, there is just not enough time during a normal development phase to sufficiently test the devices behavior over this time span.

This paper explores the possibilities for energy predictions of energy-limited application specific constrained devices in product development and its synergies with adaptive in-situ functionalities. As such it will be a first look into early research and a first step towards sufficient tooling for developers of energy-limited constrained devices.

## 2 State of the art

### 2.1 Energy consumption

In order to predict energy consumption, it is generally necessary to determine both the CPU power consumption of individual instructions as well as hardware

components. Because of the differences in usage scenarios and underlying hardware, energy consumption at an instruction level is much more relevant on mobile devices [3] than on constrained devices. Energy-limited constrained devices can also easily accumulate small inaccuracies, because unlike mobile devices, they are not regularly recharged. However, they do spend most of their lifetime in deep sleep modes, interrupted by repeating usage patterns, which reduces the overall complexity of the required models. Furthermore, due to significant difference in available resources, mobile devices can track the various hardware states in the field and are also able to measure the remaining battery charge [4], which due to per device cost is rarely an option for constrained devices. For instruction level models, determining data dependent dynamic power consumption of the worst case energy consumption was recently shown to result in a NP-hard problem, where an approximation cannot be made to a usable degree [5]. The examined class 2 constrained devices showed a variation in the data dependent power consumption of nearly half of a cores power dissipation. Another study also documented SoCs with a similar ratio [6]. More relevant for constrained devices are finite state machines (FSM) where the target hardware is modeled by different power states and their connecting transitions, determined through a measurement cycle. The parameters that influence the power consumption of a state and their behavior can then be identified by using regression analysis to create approximation functions for the parameter-dependent energy consumption of peripheral devices [7]. Transition triggers are identified by power bursts, since power states often do not line up with the utilization in software [8]. Automatically creating and refining such state machines is still an open research question [9]. A more detailed overview for both types of models can be found in earlier work [10].

There is also the established practice of using power consumption measurements during Hardware in the Loop (HIL) testing [11]. Woehrle et al. [12] validated WSN nodes utilizing HIL tests and the testbed Flocklab [13] also supports automated power tests for WSN nodes. This however puts practical limits on the test coverage, since the constrained devices are designed to run for years, making the coverage minimal in a normal development time frame. "Wearables" face this problem to a lesser degree. "Rocketlogger" is designed for in-situ measurements of the energy consumption through a normal wearable lifecycle of a few days [14].

One aspect that complicated past work on energy consumption models is the consideration of production irregularities and a variance in energy consumption in different power modes [15, 16]. These variances are usually considered in energy consumption models by factoring in an error margin during the creation. The variance has to be considered for any measurements on the actual hardware, be it HIL tests or in the creation of energy consumption models.

Energy-aware software engineering is used for mobile devices. This includes the identification of energy bugs and hotspots [17] and static analyses [18], which can also be used for general software development [19]. Energy consumption models and measurements are also utilized for compiler optimization [20].

## 2.2 Energy input

There is ample work on models for the remaining battery capacity [21], but the ambient temperature needs to be considered for more accurate results [22]. Battery capacity approximation also exists for energy harvesting powered constrained devices [23]. To predict the energy input of solar cells, both static and dynamic factors have to be considered. The static input consists of the efficiency of the cell for the different light sources and their respective wavelength. Information about the cells efficiency is readily available from the manufacturer and can be verified by confirmation measurements published in the biannual solar cell efficiency tables [24]. The dynamic input is dependent on environmental conditions during runtime, for example on the available illuminance level. A list of expected influencing factors indoors is listed in figure 1. A list for factors influencing the efficiency of photovoltaic systems in general can be found in [25]. Weather reports have also been used for short-term [26] and long-term [27] predictions of the energy input of solar powered constrained devices operating outdoors.

Positioning (direct)	Positioning (indirect)	Hardware properties	Hardware condition
Light composition	Ambient temperature	Efficiency of the solar cell at different wavelength	Cell age
Light level	Partially shaded cells	Efficiency of energy-harvesting IC	Dust and dirt
Cell orientation and angle	Weather and user interaction (e.g. shades)	Availability of maximum power point tracking	Hardware failure

**Fig. 1.** Factors influencing the energy input of solar powered constrained devices

Sufficient accurate data-sets for the prediction of indoor light levels in living- and functional buildings are currently lacking. Existing architectural information can not be used, as the field focuses on the indoor light level relative to the outdoor light level [28]. Consumer protection studies [29] can be useful but generally focus on living spaces. This information gap will be filled over the course of the LOEWE3 project "LONG MOVE" with continuous measurements over 18 month in a wide variety of functional buildings. Important aspects are consideration of both incoming sunlight over different seasons, as well as artificial light sources and information about the placement of the devices.

## 2.3 Energy management and adaptive functionality

Energy management functionalities are typically used to track and manage hard to predict power consumption in the field of mobile devices. For mobile devices, this is largely caused by the impact user interaction and usage pattern have on

the device power consumption. Such an approach was adopted by Tamkittikhun et al. [30] for solar powered embedded devices. Different to mobile devices, direct user interaction is rarely possible for these devices. However, since user intervention is not an option to achieve lifetime goals through regular charging, adaptive behavior to reduce functionality or increase sleep durations is often necessary. They are also faced with hard to predict power consumption in the form of processing incoming transmissions. For this Tamkittikhun et al. combined power measurements of individual functionalities before deployment with a functionality counter during runtime to estimate energy consumption on the device. This allows for lifetime predictions, as well as adaptation of the device functionality to meet lifetime goals. Since energy-limited constrained devices achieve their lifetime goal by remaining in sleep states, the transceiver is also powered off. As such energy consumption caused by processing incoming transmissions rarely has to be considered.

For energy-limited constrained devices there are also research activities into adapting the device behavior depending on the available energy [16,31,32]. Here the prediction problems often stem from hardware variances as well as the dependence of the energy-input on the environment.

Lachenmann et al. [31] proposed a programming abstraction for constrained devices, associating different sets of functionalities with different energy levels. For that the energy consumption of different functionalities is measured during development and associated with these levels. Depending on the remaining charge, the devices then adapt their behavior in the field to a functionality level with a lower energy consumption. For this a battery monitor hardware and a battery model mapping the voltage to the remaining battery capacity was used. Sieber [16] followed a similar approach but focused on more dynamic performance adjustments in the field based on an energy bucket concept in addition to the energy consumption model. For that application and system functionalities are defined with an energy priority and individual energy consumption limits. Instead of the use of measurement hardware on the device, the remaining battery charge is estimated based on a model of linearized remaining charge values based on measurements taken of the battery before deployment. This approach is motivated by profiting off and coping with production variances. Geissdoerfer et al. [32] utilized local energy input predictions on energy harvesting powered devices during run time, simultaneously taking into account the batteries current state of charge. The underlying model gets adjusted through state of charge feedback to the predictions.

In general, it is possible to identify several different approaches to adaptive energy management. There are those solutions that collect additional information about their environment, be it light intensity, temperature or state of charge, and act on that information. This often requires additional hardware, resulting in increased cost and software overhead for each device. Other approaches [16,30] do not collect environmental information at runtime, and are instead shipped with, and act on, energy-consumption- and -input-models created before deployment. These models can come in a wide variety of forms, from battery-models to

simply measuring the power consumption of specific functions or features. While these approaches do not require additional hardware, they still produce software overhead on each device. However, if this approach is applied to constrained devices rather than to more complex embedded devices, then it should be possible to move large parts of the computation into the development phase, reducing the required overhead or even making the solution completely application-neutral. Lastly, there are managed devices that can communicate with a management server. In addition to the possibility to report environmental conditions and off-load costly model based computations, it is also possible to provide a device with additional information, like weather forecasts to predict the energy input of solar powered devices [26]. This is however achieved by additional overhead for the management of the device as well as running costs for the management infrastructure. If however predictions about the necessity of "over the air" firmware updates come true, the additionally required overhead on the device would be minimal. Current management options for constrained devices, like "Lightweight Machine to Machine" (LWM2M) [33], have not found wide acceptance, likely due to the introduced overhead [34]. However, the current IETF draft for "Software Updates for the Internet of Things" (SUIT) [35] could provide such management options in the long term future, though likely first for devices with a permanent power supply. Another advantage of this approach is, that with the energy consumption model centralized in one location, the model could be adapted and refined over the lifetime. The categories listed are not mutually exclusive. Hybrid approaches like Lachenmann et al. [31] rely on both energy consumption models and additional measurement hardware on the device.

### 3 Outlook

This paper summarizes the literature review giving a comparison of current approaches. Based on this review, three research questions are derived to be answered in the next steps of research: *Q1*: How can energy consumption models be utilized for simulating power consumption of energy-limited constrained device? *Q2*: How can energy-input predictions be utilized for simulating the energy-input variability of energy-limited constrained devices? *Q3*: How can energy-input- and -consumption-information be utilized for lifetime prediction during the software development process for application specific constrained devices?

To find answers to these questions the design science research process is used by creating and evaluating proof of concept prototypes iteratively. These allow to determine the usability and precision of the individual models as well as forecasts based on a combination of both models. As a result, it will be possible to determine the prediction accuracy of each element within the forecast depending on the phase of the development process. Identifying the limitations of the approach allows defining where adaptive in-situ behavior can address deviations found over long time periods. This leads to determining a minimum requirement profile.

Based on these answers, it should become possible to further evaluate how the presented approach of forecasts can be integrated in today's product development processes.

This work was supported by the Hessenagentur within the Project "LONG MOVE - Leistungsadaptation und ortsbezogene Verhaltensregeln für eine nachhaltige IoT-Sensorik in der Gebäudeausstattung zur modularen Vernetzung von Einheiten" (HA-Projekt-Nr. 802/19-122).

## References

1. Middleton, P.: Forecast analysis: Internet of things endpoints, worldwide, 2017 update. Tech. rep., Gartner Research (2017), <https://www.gartner.com/en/documents/3841268/forecast-analysis-internet-of-things-endpoints-worldwide>
2. Bormann, C., Ersue, M., Keranen, A.: Terminology for constrained-node networks. RFC 7228, RFC Editor (May 2014), <http://www.rfc-editor.org/rfc/rfc7228.txt>
3. Shukla, N.K., Pila, R., Rawat, S.: Utilization-based power consumption profiling in smartphones. In: 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I). pp. 881–886 (Dec 2016). <https://doi.org/10.1109/IC3I.2016.7919046>
4. Dong, M., Zhong, L.: Self-constructive high-rate system energy modeling for battery-powered mobile systems. In: Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services. pp. 335–348. MobiSys '11, ACM, New York, NY, USA (2011). <https://doi.org/10.1145/1999995.2000027>
5. Morse, J., Kerrison, S., Eder, K.: On the limitations of analyzing worst-case dynamic energy of processing. *ACM Trans. Embed. Comput. Syst.* **17**(3), 59:1–59:22 (Feb 2018). <https://doi.org/10.1145/3173042>
6. Ascia, G., Catania, V., Palesi, M., Sarta, D.: An instruction-level power analysis model with data dependency. *VLSI Design* **12**(2), 245–273 (2001). <https://doi.org/10.1155/2001/82129>
7. Friesel, D., Buschhoff, M., Spinczyk, O.: Parameter-aware energy models for embedded-system peripherals. In: 2018 IEEE 13th International Symposium on Industrial Embedded Systems (SIES). pp. 1–4 (June 2018). <https://doi.org/10.1109/SIES.2018.8442096>
8. Pathak, A., Hu, Y.C., Zhang, M., Bahl, P., Wang, Y.M.: Fine-grained power modeling for smartphones using system call tracing. In: Proceedings of the Sixth Conference on Computer Systems. pp. 153–168. EuroSys '11, ACM, New York, NY, USA (2011). <https://doi.org/10.1145/1966445.1966460>
9. Buschhoff, M., Friesel, D., Spinczyk, O.: Energy models in the loop. *Procedia Computer Science* **130**, 1063 – 1068 (2018). <https://doi.org/10.1016/j.procs.2018.04.154>, the 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops
10. Hoss, M., Westmeier, F., Akelbein, J.: Low cost high resolution ampere meter for automated power tests for constrained devices. In: Bleimann, U., Burkhardt, D., Humm, B., Löw, R., Regier, S., Stengel, I., Walsh, P. (eds.) Proceedings of the 5th Collaborative European Research Conference (CERC 2019), Darmstadt, Germany, March 29-30, 2019. CEUR Workshop Proceedings, vol. 2348, pp. 81–88. CEUR-WS.org (2019)

11. Waeltermann, P.: Hardware-in-the-loop: The technology for testing electronic controls in vehicle engineering. Tech. rep., dSPACE (2016), [https://www.dspace.com/files/pdf1/dspace-paper\\_hil\\_overview\\_waeltermann\\_e\\_160405.pdf](https://www.dspace.com/files/pdf1/dspace-paper_hil_overview_waeltermann_e_160405.pdf)
12. Woehrle, M.: Testing of Wireless Sensor Networks. Ph.D. thesis, TU Karlsruhe (2010). <https://doi.org/10.3929/ethz-a-006206723>
13. Lim, R., Ferrari, F., Zimmerling, M., Walser, C., Sommer, P., Beutel, J.: Flocklab: A testbed for distributed, synchronized tracing and profiling of wireless embedded systems. In: 2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). pp. 153–165 (April 2013). <https://doi.org/10.1145/2461381.2461402>
14. Sigrist, L., Gomez, A., Lim, R., Lippuner, S., Leubin, M., Thiele, L.: Measurement and validation of energy harvesting iot devices. In: Proceedings of the 2017 Design, Automation & Test in Europe Conference & Exhibition (DATE 2017). Lausanne, Switzerland (Mar 2017)
15. Hurni, P., Nyffenegger, B., Braun, T., Hergenroeder, A.: On the accuracy of software-based energy estimation techniques. In: Marrón, P.J., Whitehouse, K. (eds.) Wireless Sensor Networks. pp. 49–64. Springer Berlin Heidelberg (2011), [https://link.springer.com/chapter/10.1007%2F978-3-642-19186-2\\_4](https://link.springer.com/chapter/10.1007%2F978-3-642-19186-2_4)
16. Sieber, A.: Energiemanagement fuer drahtlose tief eingebettete Systeme : Feingranulares, faires Management zum Erreichen von Lebenszeitzielen bei knappen Energieressourcen. doctoralthesis, BTU Cottbus - Senftenberg (2016), <https://opus4.kobv.de/opus4-btu/frontdoor/index/index/docId/3928>
17. Baek, D., Park, J.H., Lee, J.W.: An energy efficiency grading system for mobile applications based on usage patterns. The Journal of Supercomputing **74**(12), 6502–6515 (Dec 2018). <https://doi.org/10.1007/s11227-018-2439-x>
18. Banerjee, A., Chong, L.K., Chattopadhyay, S., Roychoudhury, A.: Detecting energy bugs and hotspots in mobile apps. In: Proceedings of the 22Nd ACM SIGSOFT International Symposium on Foundations of Software Engineering. pp. 588–598. FSE 2014, ACM, New York, NY, USA (2014). <https://doi.org/10.1145/2635868.2635871>
19. Gastel, P.v., Gastel, B.v., Eekelen, M.v.: Detecting energy bugs and hotspots in control software using model checking. In: Conference Companion of the 2Nd International Conference on Art, Science, and Engineering of Programming. pp. 93–98. Programming Companion, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3191697.3213805>
20. Roth, M., Luppold, A., Falk, H.: Measuring and modeling energy consumption of embedded systems for optimizing compilers. In: Proceedings of the 21st International Workshop on Software and Compilers for Embedded Systems. pp. 86–89. SCOPES '18, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3207719.3207729>
21. Rodrigues, L.M., Montez, C., Budke, G., Vasques, F., Portugal, P.: Estimating the lifetime of wireless sensor network nodes through the use of embedded analytical battery models. Journal of Sensor and Actuator Networks **6**(2) (2017). <https://doi.org/10.3390/jsan6020008>
22. Rodrigues, L.M., Montez, C., Moraes, R., Portugal, P., Vasques, F.: A temperature-dependent battery model for wireless sensor networks. Sensors **17**(2) (2017). <https://doi.org/10.3390/s17020422>
23. Buchli, B., Aschwanden, D., Beutel, J.: Battery state-of-charge approximation for energy harvesting embedded systems. In: Demeester, P., Moerman, I., Terzis, A. (eds.) Wireless Sensor Networks. pp. 179–196. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)

24. Green, M.A., Dunlop, E.D., Levi, D.H., Hohl-Ebinger, J., Yoshita, M., Ho-Baillie, A.W.: Solar cell efficiency tables (version 54). *Progress in Photovoltaics: Research and Applications* **27**(7), 565–575 (2019). <https://doi.org/10.1002/pip.3171>
25. Rao, R.R., Mani, M., Ramamurthy, P.C.: An updated review on factors and their inter-linked influences on photovoltaic system performance. *Heliyon* **4**(9), e00815 (2018). <https://doi.org/https://doi.org/10.1016/j.heliyon.2018.e00815>
26. Kraemer, F.A., Ammar, D., Braten, A.E., Tamkittikhun, N., Palma, D.: Solar energy prediction for constrained iot nodes based on public weather forecasts. In: *Proceedings of the Seventh International Conference on the Internet of Things*. pp. 2:1–2:8. *IoT '17*, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3131542.3131544>
27. Buchli, B., Sutton, F., Beutel, J., Thiele, L.: Towards enabling uninterrupted long-term operation of solar energy harvesting embedded systems. In: *Proceedings of the 11th European Conference on Wireless Sensor Networks - Volume 8354*. pp. 66–83. *EWSN 2014*, Springer-Verlag New York, Inc., New York, NY, USA (2014). [https://doi.org/10.1007/978-3-319-04651-8\\_5](https://doi.org/10.1007/978-3-319-04651-8_5)
28. Mardaljevic, J., Christoffersen, J.: climate connectivity in the daylight factor basis of building standards. *Building and Environment* **113**, 200 – 209 (2017). <https://doi.org/https://doi.org/10.1016/j.buildenv.2016.08.009>, advances in daylighting and visual comfort research
29. Wold, C.: Analysis of background illuminance levels during television viewing. Tech. rep., Collaborative Labeling and Appliance Standards Program (CLASP) (2011), [https://www1.eere.energy.gov/buildings/appliance\\_standards/pdfs/analysis\\_of\\_background\\_illuminance\\_levels\\_during\\_television\\_viewing.pdf](https://www1.eere.energy.gov/buildings/appliance_standards/pdfs/analysis_of_background_illuminance_levels_during_television_viewing.pdf)
30. Tamkittikhun, N., Hussain, A., Kraemer, F.A.: Energy consumption estimation for energy-aware, adaptive sensing applications. In: Bouzeffrane, S., Banerjee, S., Sailhan, F., Boumerdassi, S., Renault, E. (eds.) *Mobile, Secure, and Programmable Networking*. pp. 222–235. Springer International Publishing, Cham (2017). [https://doi.org/10.1007/978-3-319-67807-8\\_17](https://doi.org/10.1007/978-3-319-67807-8_17)
31. Lachenmann, A., Marrón, P.J., Minder, D., Rothermel, K.: Meeting lifetime goals with energy levels. In: *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*. pp. 131–144. *SenSys '07*, ACM, New York, NY, USA (2007). <https://doi.org/10.1145/1322263.1322277>
32. Geissdoerfer, K., Jurdak, R., Kusy, B., Zimmerling, M.: Getting more out of energy-harvesting systems: energy management under time-varying utility with preact. In: *Proceedings of the 18th International Conference on Information Processing in Sensor Networks, IPSN 2019*, Montreal, QC, Canada, April 16-18, 2019. pp. 109–120 (2019). <https://doi.org/10.1145/3302506.3310393>
33. Rao, S., Chendanda, D., Deshpande, C., Lakkundi, V.: Implementing lwm2m in constrained iot devices. In: *2015 IEEE Conference on Wireless Sensors (ICWiSe)*. pp. 52–57 (Aug 2015). <https://doi.org/10.1109/ICWISE.2015.7380353>
34. Schmelzer, P., Akelbein, J.: Evaluation of hardware requirements for device management of constrained nodes based on the LWM2M standard. In: *Proceedings of the 5th Collaborative European Research Conference (CERC 2019)*, Darmstadt, Germany, March 29-30, 2019. pp. 103–110 (2019), <http://ceur-ws.org/Vol-2348/short02.pdf>
35. Moran, B., Meriac, M., Tschofenig, H., Brown, D.: A firmware update architecture for internet of things devices. *Internet-Draft draft-ietf-suit-architecture-05*, IETF Secretariat (April 2019), <http://www.ietf.org/internet-drafts/draft-ietf-suit-architecture-05.txt>

## Chapter 2

---

# Data Computing and Artificial Intelligence



# Development and Validation of an Algorithm for Emissivity-Corrected Pyrometry Independent of Material Properties

Denise Reichel<sup>1,2</sup>, 

<sup>1</sup>SRH Hochschule Heidelberg, Bonhoeffer Straße 11, 69123 Heidelberg, Germany  
denise.reichel@srh.de

<sup>2</sup>Steinbeis Transferzentrum Sensorik und Informationssysteme, Moltkestraße 30, 76133 Karlsruhe, Germany  
denise.reichel@sensin.eu

**Abstract.** True contactless temperature measurement or pyrometry is closely connected to the knowledge of the material's emissivity which is generally a function of the material surface, wavelength, temperature and angle. The present paper will show how this indispensable parameter for contactless thermography can be deduced from real temperature data without any knowledge of and any demands on the material's chemical or physical structure. An algorithm was developed to deduce the unknown emissivity value by an iterative process involving random data taken with a virtual ratio or two-colour pyrometer within a temperature regime of 1100 K +/- 100 K. Possible influences on the simulation, namely level of true emissivity, emissivity ratio, measurement uncertainty and detection wavelength, were studied to find an increasing mismatch between the simulated and true emissivity for larger uncertainties and wavelengths which had been expected following Planck's Law of Thermal Radiation. An increased error has also been found for larger values of the true emissivity whereas no significant influence could be shown for the emissivity ratio of the two channels of a ratio pyrometer. The simulation was considered to be completed for results corresponding to a maximum deviation of 8 K from a given true temperature of 1100 K.

**Keywords:** Pyrometry, Thermography, Temperature Measurement, Emissivity, Algorithm, Ratio Pyrometry, Two-colour Pyrometry

## 1 Introduction

### 1.1 Importance of Emissivity for Pyrometry

Sensors for contactless temperature measurement are indispensable for applications that do not allow for direct contact due to surrounding circumstances, e.g. rotation bodies, or due to purity requirements of the material, e.g. semiconductors for microelectronic use. For this reason measuring devices (here: pyrometers) are needed which are able to detect thermal radiation from a distance.

Each body emits thermal radiation due to its own temperature and pyrometers measure this radiation in different spectral regimes (spectral pyrometers) or across the entire spectrum (broadband pyrometers). The detection regime, however, usually only stretches from the visual to the mid-infrared regime due to typical temperatures between 0 °C and 3000 °C. Planck's Radiation Law [1]

$$M_{bb}^{\lambda} = \frac{8\pi hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1} \quad (1)$$

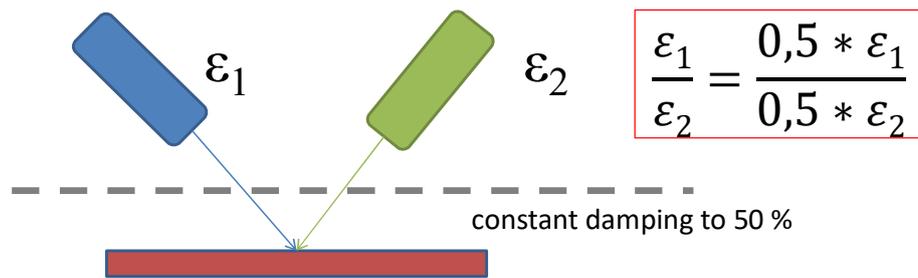
connects the object's temperature  $T$  to the detected thermal radiance  $M$  as a function of the detection wavelength  $\lambda$ .  $h$  and  $k_B$  denote the Planck and Boltzmann constant, respectively. However, this only holds for black bodies (here denoted by  $bb$ ), i.e. objects that absorb and – according to Kirchhoff's Law [2] – emit 100 % of the incident radiation ( $\varepsilon = 1$ ). Generally speaking, Kirchhoff's Law states that the spectral and directional absorptivity  $\alpha_{s,d}$  equals  $\varepsilon_{s,d}$  at thermal equilibrium:

$$\alpha_{s,d}(\lambda, \gamma)_T = \varepsilon_{s,d}(\lambda, \gamma)_T \quad (2)$$

This is true for all objects, but hypothetical black bodies show a constant emissivity over wavelength and angle. Real bodies, however, are so called “coloured”, which means that their absorption and thus emission actually differs as a function of wavelength (cf. (2)). For those bodies (1) is incomplete, it needs to be expanded by the emissivity  $\varepsilon$ . This is a very crucial parameter as it determines a true temperature measurement which makes it the most desired material parameter for pyrometry. In addition to the variables named in (2)  $\varepsilon$  also depends on the sample surface. Due to diffuse reflection which leads to an increased probability that incoming radiation is absorbed and thus to a greater absorptivity emissivity is larger for rough than for smooth surfaces. Moreover, emissivity may be a function of temperature itself. Further, dust or smoke may change the effective emissivity for the pyrometer. This shows the importance of the emissivity for temperature determination beyond its existence as a material parameter. Without the knowledge of an effective emissivity for the measurement pyrometry is rather radiometry of thermal radiation.

## 1.2 Ratio Pyrometry or Two-Colour Pyrometry

If an object's emissivity is smaller than *one* and constant independently of the spectral position, the object is said to be “grey”. However, this reduction in emittance may also be effectively caused by any undesired surrounding appearances such as smoke that is assumed to act like a neutral density filter, i.e. reducing the power density and thus the apparent emittance of the material by a single factor also independently of the considered wavelength.



**Fig. 1.** Working principle of any ratio pyrometer for background-corrected temperature measurement of a hot object (shown in red) if the background meant constant damping illustrated as a dashed line. The two channels of a ratio pyrometer are represented by the green and blue bodies respectively. The damping affects both channels equally and is thus cancelled out of the quotient.

If two detectors 1 and 2 are applied to measure at the same position on the object's surface, but at different spectral regimes, and the corresponding emissivities  $\varepsilon_1$  and  $\varepsilon_2$  do not change during the measurement then their ratio remains unaffected by a constant damping as it is visualized in **Fig. 1**. The damping factor can be cancelled out from the emissivity ratio thus allowing for emissivity-independent temperature measurement. [3, 4]

Soon this idea had been applied to grey bodies which for a detector behave in the exact same way, namely reducing the emittance at all spectral positions by a constant factor which finally also leads to emissivity-corrected temperature measurement. Theoretically, it works for coloured bodies similarly well if the two detecting wavelengths are as close to each other as possible. In theoretically infinitely close vicinity any change in emissivity as a function of wavelength approaches *zero* which means that for an infinitely small spectral range any coloured body is also grey. However, this places high demands on measurement technology which these days cannot be met, yet, at justifiable costs. Current solutions include assumptions on the relationship between both temperature and emissivity variations [5, 6] or wavelength. The latter includes – primarily for multi-wavelength-pyrometry – a linear and log-linear emissivity model, which refer to the respective relationship between emissivity and wavelength [7, 8].

Finally, ratio pyrometers may be used if a hot sample is smaller than the detection spot of the pyrometer. Assuming a “cold” background, i.e. a surrounding temperature that is much smaller than the object's temperature, then temperature measurement will only be marginally affected by the spot size mismatch as it corresponds effectively to a constant damping. Please note, that the temperature reading would significantly change if a single-wavelength pyrometer was used.

### 1.3 Status Quo

There have been a few approaches to determine the true sample temperature using pyrometry without possessing any knowledge on the emissivity. A review may be found here [9]. However, at present additional information is required. This may be accomplished through reflectometry which yields a measure for the material's emis-

sivity using Kirchhoff's Law and conservation of radiation if transmission can be excluded [10, 11, 12]. Other approaches use a combination of ratio pyrometry for grey bodies as described above and some knowledge on the material properties [13, 14]. In particular, ratio pyrometry is helpful if one can use literature data on the properties of the material assuming that the emissivity ratio does not change with modified process conditions. Moreover, it may be successfully applied to surfaces that are or become grey during thermal treatment, namely oxidizing steel surfaces [15]. Although the time slot during oxide growth affects the measurement significantly while the true surface temperature may stay the same, but the changing readings can be vice versa interpreted in terms of increasing oxide thickness as a function of time [16]. Multi-wavelength pyrometry uses more than two detecting wavelengths to replicate the course of the thermal radiation spectrum as the full spectrum unambiguously determines the object's temperature regardless of its emissivity [17, 18].

However, considering widely commercially available ratio pyrometers all of the approaches that do not utilize additional measurements (e.g. reflectometry) rely on further information to give a true temperature. Indeed, it is possible to determine any object's temperature within the limits to be discussed if despite the ratio channel the two underlying single channels are simultaneously measured. An iterative algorithm that will be presented then gives the only possible combination of  $\varepsilon_1$  and  $\varepsilon_2$  and yields a true temperature reading.

## 2 Algorithm

If the detection spot is fully filled by the hot object whose temperature is to be measured and no damping needs to be considered then both the single-wavelength signals as well as the ratio measurement from a two-colour pyrometer may be used to determine the material emissivity and thus the object's true temperature without any additional information. Further, this iterative algorithm assumes an exact overlap of the detection spot for all involved wavelength regimes. If these basic requirements are met, than all three (single-wavelength and ratio) channels of a two-colour pyrometer should "see" the same temperature, i.e. all three readings should match. This is mathematically described by the following equation system:

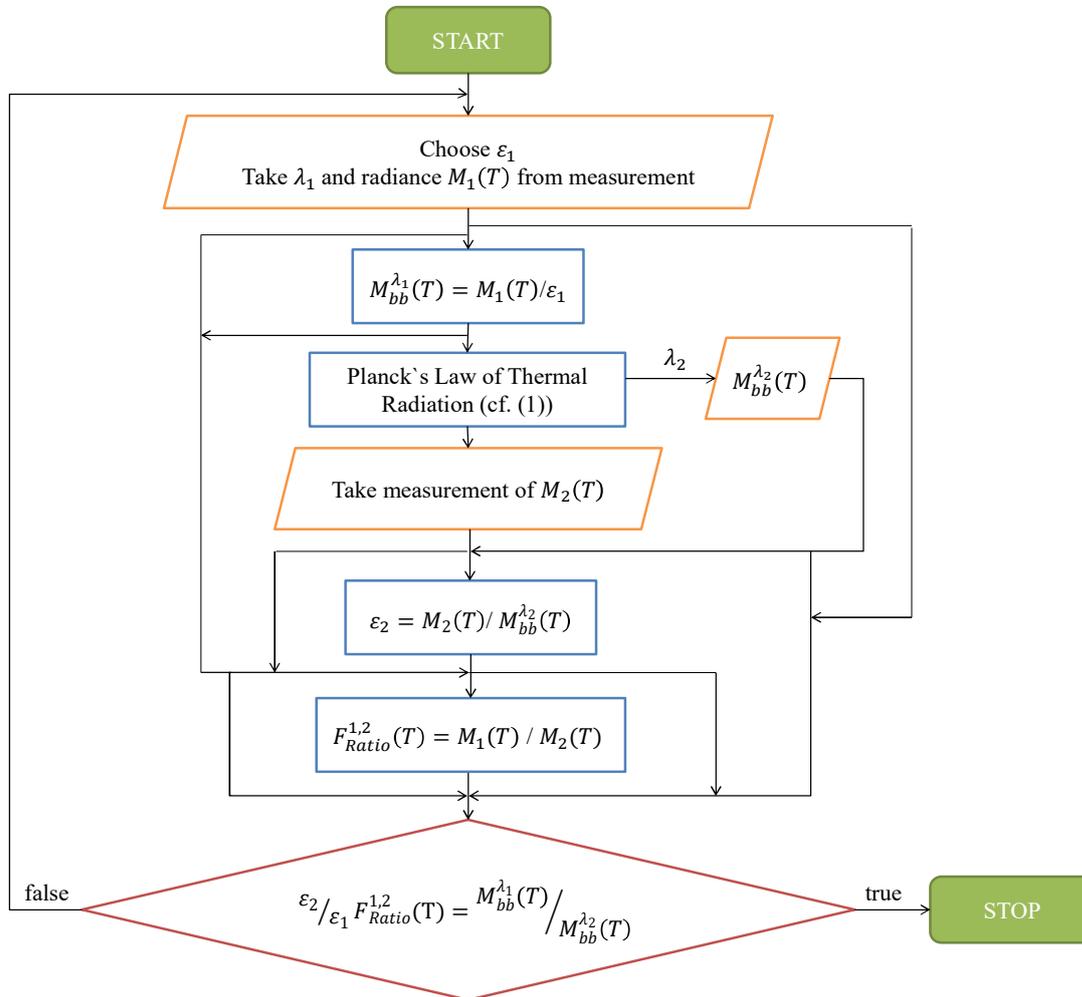
$$\varepsilon_1 M_{bb}^{\lambda_1}(T) = M_1(T) \quad (3)$$

$$\varepsilon_2 M_{bb}^{\lambda_2}(T) = M_2(T) \quad (4)$$

$$\varepsilon_2 / \varepsilon_1 F_{Ratio}^{1,2}(T) = M_{bb}^{\lambda_1}(T) / M_{bb}^{\lambda_2}(T) \quad (5)$$

This threefold equation system – consisting of the emissivities  $\varepsilon_1$  and  $\varepsilon_2$  which correspond to the two single-wavelength-channels at  $\lambda_1$  and  $\lambda_2$ , of the respective real

spectral radiances  $M_1(T)$  and  $M_2(T)$ , of their ratio  $F_{Ratio}^{1,2}(T)$  as well as of their black body equivalents  $M_{bb}^{\lambda_1}(T)$  and  $M_{bb}^{\lambda_2}(T)$  – can be solved by an iterative algorithm. It is straightforward to show that for the right combination of  $\varepsilon_1$  and  $\varepsilon_2$  equation (5) results in a true statement.

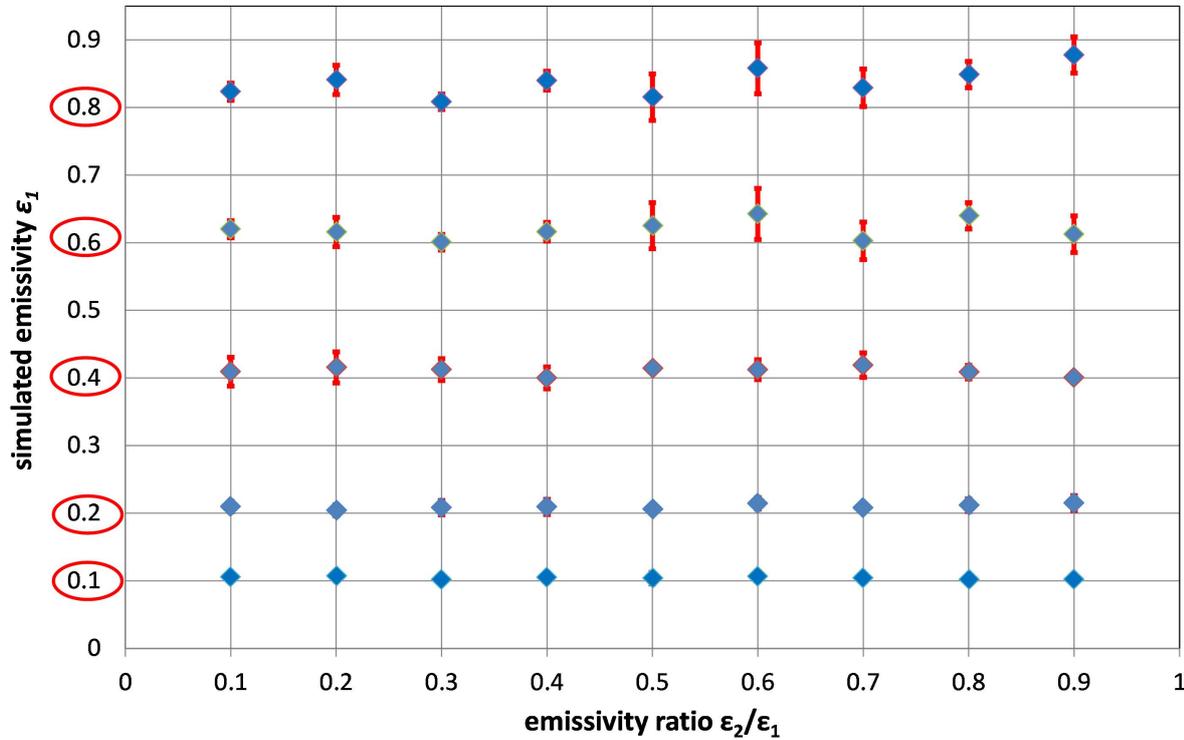


**Fig. 2.** Visualization of an algorithm determining the true emissivity of an object by evaluating all three attainable channels from a two-colour pyrometer.

Even if no information is given on the material properties the true temperature can be gained by assuming a starting value for  $\varepsilon_1$ . An iterative approach (cf. **Fig. 2**) follows which may be shortened by an educated guess on  $\varepsilon_1$ . Consequently,  $M_{bb}^{\lambda_1}(T)$  may be calculated taking the measured data from  $M_1(T)$  and knowing the detecting wavelength of the first channel. As it must be safe to assume  $T$  is equal for all channels  $M_{bb}^{\lambda_2}(T)$  can be easily gained and using the measured data of  $M_2(T)$   $\varepsilon_2$  is readily obtained. Finally, the ratio of  $\varepsilon_1/\varepsilon_2$  and  $F_{Ratio}^{1,2}(T)$  are compared to  $M_1(T) / M_2(T)$ . If this comparison does not satisfy equation (5) the iteration is repeated with a different starting emissivity  $\varepsilon_1$  until a true statement is achieved.

### 3 Results and Discussion

The above described iteration process was run for nine different hypothetically real emissivity ratios at a wavelength of  $(1.0 \pm 0.1) \mu\text{m}$  for temperatures of  $(1100 \pm 100) \text{K}$  and a measurement uncertainty of 25 K.



**Fig. 3.** Simulated emissivity as a function of a true emissivity ratio. The real object's emissivities that were used to obtain measurement data are circled in red. Please note that only the measured results from the two respective single-wavelength channels and from the ratio channel enter the simulation. The stopping condition was set to differences in the calculated and measured ratio factor  $F_{Ratio}^{1,2}(T)$  of less than  $1\text{E-}3$ . This means that at a temperature of 1100 K a maximum error of 8 K needs to be accepted.

The results presented in **Fig. 3** do not show a significant influence of the emissivity ratio on the simulation results. There seems to be an increased error in temperature due to a rise in the simulation uncertainty for larger nominal emissivities. In fact, this is a consequence of the exponential relationship of thermal radiation and temperature.

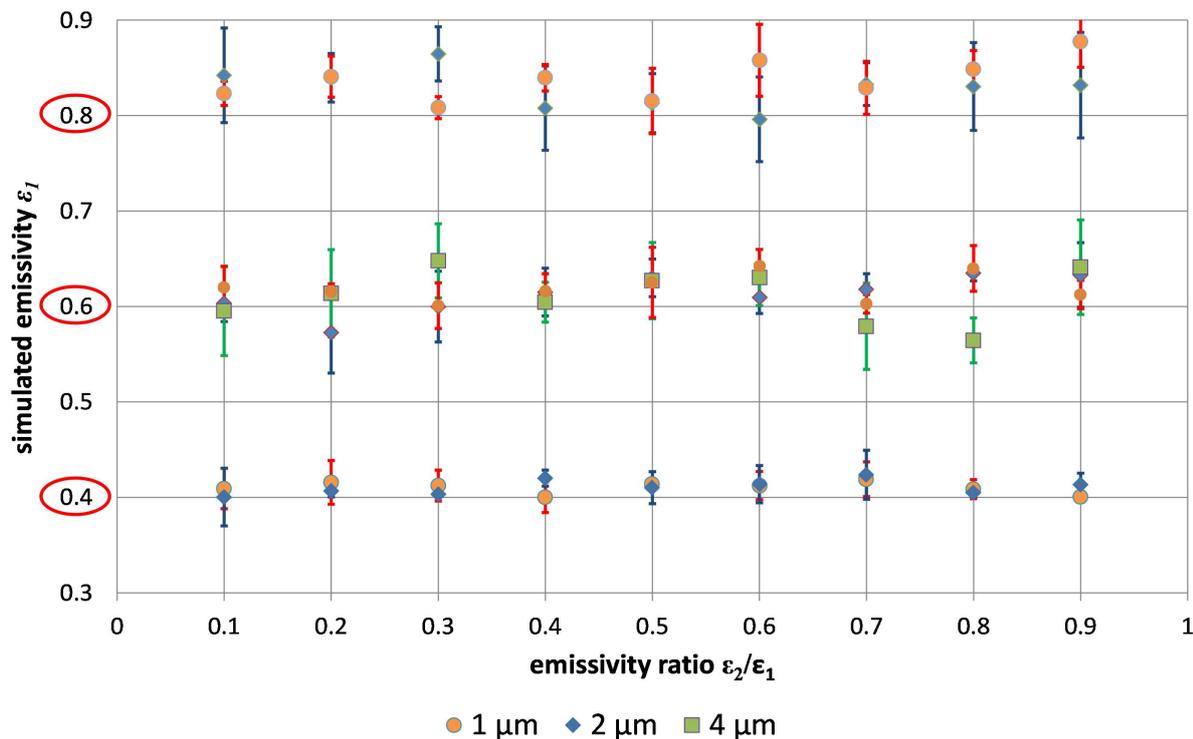
**Table 1** shows the maximum temperature variation for the maximum deviation of the simulated and true emissivity at a temperature of 1100 K and at a detecting wavelength of  $1 \mu\text{m}$ . It can be seen that despite the small standard deviation at a true emissivity of 0.1 the resulting temperature error is indeed larger because the percentage deviation has equally increased which is the decisive contribution. Moreover, the presented errors attribute to less than 1.2 % for an assumed object temperature of 1100 K which is in the order of the set temperature uncertainty of 25 K. However, the stopping condition which accepts a maximum temperature error of 8 K corresponding

to 0.7 % at a temperature of 1100 K, needs to be beared in mind although this still does not exceed the afore mentioned uncertainty of 25 K.

**Table 1.** Maximum temperature error for the considered true emissivities according to **Fig. 3** if the maximally deviated simulation result is used for temperature determination

True Emissivity	0.1	0.2	0.4	0.6	0.8
	12.71 K	8.94 K	8.79 K	9.44 K	9.53 K

Farther, these results at a wavelength regime of  $(1.0 \pm 0.1) \mu\text{m}$  were compared to two further spectral positions at  $(2.0 \pm 0.1) \mu\text{m}$  and at  $(4.0 \pm 0.1) \mu\text{m}$  (cf. **Fig. 4**).

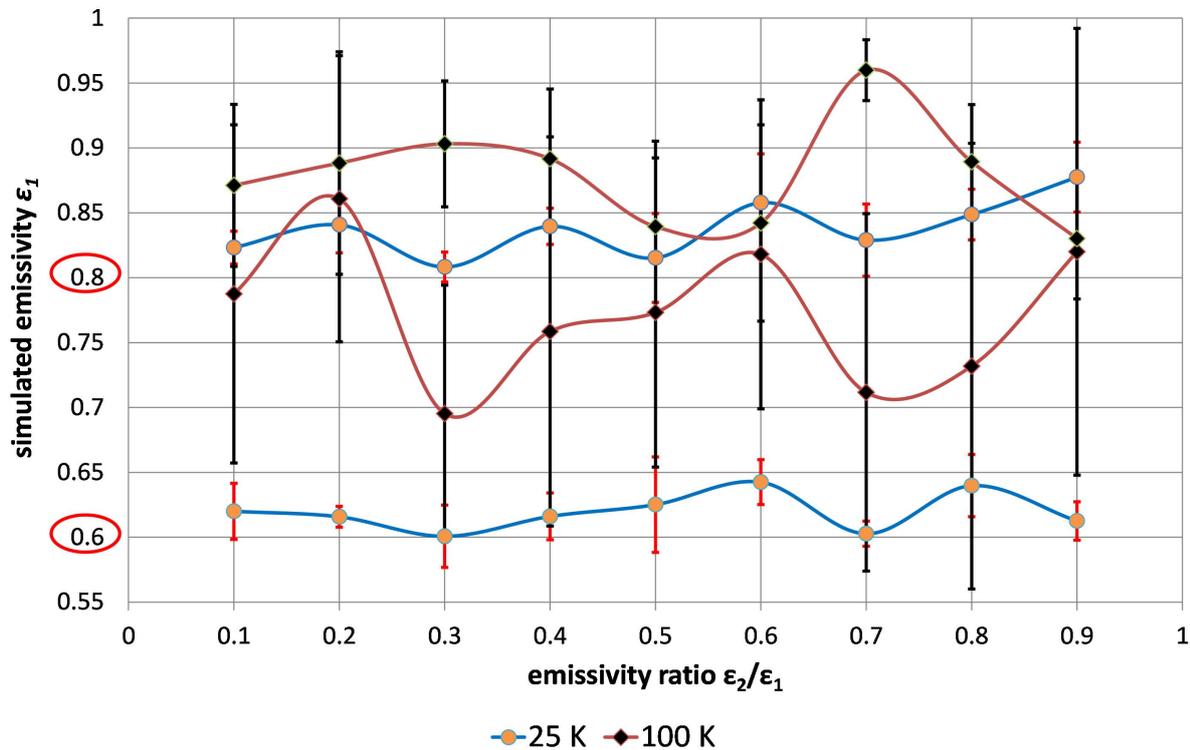


**Fig. 4.** Simulated emissivity as a function of a true emissivity ratio and the central wavelength. The real object’s emissivities that were used to obtain measurement data are circled in red.

An analysis of variances shows that the deviations are insignificant by a probability of more than 99 %. Larger standard deviations may be attributed, however, to the behaviour of Planck’s Law of Thermal Radiation for larger wavelengths. Towards larger wavelengths the maximum spectral emittance slowly declines in an asymptotic manner. The distance between different temperatures decreases likewise and even a small error in the assumed emissivity will result in comparatively large errors in temperature reading and vice versa. That means, the given temperature uncertainty of 25 K causes larger deviations in the emissivity simulation for larger wavelengths.

Finally, the measurement uncertainty was increased from 25 K to 100 K. As to be expected this led to a comparatively enormous increase both in the mean values and standard deviations (cf. **Fig. 5**). This is due to an increase in possible solutions to the equation system (3-5) which may be compensated by a stricter stopping condition

rising the number of iterations at the same time. Again no significant influence of the emissivity ratio could be found.



**Fig. 5.** Simulated emissivity as a function of a true emissivity ratio and an assumed measurement uncertainty of the pyrometer of 25 K and 100 K. The real object's emissivities that were used to obtain measurement data are circled in red.

Ratio pyrometers can be implemented in different ways. One may be the use of two actually individual detectors. They may be either arranged at different positions or stacked one behind the other (“sandwich detector”). The latter may lead to an undesired crosstalk behaviour [19]. Further, there may be only one detector that is filtered at different wavelengths which will automatically lead to a temporal mismatch between the channels. However, due to a finite sampling time a similar mismatch may likewise occur even if different detectors are used. During the simulation such a mismatch has been accounted for by the measurement uncertainty which acts randomly on the temperature measurement of both channels. Moving objects, which are a common application of pyrometry are accounted for by a random change in temperature of a total of 200 K. If larger changes are to be expected such as it would be the case for active heating and cooling the simulation time may exceed a critical change in temperature thus increasing the simulation error.

For practical use, it is important to point out that a real detector cannot operate at a single wavelength. On the other hand, this detector only gives integral values of the thermal radiance within its spectral range. This excludes an analytical determination of the temperature  $T$  and asks for a secondary simulation to find an  $M(T)$  which converges towards the measured value  $M(T)$ . In the end, a balance needs to be found between simulation time and accuracy.

## 4 Summary

The present work has presented an algorithm to determine an object's emissivity without any knowledge on the material properties. The sole prerequisite is that both "detectors" – may this be physically different devices or one detector with different filters – view the exact same position on the object's surface such that one temperature can be assumed for all channels at a time. In this case the emissivities at both spectral positions  $\varepsilon_1$  and  $\varepsilon_2$  and the true temperature  $T$  form a threefold equation system that – once the correct combination of  $\varepsilon_1$  and  $\varepsilon_2$  has been found – returns a true statement.

A dependence on the emissivity ratio has not been found and resulting temperature errors have not shown a significant influence of the nominal value of the true temperature although an increase in standard deviation for rising emissivities can be seen. The latter is, however, in good accordance with Planck's Law of Thermal Radiation.

Though it could be seen that larger wavelengths result in a higher uncertainty whether the true emissivity had been found in the simulation as predicted by Planck's Law the means have not shown to differ significantly. However, raising the pyrometer's measurement uncertainty from 25 K to 100 K was found to increase the simulation error by almost an order of magnitude.

Considering real measurement conditions the present paper has proven the concept of evaluating an object's true emissivity by a numerical solution of the equation system that results from the working principle of a ratio pyrometer.

## List of Nomenclature

$M_{bb}^\lambda$	thermal radiance of a black body at a detection wavelength $\lambda$
$h$	Planck constant
$c$	speed of light
$\lambda$	detection wavelength
$k_B$	Boltzmann constant
$T$	temperature
$\alpha$	absorptivity
$\alpha_{s,d}(\lambda,\gamma)_T$	spectral and directional absorptivity as a function of detection wavelength $\lambda$ and angle $\gamma$ at a temperature $T$
$\varepsilon$	emissivity
$\varepsilon_{s,d}(\lambda,\gamma)_T$	spectral and directional absorptivity as a function of detection wavelength $\lambda$ and angle $\gamma$ at a temperature $T$
$F_{Ratio}^{1,2}$	ratio of the real-body spectral radiances $M_1(T)$ and $M_2(T)$

## References

1. Planck, M.: Über eine Verbesserung der Wien'schen Strahlungsgleichung. Verhandlungen der Deutschen Physikalischen Gesellschaft (1900)
2. Kirchoff, G.: Über den Zusammenhang zwischen Emission und Absorption von Licht und Wärme. Monatsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, 783 (1859)
3. Gibson, A. F.: A two-colour infra-red radiation pyrometer. Journal of Scientific Instruments 28(5), 153 (1951)
4. Araújo, A.: Multi-spectral pyrometry – a review. Journal of Measurement Science and Technology 28, 082002 (2017) DOI: <https://doi.org/10.1088/1361-6501/aa7b4b>
5. Sun, X. G.: Processing Method of Multi-Wavelength PyrometerData for Continuous Temperature Measurements. International Journal of Thermophysics 26(4), 1255 (2005) DOI: 0.1007/s10765-005-6724-6
6. Xing, J. et al.: A data processing algorithm for multi-wavelength pyrometry-which does not need to assume the emissivity model in advance. Journal of the International Measurement Confederation 67, 92 (2015) DOI: <https://doi.org/10.1016/j.measurement.2014.11.019>
7. Wen, C.-D., Chai, T.-Y.: Examination of multispectral radiation thermometry using linear and log-linear emissivity models for aluminum alloys. Journal of Heat and Mass Transfer 47, 847 (2011) DOI: <https://doi.org/10.1007/s00231-011-0769-9>
8. Araújo, A., Silva, R.: Surface temperature estimation in determined multi-wavelength pyrometry systems. Review of Scientific Instruments 91, 054901 (2020) DOI: <https://doi.org/10.1063/5.0005676>
9. Reichel, D.: Temperature Measurement in Rapid Thermal Processing with focus on the application to Flash Lamp Annealing. Critical Reviews in Solid State and Material Sciences 36(2), 102-128 (2011) DOI: <https://doi.org/10.1080/10408436.2011.572757>
10. Nulman, J. et al.: Pyrometric Emissivity Measurements and Compensation in an RTP Chamber. In: MRS Proceedings, pp. 146 – 461. Cambridge University Press, Cambridge (2011) DOI: 10.1557/PROC-146-461
11. Sentenac, T.: Trichromatic thermoreflectometry for an improved accuracy of true temperature field measurement on a multi-material part. International Journal of Thermal Sciences 145, 105980 (2019) DOI: <https://doi.org/10.1016/j.ijthermalsci.2019.105980>
12. Fu, T.: Measurements of the directional spectral emissivity based on a radiation heating source with alternating spectral distributions. International Journal of Heat and Mass Transfer 90, 1207 (2015) DOI: <https://doi.org/10.1016/j.ijheatmasstransfer.2015.07.064>
13. Madura, H., Piatkowski, T.: Emissivity compensation algorithms in double-band pyrometry. Journal of Infrared Physics and Technology 46(1-2), 185-189 (2004) DOI: <https://doi.org/10.1016/j.infrared.2004.03.024>
14. Müller, B., Renz, U.: Development of a fast fiber-optic two-color pyrometer for the temperature measurement of surfaces with varying emissivities. Review of Scientific Instruments 72, 3366 (2001) DOI: <https://doi.org/10.1063/1.1384448>

15. Muller, M. Fabbro, R.: Temperature measurement of laser heated metals in highly oxidizing environment using 2D single-band and spectral pyrometry. *Journal of Laser Applications* 24(2), 022006 (2012)  
DOI: <https://doi.org/10.2351/1.3701400>
16. Roebuck, B., Edwards, G., Gee, M. G.: Characterisation of oxidising metal surfaces with a two colour pyrometer. *Journal of Materials Science and Technology* 21(7), 831 (2013)  
DOI: <https://doi.org/10.1179/174328405X46169>
17. Coates, P.B.: Multi-Wavelength Pyrometry. *Metrologia* 17(3), 103 (1981)  
DOI: <https://doi.org/10.1088/0026-1394/17/3/006>
18. Madura, H., Kastek, M., Piatkowski, T.: Automatic compensation of emissivity in three-wavelength pyrometers. *Journal of Infrared Physics and Technology* 51(1), 1-8 (2007)  
DOI: <https://doi.org/10.1016/j.infrared.2006.11.001>
19. Luryi, S., Xu, J., Zaslavsky, A.: *Future Trends in Microelectronics: Up the Nano Creek*. 1<sup>st</sup> edn. John Wiley & Sons, Inc., Hoboken New Jersey (2007)

# Detecting Sarcasm in News Headlines

Onyinye Chudi-Iwueze<sup>1</sup> and Haithem Afli<sup>2</sup>

ADAPT Centre,  
Cork Institute of Technology, Cork, Ireland  
<sup>1</sup>[o.chudi-iwueze@mycit.ie](mailto:o.chudi-iwueze@mycit.ie), <sup>2</sup>[haithem.afli@cit.ie](mailto:haithem.afli@cit.ie)

**Abstract.** The use of sarcasm dates as far back as communication goes. Sarcasm is expressed in words, facial expression and can be noticed in the intonation of voice. In this digital age, sarcastic comments are passed everyday, via tweets, comment sections of different social media outlets, and even in news headlines. Some of these headlines can be misunderstood and taken to mean something different than the original intentions. This leads to a need to detect sarcasm especially in the news and on social media. Detection of sarcasm from text has its challenges because text lacks intonation and deflection of voice that occurs when a sarcastic statement is made vocally by a human. This paper concentrates on the effect of different feature encoding techniques applied to text for feature extraction for machine learning models. A deep learning model is also applied and the results are compared. Prior to feature extraction, data pre-processing techniques like tokenization, removal of stop-words and punctuation are applied by researchers in any work involving text analysis. These pre-processing techniques are widely used and accepted and are also applied in this project. Different feature extraction methods like Count Vectorizer, Term Frequency-Inverse Document Frequency and word embedding were implemented in this experiment. The Support Vector Machine, Naive Bayes and Logistic Regression were the traditional machine learning algorithms used in this research. Convolutional Neural Network was also used. Results of these algorithms are recorded in this paper and these include the F1-score, precision, recall and the accuracy scores. The results of these different algorithms for the methods of feature extraction examined in this paper show that a combination of more than one technique is better suited for the purpose of classifications. The results of this paper show

**Keywords:** Sarcasm Detection · Convolutional Neural Networks (CNN) · Feature Extraction · Machine Learning.

## 1 Introduction

In the last decade, there has been an increase in the use of online resources for dissemination of information. These texts have different characteristics that are explored by the use of Natural Language Processing techniques of machine learning for knowledge and insight. One of such important characteristics is the presence of sarcasm in text. Sarcasm is a complex act of communication that

allows speakers the opportunity to express sentiment-rich opinions in an implicit way [1]. There is an abundance of sarcasm in texts distributed online and on social media. This includes news headlines and other media. Sarcasm is classically defined as the process of intentional misuse of words to convey some meaning (usually the opposite of what is said) [1]. For example, if a person says ‘*How lucky am I? I caught the corona virus!*’ It is clear that although the words used are positive, the real meaning is negative, making the statement a sarcastic one. Sarcasm is characterised by the use of irony that reflects a negative meaning.

Although sarcasm is well known and widely used, it is challenging not only for computers but also for humans to detect promptly. Some humans find it difficult to identify and understand the use of sarcasm [2]. Due to this fact, the presence of sarcasm, if not detected and accounted for can affect other machine learning ventures like sentiment analysis and opinion mining. This makes the detection of sarcasm a crucial task. Automated sarcasm detection can be seen as a text classification problem. Text data is one of the simplest forms of data. Machine learning algorithms are unable to process non-numerical data and as such the need for feature extraction arises. It is important to extract useful information from any forms of data, especially unstructured forms like text data. Feature extraction and the proper representation of text for classification purposes is an important factor that affects the accuracy of classification. This paper explores the use of the Count Vectorizer, Word Level TF-IDF, Character Level TF-IDF and N-Grams Level TF-IDF on different supervised learning algorithms and the accuracy of these models is measured and compared. The Word2Vec is also an efficient feature extraction technique which is widely used. The Word2Vec is used for the Deep Learning Algorithm explored in this paper. Other methods like Doc2Vec and LDA are not discussed in this paper as they are outside the scope of this project.

The rest of this paper is arranged as follows: Section 2 goes over some related work both in sarcasm detection and feature extraction methods. Section 3 details the methodology and which models were applied for the analysis. Section 4 outlines the results of the application of the different feature extraction techniques to the chosen models and describes which performed best.

## 2 Related Work

### 2.1 Feature Extraction

Text extraction describes the process of taking words from text data and transforming them into a numerical feature set that is useful for a machine learning classifier [13]. Extracting features is useful for all analyses involving text data in all domains. In the work detailed in [8] the impact of pre-processing is discussed and reviewed. The researchers determine the importance of ‘slang’ and correct spelling. SVM classifier is applied for their experiment. Another researcher uses vector representations to provide the solution to the problem of

sentiment analysis [9], obtaining an accuracy of (86%) which is relatively high. Another study considers the use of four data sets and the the use of Bag-Of-Words Features,lexicon and part of speech based features [10]. An ensemble model of SVM, Logistic Regression and Naive Bayes was implemented for the analysis. The authors in [11] applied three levels of feature extraction techniques. They also applied three classifiers in their analysis. Another group of researchers in the work detailed in [12] applied ten different feature extraction techniques to a sentiment analysis problem. They concluded that feature extraction techniques applied for a problem have the potential to improve the performance of a classifier in any given problem. Paper [7] compares the results of 6 algorithms when features are extracted using Bag-Of-Words and TF-IDF. They conclude that the TF-IDF gives better results of up to 4% difference. Much of the research experiments are done with the SVM, Logistic Regression and Naive Bayes algorithms for the traditional supervised machine applications. This informed the use of these algorithms for this analysis. A CNN model is also run at the end of the analysis using the Count Vectorizer vector as the features and also using features generated with word2vec.

## 2.2 Sarcasm Detection

The detection of sarcasm is very crucial in the domain of sentiment analysis and opinion mining. Different machine learning algorithms have been applied to the problem. Some researchers used the Naive Bayes Classifier and Support Vector Machines for analyses of social media data in Indonesia [5]. The classifiers used in the work detailed in [5] performed well for the task and that informed the decision to use Naive Bayes and the SVM for this analysis. Another body of work that applied traditional supervised techniques to the problem is described in [7]. They used support vector machines with Tf-IDF and Bag-Of-Words. The work by Davidov et.al [6] details a technique of sarcasm detection using semi-supervised methods on two different data set, one containing the reviews on Amazons products and one with tweets collected from Twitter. These researchers concentrated on features like punctuation, sentence syntax, hashtags used etc. The proposed system had an accuracy of over 75% [6]. The work by [14] shows the use of Support Vector Machines also for the detection of sarcasm from news headlines. The optimal method of feature extraction used in this work results in an accuracy score of about 80%. Some other researchers applied deep learning techniques to the problem [2]. Gathering data for the purpose of sarcasm detection is challenging and even more so for supervised learning. This is especially the case for social media data. This research uses some of the methods described in the previous works with better results for accuracy and other measures as well.

### 3 Methodology

#### 3.1 Data set

The data used for this analysis was taken from Kaggle. The news headlines were curated into two different files for competitions on Kaggle. Each file was a JSON file, containing news headlines and the label named 'is\_sarcastic' that indicates the presence or absence of sarcasm in that headline. The two files were joined together to create a bigger data set for the analysis. The complete data used in the analysis contains over 50000 headlines. The data set is not significantly imbalanced but there were a few more non-sarcastic news headlines than the sarcastic headlines. Table 1 below shows a few rows in the data and the labels for those data points.

	Headline	Is_sarcastic
1	former versace store clerk sues over secret 'black code' for minority shoppers	0
2	why writers must plan to be surprised	0
3	boehner just wants wife to listen, not come up with alternative debt-reduction ideas	1
4	remembrance is the beginning of the task	0
5	4 lessons prison taught me about power and control	0
6	top snake handler leaves sinking huckabee campaign	1
7	courtroom sketch artist has clear manga influences	1
8	stock analysts confused, frightened by boar market	1
9	gillian jacobs on what it's like to kiss adam brody	0
10	diy: sports equipment closet	0

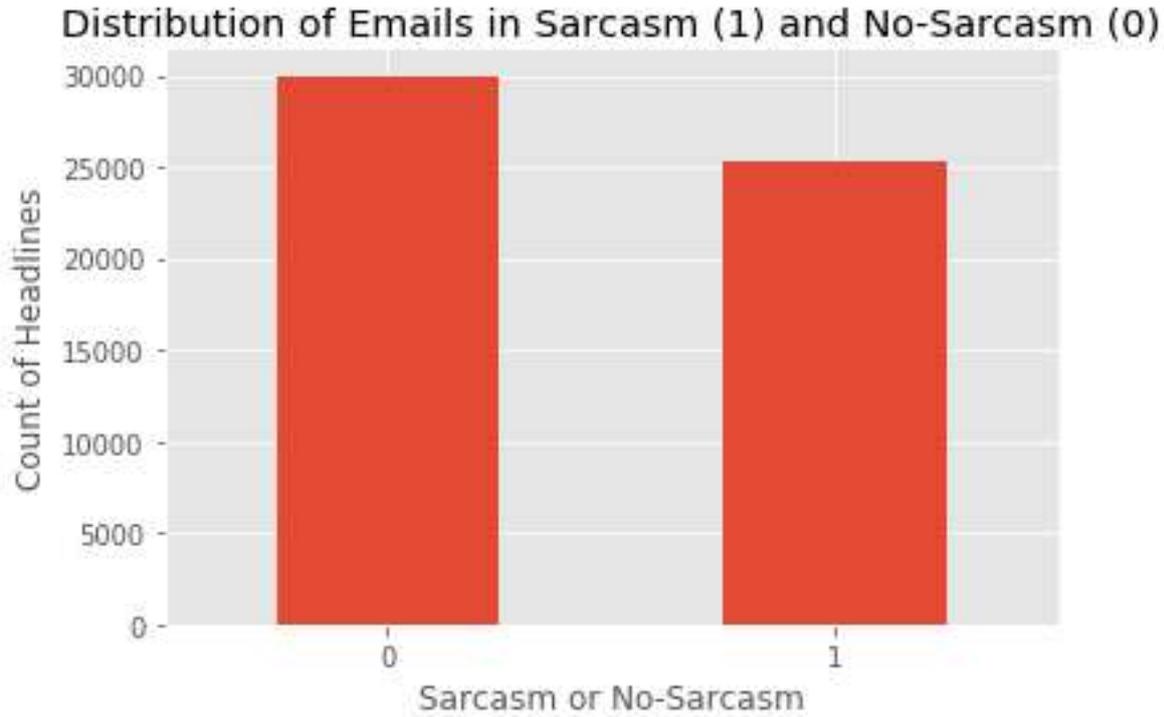
**Table 1.** Cross section of some headlines in the data set

Table 2 below shows the distribution of headlines and tokens among the train and test sets.

Sarcasm Data Set	
Data	Number of Tokens
Train	Sarcastic: 19010
	Non-Sarcastic: 22486
Test	Sarcastic: 6348
	Non-Sarcastic: 7484

**Table 2.** Table showing distribution of headlines between test and train

Figure 1 below shows the distribution of headlines in the data set. There are more non-sarcastic emails but the classes are not grossly imbalanced.



**Fig. 1.** Distribution of Headlines in the Data set.

### 3.2 Data Pre-processing

The first step in the pre-processing pipeline was the conversion of all the text to lower case. This is done to achieve uniformity of words. Unwanted symbols like digits and newlines were removed also. All punctuation were also removed from the data set. After these, the texts were tokenized. Tokenization is the breaking down of sentences into words [7]. Stopwords are words that are commonly used in any language. They do not contribute any contextual meaning to the data and as such should be eliminated. Removal of stopwords was the last step in the pre-processing pipeline and the data was thus ready for further preparation prior to model building.

### 3.3 Feature Extraction

**Count Vectorizer** The Count Vectorizer is a simple means used to tokenize and build vocabulary of known words. It is also used to encode new documents using the vocabulary that has been created. The product of the count vectorizer is an encoded vector which has the length of the entire vocabulary of the document and a count for the number of times each word appeared in the document. Count Vectorizer was implemented using scikit-learn for this project.

**Term Frequency – Inverse Document Frequency (TF-IDF)** TF-IDF is a popular and well recognized method for the evaluation of the importance of

a word in a document. Term Frequency measures the frequency of a term in a document. For a variety of documents with different lengths, the probability of a word occurring more times in a longer document is present. Due to this fact, normalization is required.

$$TF(t) = \frac{N}{T}$$

where  $N$  = Number of times term  $t$  appears in a document  
and  $TN$  = Total number of terms in the document

The Inverse Document Frequency is the measure of importance of a term. This is done to reduce the influence of words e.g 'of', 'the' that could be used multiple times in a document but have no real meaning and importance in context.

$$IDF(t) = \log_e \frac{TD}{ND}$$

where  $TD$  = Total number of documents  
and  $ND$  = Number of documents with term  $t$  in it

For most traditional applications the bag-of-words model was used. However, words that occurred frequently in the document were given significant weights in the Bag-OF-Words model and this affected the accuracy of these results. The use of TF-IDF was introduced to solve these problems of the typical Bag-Of-Words Model. The IDF of an infrequent term due to the log applied is high and the IDF for common words is low [13]. Three levels of TF-IDF were used for this analysis. They are the Word-level, N-Gram and Character Level TF-IDF. All these are implemented using the scikit-learn package of python.

**Word2Vec** Word embeddings are vector representations of words and word2Vec is one of the most popular methods of word embedding. The word2Vec method was first proposed in the work of Mikolov et al [15] as a method of feature extraction for Natural Language Processing (NLP). It is an improvement on the traditional methods of feature extraction. The construction of word embedding using neural networks and can be done using either the Common Bag of Words or Skip Gram. According to [15], the Skip Gram method performs better for smaller data set. The word2Vec vectors applied in this experiment are trained on the words available in the data set.

### 3.4 Classification algorithms

A couple of classification algorithms were chosen for this experiment. These algorithms are briefly described below:

**Support Vector Machines** The Support Vector Machine is efficient for both classification and regression. It is known to give good results for classification. The classes are separated by a hyper plane found by the algorithm. The LinearSVC from scikit-learn was implemented for this analysis.

**Naive Bayes** The Naive Bayes is a powerful algorithm used for classifying data based on probabilities [7]. This algorithm is based on the Bayes theorem in statistics. The classification of the data is done using various probabilities. It is fast and scalable but is not without disadvantages as it assumes independence among predictors. It works well for small data sets and has been known to give great results. The Naive Bayes Classifier employed in all stages of the analysis are built using the scikit-learn package available in python.

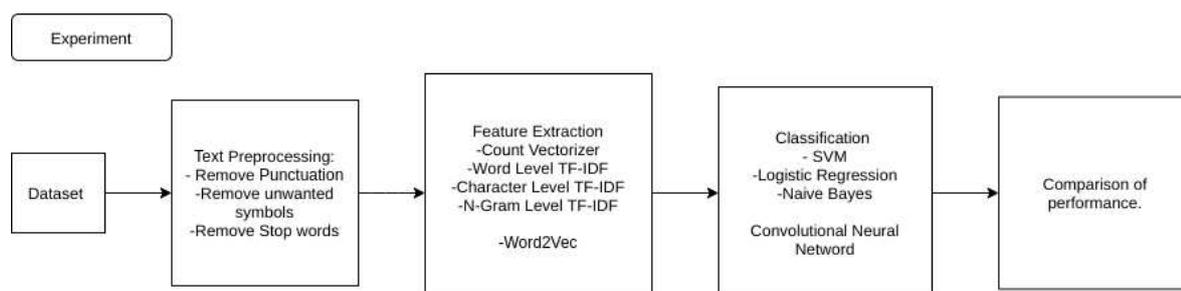
**Logistic Regression** Logistic Regression is a popular classification algorithm. It belongs to the class of Generalized Linear Models. Its loss function is the sigmoid function which minimizes the results to be a value between 0 and 1. The Logistic Regression Model was implemented for this analysis using the function available in the scikit-learn package.

**Convolutional Neural Networks (CNN)** CNNs are very commonly used in text classification problems due to their success and great results. This has contributed to the decision to use CNN for this aspect of the experiment. In the first experiment with CNN, one CNN model is used on the features extracted using the Count Vectorizer. Another CNN model is also trained on the features extracted using Word2Vec. The results of these two models are plotted and compared.

## 4 Experiments and Results

### 4.1 Experiments

Figure 2 below is a pictorial representation of the processes involved in this experiment. The first step taken was the amalgamation of the two data sets obtained from Kaggle. Text pre-processing techniques were applied to the data as described above. The main body was an iterative process involving the application of one type of feature extraction technique accompanied with the building of the model and recording the results. Cross Fold validation was applied using the GridSearch function of python's sklearn for the final results for the supervised learning models. The results of these experiments are recorded and reviewed in the section that follows.



**Fig. 2.** Experiment Structure

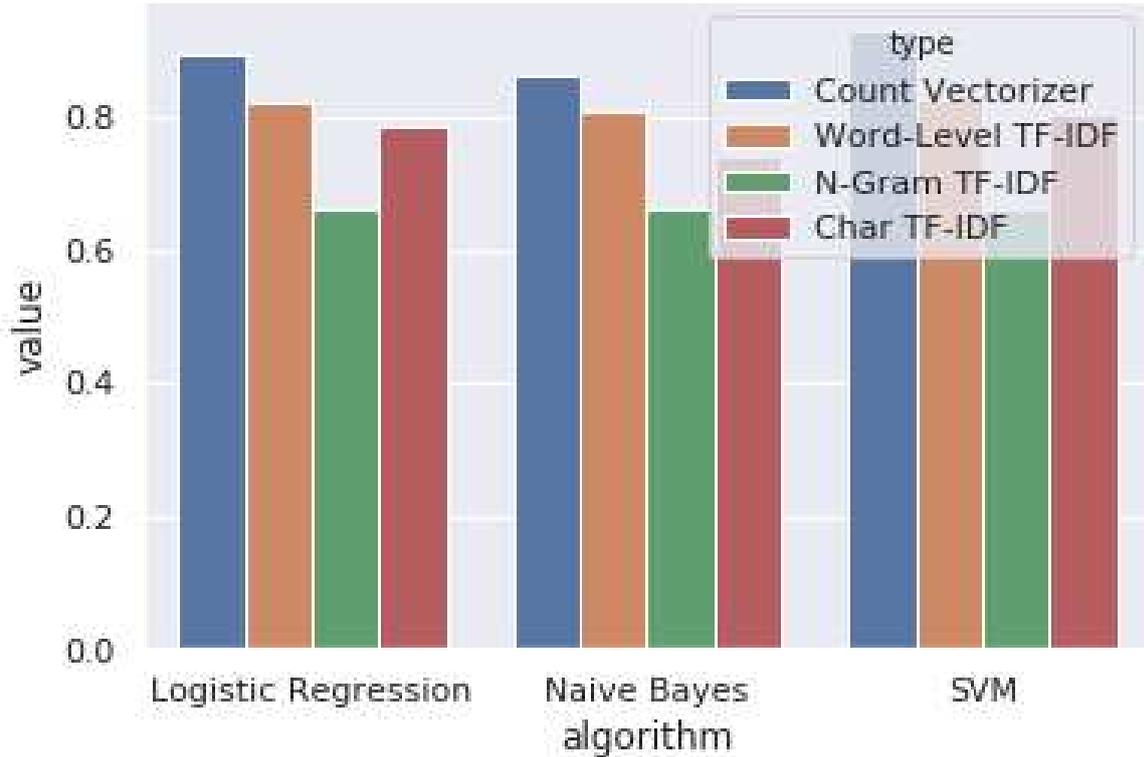
## 4.2 Results

Table 3 contains the results of the experiments. The metrics of measurement for each algorithm and the accompanying scores are detailed in that table.

	F1 Score	Accuracy	ROC AUC Score
<b>Count Vectorizer</b>			
Linear SVM	0.94 for 0 0.92 for 1	0.93	0.927127083461732
Logistic Regression	0.90 for 0 0.88 for 1	0.89	0.891574615638756
Naive Bayes	0.87 for 0 0.85 for 1	0.86	0.862782631933632
<b>Word Level TF-IDF</b>			
Linear SVM	0.84 for 0 0.81 for 1	0.83	0.826282163974597
Logistic Regression	0.84 for 0 0.80 for 1	0.82	0.802490934661872
Naive Bayes	0.83 for 0 0.78 for 1	0.81	0.81771997863453
<b>Character Level TF-IDF</b>			
Linear SVM	0.82 for 0 0.78 for 1	0.8	0.800651976895386
Logistic Regression	0.81 for 0 0.76 for 1	0.79	0.7817870730821
Naive Bayes	0.77 for 0 0.70 for 1	0.74	0.736702613127708
<b>N-Gram Level TF-IDF</b>			
Linear SVM	0.75 for 0 0.49 for 1	0.66	0.638295324080576
Logistic Regression	0.75 for 0 0.49 for 1	0.66	0.63666163513879
Naive Bayes	0.75 for 0 0.49 for 1	0.66	0.637254498317267

**Table 3.** Table showing results of supervised learning

For the supervised learning techniques, the results to be examined are the results of the SVM, Naive Bayes and Logistic Regression when text features are extracted using the Count Vectorizer technique and different levels of the TF-IDF technique. These results are detailed in Table 3 above and can be seen graphically in 3 below.



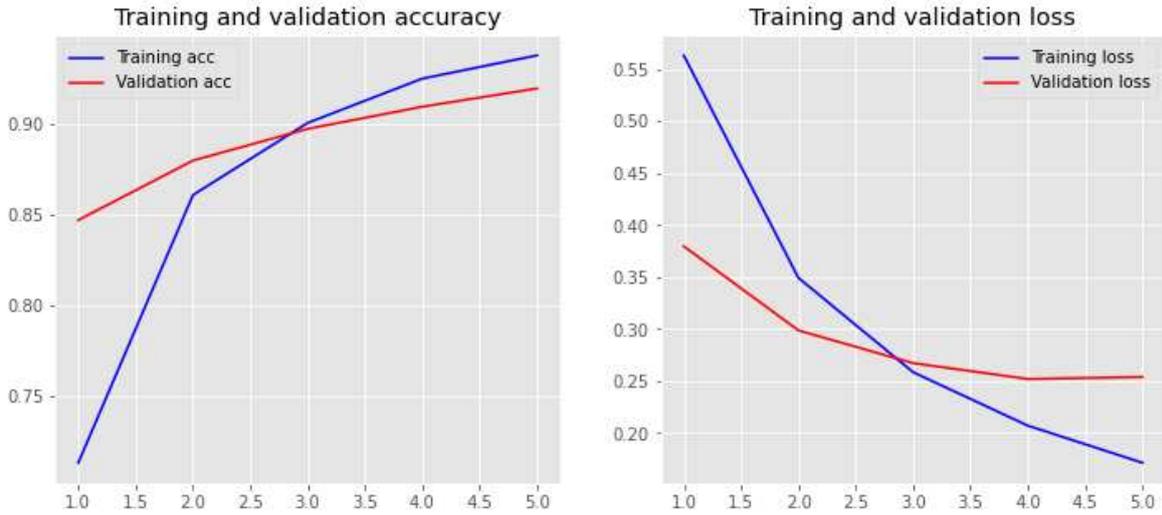
**Fig. 3.** Supervised Learning Results Diagram.

The bar plot that follows in Figure 3 is pictorial representation of the results of the analysis performed using the supervised learning methodologies. It is clear that the SVM outperforms both the Logistic Regression and the Naive Bayes in all the cases, except the N-Gram level TF-IDF where they all perform similarly poor. The figures that follow concentrate mainly on the results of the SVM algorithm due to this fact.

The results of the work done in our research shows that the use of Count Vectorizer and TF-IDF gives an accuracy of 93% when this feature set is fed to SVM. This is an improvement on the work of [14] as the methods used here provide a higher accuracy all round.

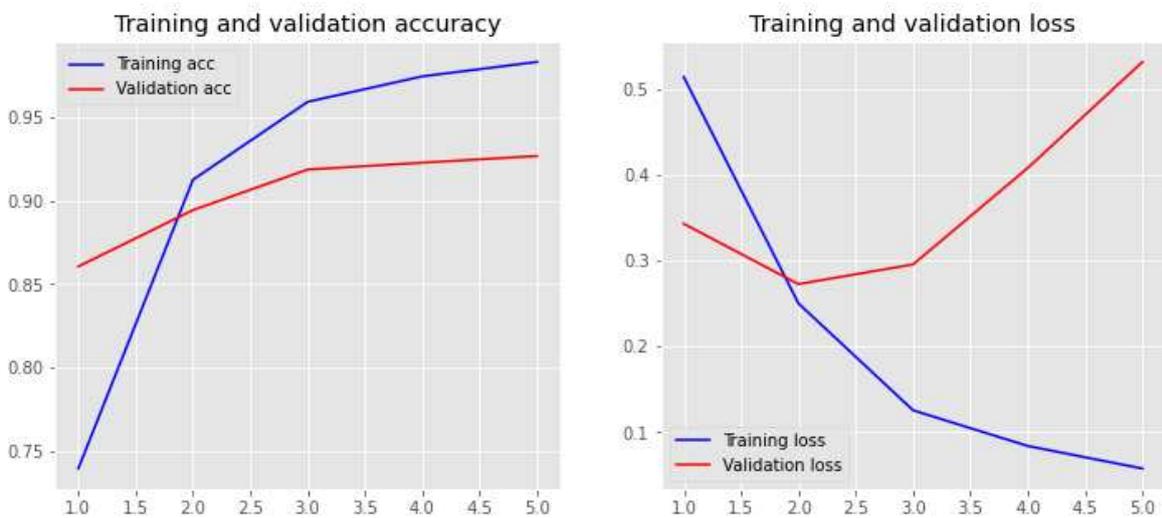
The results as seen above corroborate the known fact that Support Vector Machine usually outperforms other algorithms for classification purposes, depending on the measure of data available. This is clear in the measure of the accuracy as seen in the bar plots in Figure 3. The accuracy of the SVM with the Count Vectorizer is about 90%. The results for the n-gram level TF-IDF are the worst of all. This is similar to the Bag-Of-Words model and as such the bad results are expected.

Figure 4 below shows the training and validation accuracy scores for the model trained over 5 epochs. Unsurprisingly, the accuracy of training for the CNN is very close to 100%. The validation accuracy however is about 92.5%. The loss also reduces for both the training and validation data.



**Fig. 4.** CNN Results with features extracted using the Count Vectorizer

The training and validation accuracy score is higher when word embedding is applied to the CNN model. However, the Figure in 5 shows an interesting pattern for the validation loss when word2vec is applied with the CNN. The loss reduces and then begins to increase again for the data used in validation. This suggests that the model does not actually perform as expected when word embedding is applied. This is due to the fact that word2vec is a neural network based solution and these work optimally when the volume of data is substantial.



**Fig. 5.** CNN Results with features extracted using the Word Embeddings

## 5 Conclusion

Sarcasm can be found in a wide variety of areas. To be able to accurately detect sarcasm from different aspects and topics, a large enough data set will be needed. The accuracy of prediction as seen in this paper greatly depends on the mode of feature extraction. The use of N-grams and characters produced less desirable results than the use of words. The combination of the Count Vectorizer and the Term Frequency-Inverse Document Frequency for the supervised learning techniques gave the most satisfactory performance metrics.

This results of this research favor the supervised machine methods over the deep learning methods. Although this could be due to the limited amount of data available. This goes to show that the use of simple methods can produce great results, especially in the situation that the data available is limited and not suitable for more complicated deep learning techniques. The CNN performed satisfactorily also. The use of word embedding for this task showed great performance in terms of accuracy, both in the training and validation sets so does the use of the features extracted using the count vectorizer. The overall performance of both the CNN and supervised learning methods suggests that the use of Count Vectorizer is most appropriate for this task. As stated above, the volume of data available limits the strengths of the Word2Vec method, this the Count Vectorizer outperforms in this situation. This is due to the fact that simpler methods are always a better choice than more complicated methods provided that the results are similar or better with the simpler methods.

Sarcasm can also be multi-modal, contained in images and GIFs. This is especially available on social media like Twitter where tweets can be replied with GIFs and videos. Future research would involve an extension of this system to include multi-modal methods for sarcasm detection. Social media data for sarcasm detection will also contain emoticons (emojis) and analysis of those emojis could add more meaning to the text. A bigger data set will allow for the use of feature extraction techniques that require large data sets. This paper has not explored the use of word embedding for both supervised and deep learning techniques. More exploration into these methods of feature extraction available is warranted and would be a focus of future research in this area.

## References

1. Liu, P., Chen, W., Ou, G., Wang, T., Yang, D., Lei, K.: Sarcasm Detection in Social Media Based on Imbalanced Classification In: Web-Age Information Management, pp. 459-471. Cham, (2014). [https://doi.org/10.1007/978-3-319-08010-9\\_49](https://doi.org/10.1007/978-3-319-08010-9_49)
2. Mehndiratta, P., Sachdeva, S., Soni, D. :Detection of Sarcasm in Text Data using Deep Convolutional Neural Networks In: Scientific International Journal for Parallel and Distributed Computing, vol. 18, September (2017) <https://doi.org/10.12694/scpe.v18i3.1302>
3. Zhibo, W. , Ma, L., and Zhang, Y. : A Hybrid Document Feature Extraction Method Using Latent Dirichlet Allocation and Word2Vec. In : IEEE First International Conference on Data Science in Cyberspace, June (2016). <https://doi.org/10.1109/DSC.2016.110>

4. Zhang, W., Xu, W., Chen G., Guo, J. : A Feature Extraction Method Based on Word Embedding for Word Similarity Computing. In: Communications in Computer and Information Science, vol. 496, pp. 160-167 January, (2014) [https://doi.org/10.1007/978-3-662-45924-9\\_15](https://doi.org/10.1007/978-3-662-45924-9_15)
5. Lunando, E., Purwarianti, A. : Indonesian Social Media Sentiment Analysis With Sarcasm Detection. In: Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSYS, pp. 195-198. September (2013). <https://doi.org/10.1109/ICACSYS.2013.6761575>
6. Davidov, D., Tsur, O., Rappoport, A. : Semi-Supervised Recognition of Sarcasm in Twitter and Amazon. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pp.107–116, July (2010) <https://doi.org/10.5555/1870568.1870582>
7. Ahuja, R., Chug, A., Kohli, S., Gupta, S., Ahuja, P. : The Impact of Features Extraction on the Sentiment Analysis In: Procedia Computer Science, vol. 152, pp. 341-348. Elsevier, January (2019). <https://doi.org/10.1016/j.procs.2019.05.008>
8. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau R.:Sentiment Analysis of Twitter Data In: Proceedings of the Workshop on Language in Social Media, pp. 30-38, June (2011). <https://doi.org/10.5555/2021109.2021114>
9. Fan, X., Li, X.,F. Du, F., Li, X., Wei, M.:Apply word vectors for sentiment analysis of APP reviews. In: 3rd International Conference on Systems and Informatics(ICSAI), pp. 1062-1066, (2016)
10. Fouad, M.M., Gharib, T.F., Mashat, A.S: Efficient Twitter Sentiment Analysis System with Feature Selection and Classifier Ensemble. In: The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018), vol. 723, pp. 516-527, Springer, Cham January (2018). [https://doi.org/10.1007/978-3-319-74690-6\\_51](https://doi.org/10.1007/978-3-319-74690-6_51)
11. Angulakshmi, G., Manicka Chezian, R.: Three level feature extraction for sentiment classification. In: International Journal of Innovative Research in Computer and Communication Engineering 2, vol. 8, pp. 5501-5507 (2014).
12. Prusa, Joseph, D., Taghi, M.K, and David J.D. :Impact of Feature Selection Techniques for Tweet Sentiment Classification. In: FLAIRS Conference 2015, pp. 299-304, April (2015)
13. Waykole, R.N, Thakare, A.D: A Review Of Feature Extraction Methods For Text Classification. In: International Journal of Advance Engineering and Research Development, vol. 5, April (2018).
14. Vaishvi, P. J.: Optimal Feature Extraction based Machine Learning Approach for Sarcasm Type Detection in News Headlines. In: International Journal of Computer Applications, vol. 177, March (2020).
15. Mikolov, T., Chen, K., Corrado, G., Dean, J.:Efficient Estimation of Word Representations in Vector Space.

# Chatbot based Behaviour Analysis for Obesity Support Platform <sup>\*</sup>

Yanxin Wu<sup>1</sup>, Ryan Donovan<sup>1</sup>, Binh Vu<sup>2</sup>, Felix Engel<sup>2</sup>, Matthias Hemmje<sup>3</sup>, and Haithem Afli<sup>1</sup>

<sup>1</sup> ADAPT Centre, Cork Institute of Technology, Cork, Ireland

<sup>2</sup> FTK e.V. Forschungsinstitut für Telekommunikation und Kooperation,  
Wandweg 3, 44149 Dortmund, Germany

<sup>3</sup> GLOBIT GmbH, Barsbüttel, Germany.

yanxin.wu@mycit.ie

ryan.donovan@adaptcentre.ie

binhvu@noemail.org

felix.engel@fernuni-hagen.de

matthias.hemmje@fernuni-hagen.de

Haithem.Afli@cit.ie

**Abstract.** The challenge created by the health problem of rising obesity rates requires novel approaches to solve. This paper presents an innovative approach in the form of a conversational platform aimed at helping users deal with health issues associated with obesity. The chatbot platform enables collection of personal data from users to be analysed via natural language processing and behavioural analysis to provide tailored solutions for each user based on their current states and psychological traits. The gathered and analysed data is accessible. This platform developed using Microservices architecture and chatbot technology. User can interact with the chatbot to generate personal chat data stored in the platform. The collected chat data will be used for natural language processing and behaviour analysis, along with other available data, to create a customized user model. The gathered and analysed data is accessible and usable for Health Care professionals as a mechanism to encourage healthier nutrition, and the user can benefit from the platform by getting feedback and support on their methods for improving their eating and physical activity habits by way of the chatbot.

**Keywords:** Chatbot · Microservices · emotional states · physiological activity · Behaviour analysis.

## 1 Introduction

The ongoing development of mobile technology has coincided with yearly increases in smartphone and social media usage [1] [2] [3]. The user data generated from smartphone and social media use can be collected and utilised as big

---

<sup>\*</sup> Supported by Horizon 2020 project STOP Obesity Platform.

user data-sets. Industry and research professionals have utilised such data-sets to improve the user experience of their applications or as a source of collaboration with other research parties to solve large-scale social issues, such as those seen in the public health domain.

One of the rising public health problems is Obesity. Obesity is defined here as an abnormal accumulation of body fat that is excessive, related to Body Mass Index (BMI) score greater than 30 [25]. The amount of people obese in the world has almost tripled since 1975. If this rate of increasing obesity is unabated, then almost half of the world's adult population will be overweight or obese by 2030 [5]. Obesity is not only a problem for WEIRD (Western Educated Industrialised Rich and Democratic) nations, as the percentage of people with obesity is projected to increase in 44 countries by 2015 [6]. It has proven to a problem that strikes early and remains difficult to overcome, with 80% of the adolescents suffering from obesity becoming adults suffering from obesity.

Obesity [26] has a devastating impact on public health, as it is a major risk factor for chronic diseases such as hypertension, cardiovascular, coronary heart diseases, type-2 diabetes, and certain types of cancers. Obesity is a contributor to difficult and debilitating psychological issues, as people with obesity suffer disproportional from affect disorders such as depression and anxiety [27]. It contributed to 5% of worldwide deaths in 2014. Within the EU, in most member states this problem is increasing at a rapid rate: from 2010 to 2016 the percentage of the population that is obese rose from 20.8% to 23.3% [7]. The negative impact obesity has on public health has prompted EU driven research and industry collaborations to develop innovative approaches to solve this problem.

In the market of smart health, there has been some innovative approaches developed. Some noteworthy approaches are:

- *iBitz* - An app that is aimed at children and has an active connection with a fitness tracker named *iBitz Kids Pedometer*, which tracks users footsteps and activity levels. In-app rewards are allocated to provide incentive feedback on activity levels, representing a gamification approach. The more activities the children do, the more points they gain. The points are converted result in rewards customised based on parents preferences.
- *Habitca* - An IOS and Android app. This app aids users in habit building and productivity through a gamified approach; the app allocates rewards and punishments to provide motivation for user, harnesses social networks to allow users to share their progress, and simplifies this process through a intuitive and minimal interface.
- *My Diet Coach* - This app provides a virtual coach that assists the user in their the daily diet. The app provides pictures to represent the “old” user and the “new” user in order to motivate users.

While these approaches are innovative and practical, none offer an interoperable overall support platform. Without such a platform, users are overloaded with incompatible ICT applications, making the difficult process of losing and managing weight more difficult. Additionally these applications fail to employ interoperability with state-of-the-art fitness sensing hardware as well as cutting

edge user interface paradigms (e.g. conversational chatbots). Besides this, a lot of actions exist in the market which could benefit from ICT support in many forms. This ICT support would become interoperable, if a unified platform would exist.

The paper structured as follow: section 2 detailed describe the approach of using Chatbot platform and Behaviour analysis. Section 3 shows the implementation of the platform and Behaviour analysis experiment plan. Section 4 presents the usage of the chatbot platform in the ongoing STOP project [16]. Section 5 shows some potential future work of the chatbot. Section 6 lists the support and founders.

## 2 Chatbot and Behaviour Analysis Approach

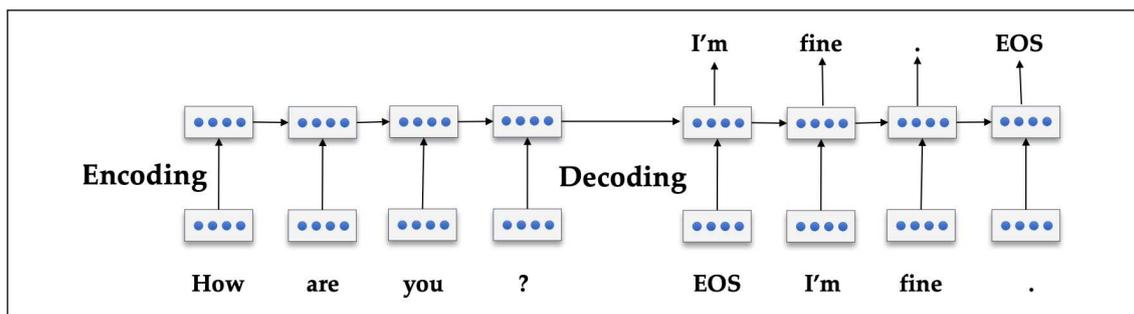
There is an overabundance of blanket style approaches tackling public health problems, where the same intervention is applied across all individuals. This is despite the fact there exists large variation in the personality characteristics and emotional experiences, both of which can significantly moderate the effectiveness of public health interventions. It is essential that each person, in respect to their idiosyncratic characteristics and needs, is treated appropriately with user-driven work plans. This paper introduces a chat-bot based approach that can meaningfully and intelligently engage with platform users, extracts information about user's characteristics and needs, and enables a more personalised and more effective health intervention.

Chatbots are systems that engage in extended conversations with the goal of mimicking the unstructured conversational or 'chats' characteristic of human-human interaction. Chat-bot systems have been used for practical purposes, such as testing theories of psychological counselling [9] and E-health applications. There are several chatbot-based application in the market.

- *Youper* - Youper [18] was created by a team of scientists and doctors. The app focuses on improving user emotional health with the personalized conversation. Youper user can view the emotion record and set different goals to achieve.
- *OneRemission* - OneRemission [19] is an app aimed at cancer patients and their loved ones. "It aims at making the lives of cancer survivors, fighters, and supporters easier, safer, and more enjoyable.". The app can offer valuable information database that based on the experts' knowledge to improve user states on both physically and mentally.
- *Your.MD* - Your.MD [20] is an app base on the idea of self-care. The application can provide a big database that user can use to self-check the specific symptom.
- *Babylon* - Babylon health [21] application offers chatbot-based symptom self-check, real doctor video/audio check, health monitor. Babylon health company also provide real GP subscription plan.

- *Sensely* - Sensely [22] is a virtual assistant app that user can use for: symptoms checking, receiving self-care information, scheduling clinician appointment and locate the nearest pharmacies. Users have several ways to interact with the app: a chatbot associated with text-to-speech and speech recognition technologies; a virtual character system that use to improve user experience.

Traditionally, chatbots are typically rule-based. As recent as 2014, Siri and Google Now still relied on handcrafted rules to find the most relevant answers. But deep learning techniques and the availability of more user generated datasets and powerful computers have opened up new possibilities with corpus-based models. These models mine large datasets of human-human conversations, which can be done by using information retrieval (IR-based systems simply copy a human’s response from a previous conversation) or by using a machine translation paradigm such as neural network sequence-to-sequence with word embeddings and attention mechanism as showed in Fig. 1, to learn to map from a user utterance to a system response [8]. There are three main components in these



**Fig. 1.** A sequence to sequence model for neural response generation in dialog. [8]

models:

- *Embedding* – Embedding can be of type word or other forms of tokens such as characters or n-grams. The embedding layer is converting the input into a vector of continuous numbers representing the input.
- *Encoder* – At this stage we are encoding the input embeddings (the vectors) to produce intermediate states which are fixed length vectors.
- *Decoder* – The decoder takes the fixed length encodings produced by the encoder and generate a variable length sentence using beam search decoding.

A number of modifications are required to adapt the basic sequence to sequence model implemented initially for Machine translation to the task of chat generation. As mentioned in [10] these models tend to produce repetitive responses like “ Sorry, I can’t help” or “I don’t know” that can end the conversation. This problem can be technically addressed by changing the training model objective function to a mutual information objective, or by modifying a beam decoder to keep more diverse responses in the beam.

## 2.1 Emotions and Personality

The capability of the chat-bot to provide insights into the affective and general psychological state of its users is crucial to develop. In order to perform a feasibility study on the utility of the chatbot to enable behavioural analysis, the important psychological phenomena and their interconnections needs to be defined and clarified. This sub-section defines these psychological phenomena, their interrelation, and what signals can be used for the chatbot system to accurately detect these phenomena.

**Basic Emotions** This chatbot platform is intended to provide analysis of the emotional states of its users. Emotions are considered in terms of typical behavioural and physiological patterns along with the subjective experience of the particular emotion. In the context of this work, the working definition of emotion, is the subjective experiencing of that particular emotion (e.g. the typical subjective experience of Anger, Disgust, Joy, Fear, Surprise, and Sadness) coupled with physiological and behavioural activity simultaneously occurring with it. This fits the Basic Emotion Theory (BET) perspective [28].

BET state that there exists a sub-set of emotions with highly consistent and reliable behavioural and physiological patterns [13]. This consistency in basic emotion signals is culturally independent and thus makes them excellent candidates for scientific research. While there is some level of disagreement on which emotions should be considered basic, this research classifies the basic emotions as: Fear, Surprise, Sadness, Joy, Anger, and Disgust. These emotions are labelled as basic not because they are simple, but because they are deeply rooted in sub-cortical areas of the brain and are the emotions that have been the most influenced by evolution. The more cognitively mediated emotions emerge from the basic emotions (e.g. guilt is seen as a mixture of sadness and disgust). The basic emotions form the basis of all emotional experiences.

However, whilst the basic emotions have been shown to correlate consistently with physiological activity, this has not been fully investigated and been used for practical purposes. Prior research has shown that the basic emotions are associated with distinct patterns of cardiorespiratory activity [14]. For the purposes of the chatbot platform, our aim is to establish the patterns between the basic emotions and such activity, in order for our application to have another measure of assessing the emotional state of the user. This could be used in cases where ambiguity is present, such that the words people use can have a positive or negative connotation. Such information then can be used as a way to assess such ambiguous cases whilst also providing confirmation for non-ambiguous cases.

**Detecting Personality Through Emotions** One of the unique contributions this chatbot can provide is harnessing emotion detection in order to build a model of the user's personality. Personality is defined here as the unique way a person feels, perceives, and behaves in the world. An accurate understanding of personality enables one to understand what makes people the same, what makes

people similar to others, and what makes people individually unique [29]. The value of personality analysis is in provides ‘in-context’ information about the person and their psychological state at any point and avoids blanket one-size-fit-approaches.

Emotions are regarded as the ”prime movers” of personality [10]. How and with what intensity we experience emotions is a key component to the uniqueness of our personality [17]. For example, studies that asked participants to describe the personality of another person showed that a high number of participant’s descriptors referred to the person’s typical emotional experience (e.g. he/she is a happy person; he/she is an anxious person). Yet the exact nature of the relationship between emotions and personality has been difficult to ascertain due to the fact that emotion research initially faced similar challenges to personality; in that it previously relied exclusively on subjective self-report.

However, one solution to this problem has been to focus on the behavioural and physiological patterns that coincide with experience of each emotion [11]. These patterns tend to be central-nervous system (CNS) activity facial expressions, type of language and pitch of voice used, behavioural expressions, amongst others. For example, the emotion anger has been shown to have a particular CNS activity [12]; facial expressions tend to be lowered eyebrows, tightly pressed lips, and bulging of the eyes; language tends to be direct and in a higher than normal pitch; people can either be physically aggressive or display signals that suggest readiness for aggressive behaviours (e.g. clenched fist). These patterns serve as signals to the subjective experience of the emotion anger.

The identification of these components enables innovative methodologies for detecting personality that can be harnessed by this chatbot platform. Thus far, no known research has been carried investigating the links between cardiorespiratory activity and personality traits, making this a novel research endeavour. The emotions will then mediate the ability of the chatbot through behavioural (semantic) and physiological data to detect personality traits of the user, which can then be used to model particular and more productive.

However, the data sources of the platform are various, from smartphones to smart wearable devices. All the devices have their own operating system and different sensors that are used to collect the user activity data. Beside using cardiorespiratory data, other user behaviour data or sensor data also can be used to associate personality traits analysing. For example, mobile phones have internal GPS to locate the user location. Though analysing location data, some specific location will be marked. Combination of location results and cardiorespiratory data will review certain habits of the user. By using behaviour data, we can build user profiles that support analyze models. This enables behavioural analysis at a more macro-level of the person.

## 3 Implementation

### 3.1 Data Usage

Based on the requirement of the chatbot platform, data to be collected and used in this work include physical activity data collected by wearable sensors such as Fitbit wrist bands, smart watches and smart mobile phones; nutrition information provided retailers and self-reporting; physiology information, such as BMI, heart rate, blood pressures by measurement; and other self-reporting data, such as physical activities that are not recorded by sensors, feedbacks.

The physical activity data shall include types of activity (such as walking, running, swimming); duration of activity; levels of activities (such as steps, speed). These can be recorded by wearable sensors (walking steps, running speed) and self-reporting (such as dance, swimming). Nutrition information will be collected by grocery calories identification via a mobile app, meal calories estimation and self reporting.

Feedback provided to users shall be friendly, easy to understand, reliable and respect the behavioural changes on the user. A conversational Chatbot trained on behaviour analysis will be able to provide a good vehicle to achieve this goal.

### 3.2 Platform Implementation

When implementing this chatbot-based platform, there are several technologies been used. One is Microservice architecture(MA). Microservice architecture is a more detailed structural style than service-oriented architecture (SOA). Same as SOA, services in MA are connected through the network. However, not like services in SOA, services in MA been broken down into many micro-services. Each microservice will do a small amount of work and "group together" to perform the whole functionality of the service. By doing the broken-down process, the overall structure becomes more loosely coupled which gain more readability and maintainability.

Another technology has also been used when implementing the chatbot platform is containerization. Containerization is a technology that allows developers to wrap the application along with the specified configuration files, libraries and required dependencies together into a container which can run in any computing environment. Because a container has all the requirements needed, it will run its own without interfering with the host environment. This characteristic of the containerization makes it works well when associated with microservice architecture. There are several companies doing containerization in the industry, the one used in chatbot platform is Docker container [15]. Compare with other containers, docker containers are easier to build and more lightweight.

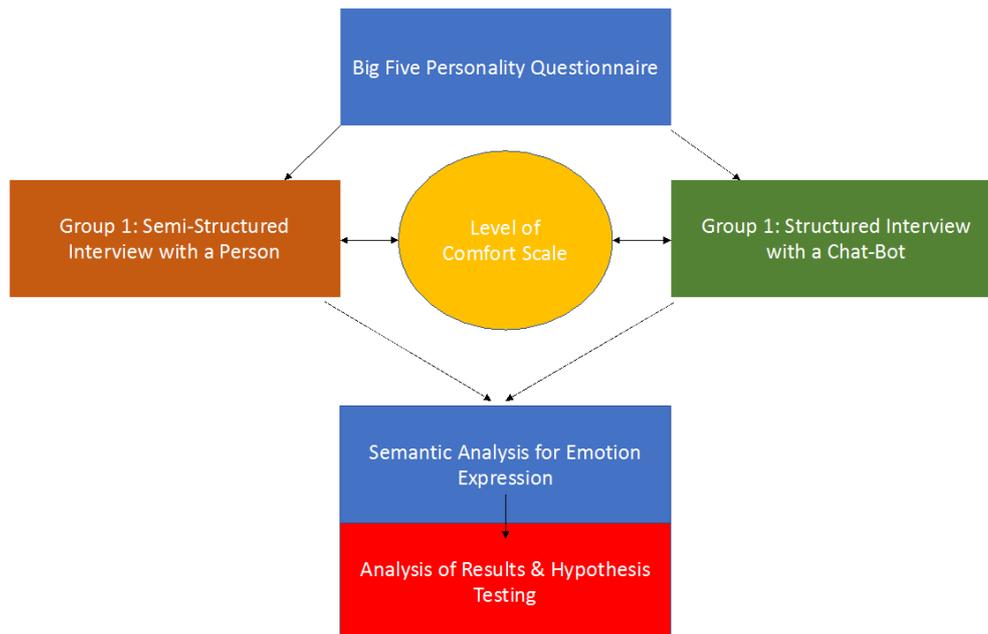
The benefit of using MA and docker container is that combining these two technologies will help the platform to face future challenges. Because the platform is in microservice architecture so different type of chatbot can be added at any point without stop and update the whole platform. Using containers also allow chatbot to been developed in different languages since the container can be running cross-platform.

### 3.3 Behaviour Analysis Plan of Action

An experimental study will be conducted. The preliminary schedule for the study is as follows:

- Step 1: Agreement on questions to be used. Finalisation of experimental materials.
- Step 2: Identification and recruitment of research participants.
- Step 3: Run Time of the Experiment.
- Step 4: Analysis of the Results.

Participants will be those who suffer from health issue. Participation in the study will be completely voluntary. Participants will be informed about the chatbot platform and the expectations of how this work can benefit those who suffer the health issue. In order to test the above, participants who suffer from



**Fig. 2.** Chatbot Experiment Procedure

obesity will be recruited in a two-group study. Both groups will also answer the Big Five Aspects Scale (BFAS). People will be separated into the following groups:

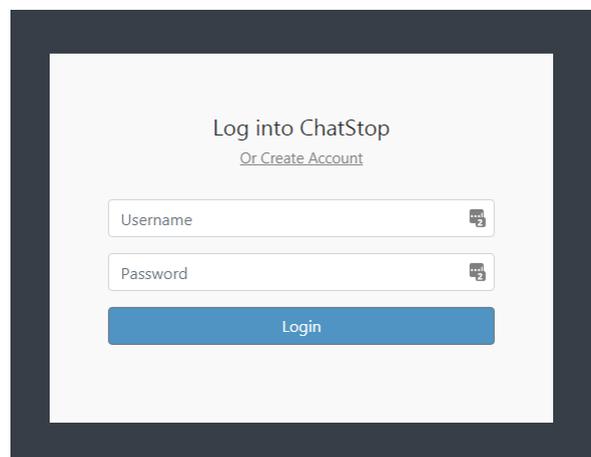
- One group will have participants that engage in conversation with a person using a set of questions about physical/emotional experiences.
- One group will have participants that engage in conversation with a chat-bot like app using a set of questions about physical/emotional experiences

The interview will be conducted in a semi-structured manner. The research interviewer/chat-bot will ask the participants each question at a time and follow

it categorically. However, in cases where participants will have more to say on any particular topic due to their own unique life experience or opinions, then there is room for exploration into these areas and related follow up questions. The questions are therefore a guide rather than a strict manual. However, both groups will have the same set of potential questions.

### 3.4 User Interface Design

The design of the chatbot platform mainly includes two parts: the web front-end that users can access, the different chatbot that user can talk to. The front-end is used to access the chatbot and other functions and has two parts: the UI and the data storage. The UI is accessible for all users that allow each user to create an account to use the website(see Fig. 3 and Fig. 4).

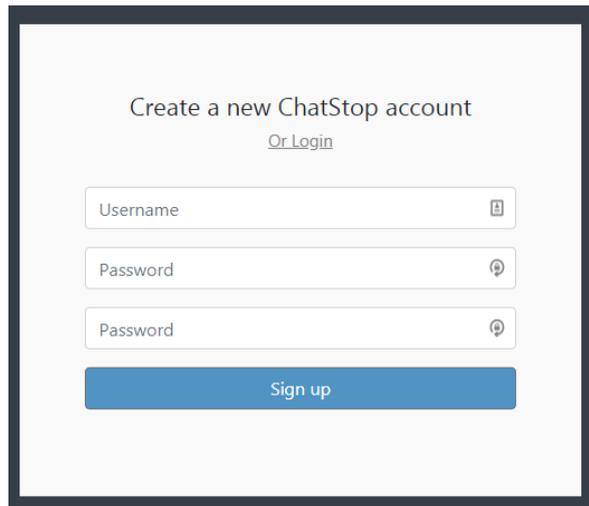


**Fig. 3.** Log in Page

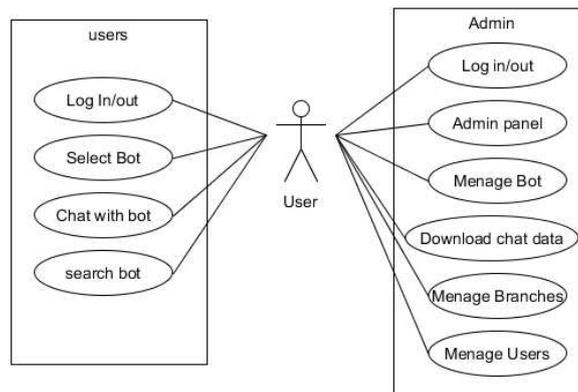
There are two types of user for the website: Common user and Admin user. Common users are the one who will use the chatbot and can use the chatbot only. Admin user can access to more function than common user such as add new chatbots, manage existing chatbots, download conversation history from chatbots and rule management of all users(see Fig. 5).

According to the data protection policy and website management, the Admin panel option will not appear if the common user log in(see Fig. 6).

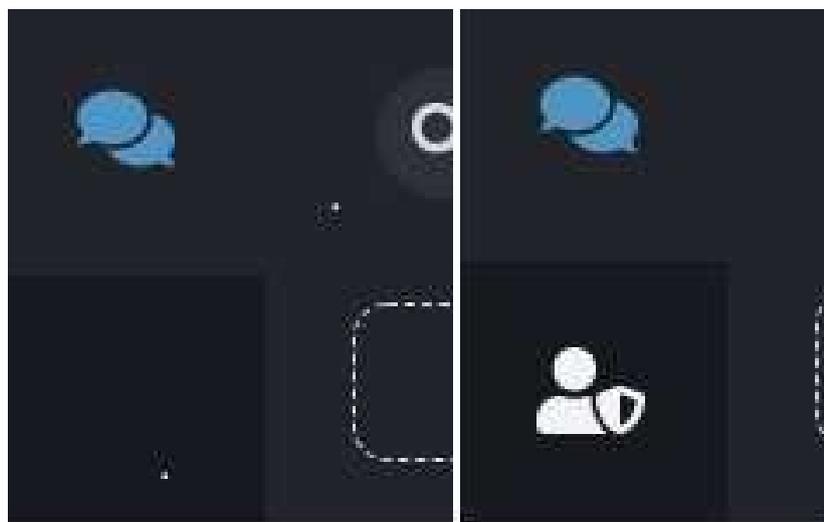
After the admin user logged in, the Admin panel can be used(see Fig. 7).



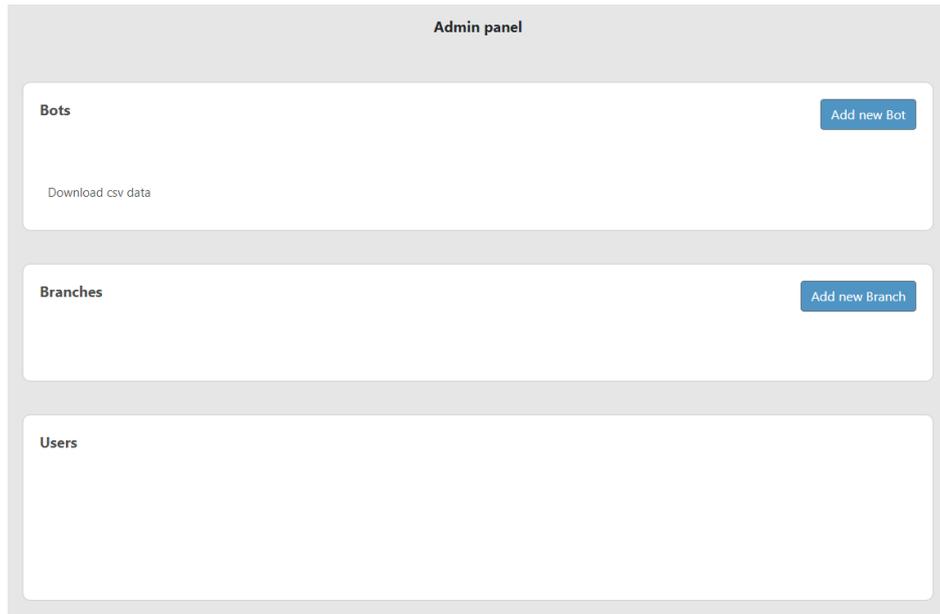
**Fig. 4.** Sign up Page



**Fig. 5.** User Type



**Fig. 6.** Panel Option



**Fig. 7.** Admin Panel

The website data storage is separately deployed as one microservice of the platform which allows UI to connect through the specified API. All the data from the website like user information, bot information and chatbot conversation been stored in the database. The chatbot will be developed separately in different platforms. In order to add the bot to the platform, the bot will be wrapped into a docker container

## 4 Application

STop Obesity Platform (STOP) is a 36-months project funded by the European Union (EU). The project aims to support persons with obesity with a better nutrition under supervision of healthcare professionals. In the STOP project, user health and activities are monitored by different smart sensors and wearable devices, such as e.g. smart watch, fitness tracker. Captured data is stored, enriched, then semantically fused to enable an unified and unique data access interface. The outcome is used as inputs for sophisticated AI data analysis in the STOP Ecosystem Platform.

In STOP, the Chatbot is trained on user's physiology information, physical activity data, and nutrition information. These inputs are tailored for each user to provide a friendly, easy to understand, reliable feedback. This helps to change user behaviour toward a healthy life style with more exercises and good nutrition.

User fitness data is monitored by their smart phone and wearable devices. Data is stored on the phone by the fitness app and uploaded to the manufacture server. Depend on the manufacture, user data can be accessed by third party applications through Web APIs or Software Development Kit (SDK) for mobile apps. Based on the Wrapper-Mediator-Architecture, the STOP Ecosystem

Platform can gather user fitness data from popular services, such as e.g. Fitbit and Google Fit, and stores it in the Fitness database. Furthermore, the platform provides a REST API interface for mobile apps to submit user data.

With the REST API interface, the Chatbot can access and modify user's fitness data on their behalf.

## 5 Future Work

The chatbot platform is being developed in the context of STOP, an EU-funded project aims to support persons with obesity with a better nutrition under supervision of healthcare professionals [16]. In this project, each user has different fitness measures, as well as, health issues. Therefore, they need supports based on their current status. The chatbot platform, in this case, can not be trained on general input data but needs to be tailored for each individual. The outcome can be used to help each user to change their behaviour toward a healthy life style with more exercises and good nutrition. Furthermore, the chatbot can be integrated into popular messaging platforms such as e.g. Facebook Messenger and WhatsApp using their Development APIs [23] [24]. This will enable users to communicate with the chatbot anytime from their familiar apps and the chatbot to reach a larger audience on these platforms. Finally, the work of developing different type of chatbots and models to adapt the future changes in the market can also be considered.

## 6 Acknowledgments

This research was conducted with the financial support of the Horizon 2020 project STOP Obesity Platform under Grant Agreement No. 823978 and at the ADAPT SFI Research Centre at Cork Institute Of Technology. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant 13/RC/2106.

## References

1. Statista, <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>. Last accessed 21 June 2020
2. Statista, <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>. Last accessed 21 June 2020
3. Statista, <https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/>. Last accessed 21 June 2020
4. European Commission, [https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu\\_en/](https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en/). Last accessed 21 June 2020
5. Tremmel, M., Gerdtham, U.G., Nilsson, P.M. and Saha, S.: Economic burden of obesity: a systematic literature review. *International journal of environmental research and public health* **14**(4),435(2017)

6. Pineda, E., Sanchez-Romero, L.M., Brown, M., Jaccard, A., Jewell, J., Galea, G., Webber, L., Breda, J.: Forecasting future trends in obesity across Europe: the value of improving surveillance. *Obesity facts* **11**(5),360-371(2018)
7. World Health Organization.: European health report 2018: more than numbers—evidence for all. World Health Organization, Regional Office for Europe. (2018)
8. Daniel Jurafsky & James H. Martin. *Speech and Language Processing* (3rd ed. draft), <https://web.stanford.edu/~jurafsky/slp3/>, Last accessed 21 June 2020
9. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM* **9**(1), 36-45 (1966)
10. Davis, K.L.,Panksepp, J.: *The emotional foundations of personality: A neurobiological and evolutionary approach*. WW Norton & Company (2018)
11. Ekman, P.: Basic emotions. *Handbook of cognition and emotion* , 45-60(1999)
12. Kreibig, S.D.: Autonomic nervous system activity in emotion: A review. *Biological psychology* **84**(3),394-421(2010)
13. Ekman, P., Cordaro, D.: What is meant by calling emotions basic. *Emotion review* **3**(4),364-370(2011)
14. Rainville, P., Bechara, A., Naqvi, N., Damasio,A.R.: Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International journal of psychophysiology* **61**(1),5–18(2006)
15. Docker, <https://www.docker.com/>. Last accessed 21 June 2020
16. CORDIS, <https://cordis.europa.eu/project/id/823978/>. Last accessed 21 June 2020
17. R. Donovan, A. Johnson, A. deRoiste, and R. O’Reilly: *Quantifying the Links between Personality Sub-Traits and the Basic Emotions*. presented at the International Conference on Computational Science and Its Applications, 2020
18. Youper, <https://www.youper.ai/>. Last accessed 21 June 2020
19. OneRemission, <https://keenethics.com/project-one-remission/>. Last accessed 21 June 2020
20. Your.MD, <https://www.your.md/about/>. Last accessed 21 June 2020
21. Babylon, <https://www.babylonhealth.com/product/>. Last accessed 21 June 2020
22. Sensely, <https://www.sensely.com/>. Last accessed 21 June 2020
23. Docker, <https://developers.facebook.com/docs/messenger-platform/>. Last accessed 05 July 2020
24. Docker, <https://www.whatsapp.com/business/api>. Last accessed 05 July 2020
25. WHO, [https://www.who.int/health-topics/obesity#tab=tab\\_1](https://www.who.int/health-topics/obesity#tab=tab_1). Last accessed 21 June 2020
26. Kissebah,A.H., Freedman,D.S., Peiris,A.N. : Health risks of obesity. *Medical Clinics of North America*,**73**(1), 111-138 (1989)
27. Gariepy, G., Nitka, D., Schmitz, N. : The association between obesity and anxiety disorders in the population: a systematic review and meta-analysis. *International journal of obesity*,**34**(3), 407-419 (2010)
28. Ekman, P: Basic emotions. *Handbook of cognition and emotion*, 45-60 (1999)
29. Little, B: *Me, myself, and us: The science of personality and the art of well-being*. Public Affairs, (2014)

# Online News Analysis on Cloud Computing Platform for Market Prediction\*

Claudia Juarez<sup>1</sup> and Haithem Afli<sup>2</sup>

ADAPT Centre,  
Cork Institute of Technology, Cork, Ireland  
<sup>1</sup>juarez.moreno@gmail.com,<sup>2</sup>haithem.afli@cit.ie

**Abstract.** Stock market price fluctuations and predictions have been widely examined; there are two approaches for analysis, fundamental analysis (data from financial records and balance sheets) and technical analysis (focused on past market action). However, past trends cannot predict stock market movement alone; external factors significantly influence them, a notable example being how Twitter comments by Elon Musk affected the stock price for Tesla and the subsequent follow-up by the SEC with the corresponding sanctions. In this highly interconnected society and 24 hours news cycles, we require an extra tool to study the stock market. This research evaluates news articles from financial publications and determines the word patterns that will help make a buy or sell decision, by identifying a combination of words or phrases that indicate if a stock price might go up or down. This research also focuses on the creation of a blueprint for the implementation using cloud technologies to house the financial information, perform the analysis and then present it as a web app for more comfortable use and interpretation. A novice trader could benefit from a simple indicator that provides information on whether the stock might go up or down, as it would facilitate the decision-making process by identifying the exact day for buying or selling, or in case there is no relevant news, then hold the position.

**Keywords:** Natural Language · Cloud Deployment · Custom Corpora · Deep Learning.

## 1 Introduction

### 1.1 Motivation

Stock market analysis is a field that has been studied from several angles, such as economics, finance and statistics, to name a few. Typically for an experimented broker, the breakdown of a stock behavior will be formed by two strategies, the technical analysis, which uses mathematical models to predict the stock variations in the stock charts, and the fundamental strategy.

---

\* Supported by Cork Institute of Technology.

The fundamental analysis comprehends the breakdown of a company's financial statements, historical data, and the interpretation of the summary provided by specialized websites [14]. It also involves plenty of reviews of news, financial articles, and awareness of the company's media presence as new developments or products might bring the stock price up, whereas a news scandal might affect the price and make it go down (albeit temporarily).

The main focus of this research paper and the contribution to the current work is twofold. First, we developed the code for the stock behavior analysis based on the news articles which the end-user will have the ability to select and customize according to their needs.

The second contribution concentrates on cloud architecture development; this paper provides a blueprint for implementing the behavior analysis code in an app that can be accessed over the internet using Python, machine learning, and a commercial cloud platform.

This paper differentiates from past research with these new approaches:

1. It has more extensive data sets, as we created our corpus using financial news articles and stock market opening and closing prices.
2. Our model uses natural language and deep learning as it works best with unstructured data.
3. We provide a practical way for the trader to interact with the predictions through a web app, where they have the ability to choose the source of data collection
4. Simultaneously, the algorithm retrains itself each time the user changes source data with an automated data collection process.

## 1.2 Contribution

This paper takes away part of the complexity of stock analysis for the average investor, by creating an easy to use web app that will take advantage of machine learning and cloud computing to deliver pertinent information such as a prediction of stock behavior. This tool will help any investor who might not have a financial background but an eagerness to manage their portfolio. It can also help the part-time stock investor to avoid contacting a career advisor and spending a percentage of the profit on these services.

## 2 Background

### 2.1 Natural Language Processing Research

Due to this subject's popularity, there has been a myriad of research papers that tested the link between media and stock price movement. Yaojun and Wan have tried to tie stock price fluctuations to social media (although restricted to Chinese social media) and tied it to a 'sentiment analysis,' this means labeling words as positive or negative; even if the result was successful, they only considered Chinese social media accounts that provided financial information,

thus having a minimal information source [15]. Nareen found in his study that sentiment directly influences a stock price; however, it did not quantify or offer a prediction [13]. Joshi's study also found a high correlation between price fluctuation and sentiment however lacked a broad set of testing data; they based their research only on historical data from yahoo.com for just one company from news aggregators [8].

Deep learning is widely used with Python as a tool that helps a program learn patterns based on data [4]. This project works based on a correlation between 'words as vectors' natural language processing and sentiment analysis to start mapping keywords/phrases and comparing them to previous stock market historical prices; it also implements word to vector (word2vec) technique to help us train a program in human language nuances such as ambiguities and grammar, while sentiment analysis provides a positive or negative connotation to text and phrases [9].

## 2.2 Existing Methods

We have taken as starting point three existing research papers and built upon them; Joshi's and Nareen's studies found a correlation between stock movement and sentiment analysis. Yaojun and Wan created a base lexicon that labeled words as positive or negative to relate stock fluctuations to social media posts. We took both ideas and methodologies and created a custom corpora that measured positive and negative in terms of stock market fluctuation and instead of social media news article headlines.

## 2.3 Cloud Computing

There have been some examples of machine learning models successfully implemented in cloud technologies, although not directly related to stock predictions. There has been evidence that utilizing cloud computing technologies provides a significant reduction of execution time for requests from stakeholders by maximizing the utilization of cloud resources [3].

One such case of analysis is in the healthcare industry. Although it can be applied to any industry, the primary tools Abdelaziz uses are machine learning algorithms deployed in virtual machines. Abdelaziz performed predictions using linear regression and neural networks, finding a considerable improvement in response times of other models using these models in conjunction with cloud capabilities [3].

Currently, there are several tools for analyzing vast quantities of text; one of them is Python NLTK (Natural Language Toolkit). It can help us extract text excerpts and detect patterns, word structure, and frequency to determine meaning and intention[6]. As Feyzkhanov states in his book, serverless deep learning deployment is a novel approach that has the advantages of being scalable, simple, and cheap to start [7]. Most of the academic work does not focus on providing a design for cloud deployment.

## 2.4 Contribution

We determined that it is possible to predict with accuracy the movement of the stock market based solely on news articles taking advantage of cloud computing architecture.

# 3 Natural Language Processing Implementation

## 3.1 Methodology

We used model manipulation and local improvement as our heuristic methodology, we changed the nature of the deep learning model to apply it to our particular test case scenario by doing a trial and experiment using different testing data sets; we also created localized improvements by starting with a feasible solution on a working deep learning model and constructing and improving iteratively upon it.[11]

For this research paper, we did not use a standardized sentiment lexicon as we are not looking for a sentiment such as ‘positive’ in the general sense of the word, but ‘positive’ as it is looking for an impact on the stock market. Instead, we created and customized for this particular research necessities’ a custom corpora using different sources and combining them to obtain the data that is relevant for our research. [11]

We analyzed the headers of various news articles from online news outlets based on a stock ticker and supplied a prediction for said stock price movement. The Natural Language Processing or NLP model is housed in cloud-based architecture and made available to the public for ease of use in a web app. We created a custom corpus that worked with the NLP prediction model forgoing using a pre-trained model.

We chose this approach because, upon closer inspection of the individual texts, they contained a lot of ‘noise’ such as advertisement, excerpts of other pieces, links, or author’s bios, which were not associated with the article on hand. Headlines of articles are a synthesis of the entire sentiment of the article.

## 3.2 Data Creation

For the creation of the data set, we used an API to retrieve the publication name, headline, source article and date of publication from Google News and then paired it with the stock movement for that particular day; if the news fell on a weekend or a holiday, we moved to the next business day.

We calculated the percentage of change for that day, either up or down on price, and since these numbers varied widely, we normalized on a scale from 0 to 1, we did this operation for every set of stock ticker and news article.

The main two discoveries we made after creating this custom corpus were that the data set was broad enough to provide information on global markets alongside the search of a single news resource and that the paid API provides more online publications even after just indicating 5 to search through.

### 3.3 Prediction Models

We selected this model because the Keras/TensorFlow Model can be used for fast prototyping and combined with TensorFlow is optimal for a production-ready system; the downside is that we do not have much control over the code as it is best for fast experimentation. We are working with neural networks that cannot work on raw data text, so the next step is to convert it to tokens, which are just integers. [12]

Tokenize; The tokenizing method works by going through the complete dataset and counting the number of times each word is used and then making a vocabulary where each word gets an index; this way our data samples are converted to numbers called tokens [12]

Pad-Truncate; The sequences are padded/truncated to make sure they are all of a standardized size by taking the average number of words in all the sequences and add two standard deviations to ensure we are aiming at 95% of the data; the padded works by adding zeros to the sequence. [12]

Embedding; Even after converting the words to tokens, a neural network still cannot work on this data due to vocabulary limitations and semantics; so, we use embedding; it works by converting integer tokens into real-valued vectors. This technique is known as representation learning, which is a way to obtain a real-valued representation from a text while preserving the semantics. [12]

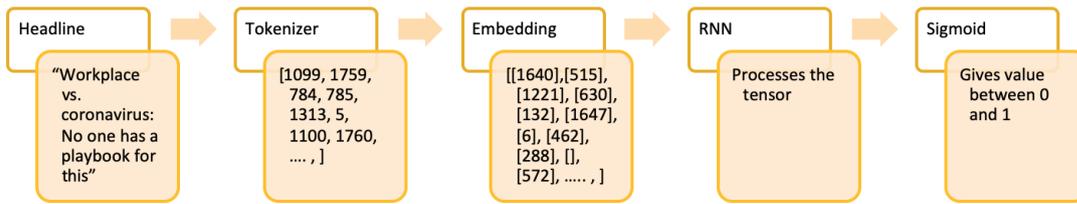
Recurrent Neural Network (RNN); The result of the embedding process is a two-dimensional matrix called a tensor, which now is in the correct format to be used in an RNN; its main characteristic is that it can process sequences of arbitrary length. Each layer is dependent on the result of the layer before it, which makes it perfect for natural language processing. [12]

Sequential Model; In this model, we run 3 activation layers plus one dense layer that provides the numerical value. [12]

Activation Function; We use sigmoid as an activation function; it is an excellent generic distribution that handles well randomly occurring events and works best with a small number of layers; also, the sigmoid activation provides us a range of values between 1 and 0. [12]

The result provided is a number between 0 and 1, where closer to zero means that the price of the stock is more likely to go down and closer to one means the price of the stock is more likely to go up.

The Deep Learning model workflow and script we use is based on Menshaw's approach to analyzing movie reviews and works the following way:



**Fig. 1.** Deep Learning models Workflow [12]

1. Load the train/test set and preprocess data to:
  - (a) Normalize result behavior between 0 and 1
  - (b) Fill empty lines with zeros.
  - (c) Select the size to train and test sets
  - (d) Convert to array
2. Tokenize train/test set data
3. Ensure all sequences have the same length
4. Pad or truncate
5. Embedding or Vector creation
6. RNN model
7. Train model against the test set
8. Evaluate model accuracy
9. Try out with an example.

The deep learning models gives an immediate prediction for the latest news article, it is printed in-screen and saves the result in a CSV file.

Input1: User input of stock name and online publication

Input2: Custom Corpora Data Set

Output1: Print value of model result, between 0 and 1.

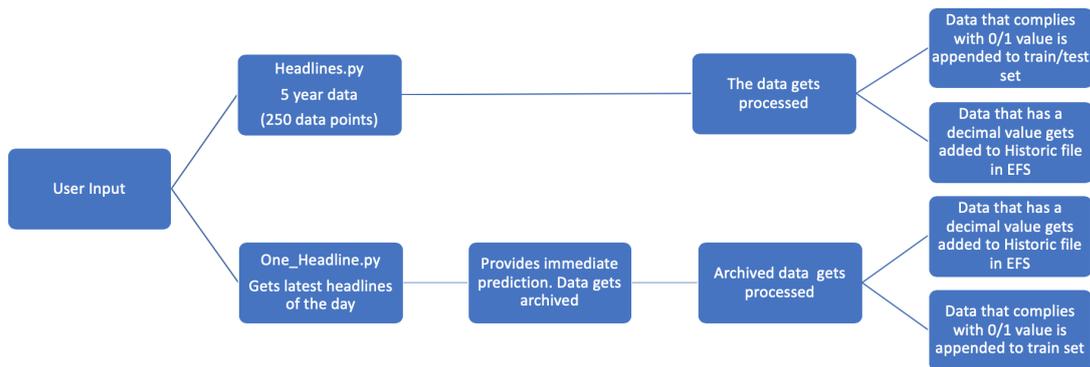
Output2: the results of the model are saved in a CSV file.

The CSV file saves all queries for later processing and ingesting into the train/test data set.

These archived results keep incrementing as the app gains popularity and more users query different stocks/publications combinations creating a bigger archiving file.

### Model Learning Cycle

The archived results get verified daily through a cron job, archived data gets processed and verified; if enough lines for the same stock (200) are recorded in the file, then said lines would get processed, reusing the script we used for gathering the initial data set. These lines will get their financial information attached to them, the percentage of change calculated, and value normalized. Then the data that has a normalized value of 1 or 0 will become part of the primary test/train data set and flagged 1, which means it was already processed; the workflow is represented on figure 2 Detailed Script Workflow.



**Fig. 2.** Prediction System Architecture

## 4 Experiments

All scripts were created using Python, Python libraries, and custom APIs; the code for the final scripts is available at the private GitHub repository: <https://github.com/claudia0juarez/Thesis>

### 4.1 Parameters

The parameters used in this research were: model accuracy and prediction results, the model accuracy will measure the performance of the model while the prediction results are calculated based on a comparison between the model prediction and the actual stock market movement.

### 4.2 Web Scrapper Development

The first step in the process was to develop the scripts that will fetch the data from online resources (news articles/web publications). We used Python and the BeautifulSoup library for the web scraper due to the ease of parsing the HTML files. For scanning the web for relevant news, we use a paid API that executes search requests in several search engines called SerpWow. [2]

### 4.3 Deep Learning Model Fine Tuning

The deep learning model fine-tuning was run using proof of concept tests and a small data sample of 9 news articles with an assessment based on stock price movement; we ran our model and then determined the level of success by comparing the results with the real data.

The deep learning model provides a result in the range of 0 to 1 where:

0 to 0.5 is a negative result meaning the stock is most likely to go down 0.5 to 1 is a positive result meaning the stock is most likely to go up

POC Test1 The first proof of concept test ran with a custom corpus size of 107,000 headlines. The model results remained pretty much in the same range; as shown on table 1 POC Test1 with Custom Data Set in page 8 we are still not getting a reliable prediction.

No	Model Accuracy	Prediction Results
1	69.21	0.49
2	69.31	0.47
3	69.27	0.46
4	69.35	0.47
5	69.22	0.48
6	69.13	0.46
7	69.21	0.47
8	69.33	0.47
9	69.15	0.47
10	69.26	0.48

**Table 1.** POC Test1 Custom Data Set

No	Model Accuracy	Prediction Results
1	71.8	0.38
2	71.92	0.37
3	71.49	0.35
4	72.08	0.37
5	71.92	0.38
6	71.68	0.38
7	72.31	0.38
8	72.29	0.35
9	71.33	0.38
10	71.66	0.39

**Table 2.** POC Test2 Custom Data Set

POC Test2 With an extended custom data set of 173,000 data points; again, the accuracy of the model greatly improved as shown on table 2 POC Test2 with Custom Data Set on page 8; however, the results were again in the same range and not providing any valuable feedback.

POC Test3 The custom data set we used has test values ranging from 0 to 1, so for the next test, we only used the data that has naturally 0 (negative) or 1 (positive) to train the model as shown on table 3 for New Testing Data on page 9.

After running the model with this new data set, we had a better accuracy result, and the actual results with the model started to vary from one heading to the next, although it was only correct 3 out of 9 times the main issue was that we had a minimal test/train data set of only 2,500 records as shown on table 4 for New Testing Data on page 9.

**Table 3.** New Testing Data

No	Stock	Headline	Date	Open	Close	Assessment Based on Stock Movement
1	Microsoft	Workplace vs. coronavirus: No one has a playbook for this	05-Mar-20	166.045	166.27	Positive
2	Microsoft	Pentagon asks to reconsider part of JEDI cloud decision after Amazon protest	12-Mar-20	145.3	139.06	Negative
3	Microsoft	Windows 10 Warning: Anger At Microsoft Rises With Serious New Failure	09-Feb-20	183.58	188.7	Positive
4	Microsoft	How corporate IT is entering the multi-cloud	14-Mar-20	140	135.42	Negative
5	Apple	Microsoft's Massive Stock Gains May Be Far From Over	09-Feb-20	314.18	321.55	Positive
6	Tesla	Tesla's Sales Fell 68% In The Netherlands And 92% In Norway In February	02-Mar-20	711.26	743.62	Positive
7	Starbucks	Starbucks Is Bringing Beyond Meat To Its Canada Locations	26-Feb-20	82.6	80.67	Negative
8	Samsung	Samsung Unveils Samsung Galaxy S20 Series With AI-Powered Camera	14-Mar-20	2209.7	2209.7	Neutral
9	Amazon	Amazon's Stock May Jump Following Quarterly Results Despite Rising Costs	26-Jan-20	1820	1828.34	Positive

**Table 4.** POC Test3 with Custom Data Set

No	Model Accuracy	Prediction Results	MODEL	ACTUAL STOCK	MODEL VS ACTUAL STOCK
1	75.17	0.61	positive	POSITIVE	
2	76.75	0.31	negative	NEGATIVE	
3	73.51	0.26	negative	POSITIVE	
4	73.57	0.69	positive	NEGATIVE	
5	72.29	0.39	negative	POSITIVE	
6	72.94	0.44	negative	POSITIVE	
7	72.61	0.38	negative	NEGATIVE	
8	76.68	0.47	negative	NEUTRAL	
9	75.19	0.48	negative	POSITIVE	

#### POC Test4

We did a round of testing converting the entire data set (173,000) and normalizing their score values to either 0 or 1 depending on if it was higher or equal than .5. With this test, we discovered that this model behaves similarly to a Naives Bayer model as it only works with naturally 0 and 1 values.

**Table 5.** POC Test4 with Custom Data Set

No	Model Accuracy	Prediction Results	MODEL	ACTUAL STOCK	MODEL VS ACTUAL STOCK
1	75.2	0.66	POSITIVE	POSITIVE	
2	72.35	0.69	POSITIVE	NEGATIVE	
3	76.18	0.62	POSITIVE	POSITIVE	
4	77.77	0.6	POSITIVE	NEGATIVE	
5	75.97	0.62	POSITIVE	POSITIVE	
6	77.4	0.6	POSITIVE	POSITIVE	
7	72.15	0.62	POSITIVE	NEGATIVE	
8	77.8	0.65	POSITIVE	NEUTRAL	
9	76.78	0.61	POSITIVE	POSITIVE	

The accuracy remained in the same range; however, the result returned to the same behavior, only giving a value within the same range as shown on table 5 POC Test4 with Custom Data Set on page 10.

The best way to obtain better results with the deep learning model is by using a good data set both in quality and quantity. Quality meaning that only truly 0 or 1 results will provide a good result, there is no use in artificially moving the value set to 0 or 1 as it will throw the overall model. Also, the quantity, with a more significant data set, the results get consistently better.

#### 4.4 Results

After running these POC and refining the model with the custom data set, we ran the custom script to get more data points (from 2500 to 3500) and re-run the test with a more prominent test set (around 272 headers). We ran 3 tests with an incrementally higher number of test sets:

##### Test1 (2505 test set)

The 272 news headlines had a 46% accuracy when tested along with the custom test set of 2505 data points, which meant that 124 out of 272 headlines were correctly predicted, and the stock price moved accordingly to said value.

Correct Prediction	124	46%
Incorrect Prediction	147	54%

##### Test2 (3521 test set)

For the second test the 272 news headlines had a 43% accuracy when tested along with the custom test set of 2505 data points, which meant that 116 out of 272 headlines were correctly predicted, and the stock price moved accordingly to said value.

Correct Prediction	116	43%
Incorrect Prediction	155	57%

### Test3 (3789 test set)

To get the final set, I needed to create 264,000 overall custom data set and then only use the 0 and 1 values; this test retrieved a 63% of accuracy that represented 170 correct predictions out of 272.

Correct Prediction	170	63%
Incorrect Prediction	101	37%

To get a better prediction result, our work shows that there is a need to keep incrementing the custom test/train data set. However, it is very resource-intensive as it takes about a day to get around 100,000 results (provided the paid APIs don't throw an error), and from this set, we will still have to filter the 0/1 results, which are about 1,000.

## 5 Cloud Architecture

Service-oriented-architecture or SOA is known for increasing the capability of an enterprise to address new business requirements with minimal cost, resources, and time overheads. We based the development of the cloud-based architecture of our prediction system on this framework.[5]

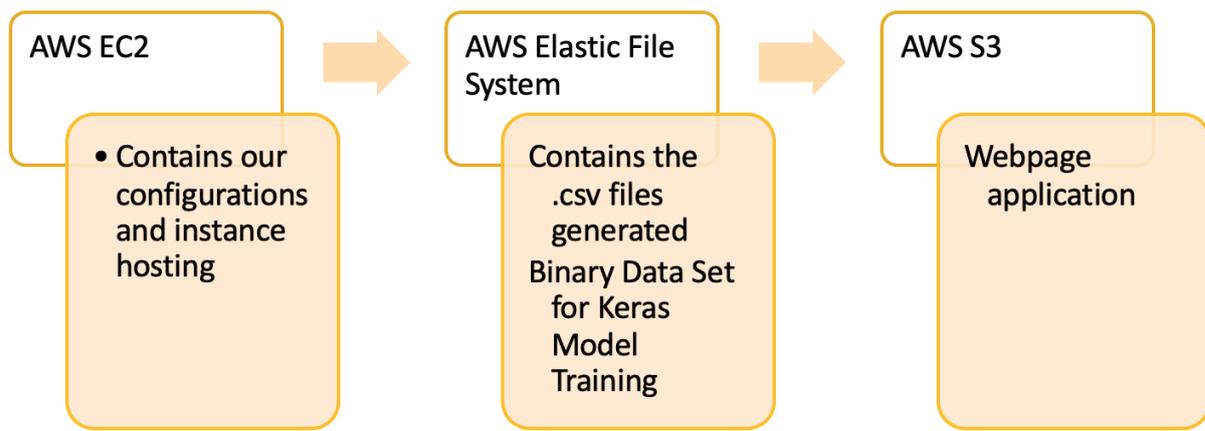
The blueprint was based in SOA to maximize the cloud benefits; it can leverage cloud computing resources as services contained within itself; it will help layout the design of services that will increase usability and durability as well as the blueprint for design, development, and deployment. [10]

Data to Cloud Roadmap:

1. Define the data. The data was taken from news publications (financial, and well-reputed sources). [5]
2. Define the services. We used AWS, although this layout might be applicable to any other cloud service provider, along with Python libraries, and paid APIs. [5]
3. Define the processes. We used ETL (Extract/Transform/Load). [5] which is a quantitative methodology based on observations and experimentation.
  - Extract; Automate the data collection with web scrapping tools for these parameters: renowned online financial publications and historical data on market prices for the past years (opening price, closing price).
  - Transform; Use data analytics tools to correlate the price changes of Company Stock vs. news articles mention on the same day. With this information, we created custom corpora and utilize NLP for training a model into determining the sentiment/vector and the effect on the price (high or low).
  - Load; Create a cloud-based web app to review different stocks, followed by the creation and documentation for a blueprint design.
4. Define governance; it is the ability to control changes to services and the usage of said services; we must control how our data is accessed, deleted, added, and altered employing processes procedures and technology. We used the AWS in-place security protocols to make sure that our Data is persistent and safe. [5]

5. Define which candidate data, services, and processes should live in the cloud and which should live on-premise (if any). As for this project, all of our data will reside in cloud services. [5]

The high-level architecture of the prediction model, as shown in figure 3 Cloud High-Level Architecture is an integration of three AWS subsystems, each of them housing a subsection of the overall process.



**Fig. 3.** Cloud High-Level Architecture

EC2 hosts the instance, application files, the network security setups, load balancer, the security groups, the VPC and the certificate manager. The EFS (Elastic File System) contains the CSV files created by the application and the data set and S3 contains the HTML web application.

The more detailed process described in figure 4 Detailed Cloud Process on page 13 starts with the public internet and the main website [www.moneyplease.trade](http://www.moneyplease.trade), the request goes through the CloudFront, which warrants that the edge services of AWS will be close to our primary users; at the moment, we are only using the EU, USA, and Canada as this is the starter level. However, we can upgrade in the future to include more edge zones.

The Cloudfront distribution is connected to an S3 bucket that contains our web app HTML file. We choose S3 because, unlike our CSV files, the content of this bucket will be static. This content is the one distributed to the static edges.

CloudFront relies on the Certificate Manager to generate trust certificates and to have a secure connection; the CM creates and updates the certificates from now on, so we do not need to worry about it.

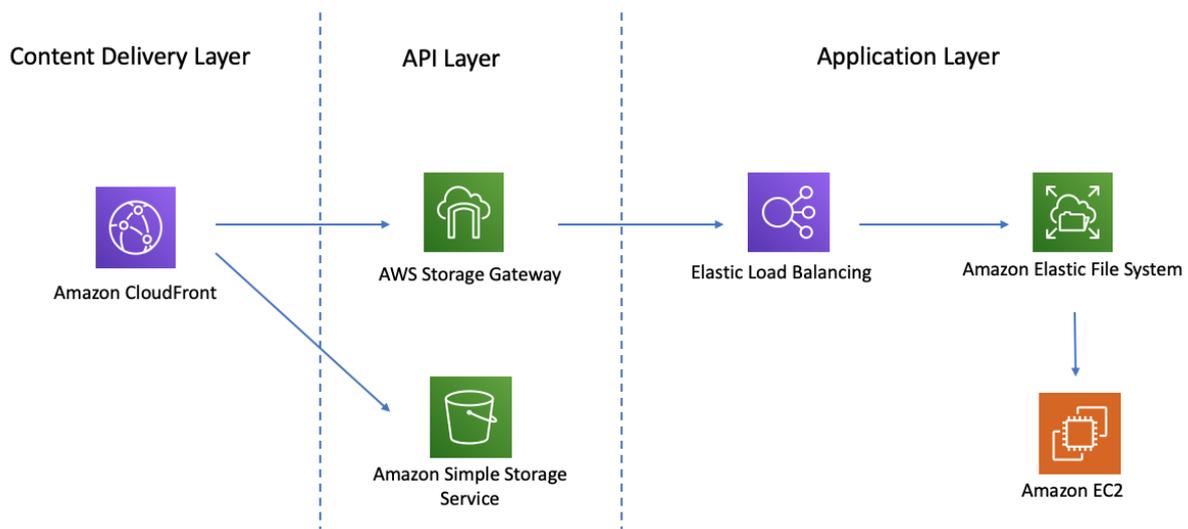
The communication between the S3 bucket/the load balancer and the EC2 instances is hosted on our virtual private cloud and not over the public internet. S3 will function as our storage system, it works best for hosting our HTML file because it is not primary for reading/writing. Whenever we require to change this file, we will need to download a copy of it, modify it and reload it again as changes in our webpage occur.

S3 will connect then to the load balancer; at this point, it is only managing a single instance. However, it can be set up for escalation (managing recurring instances in case the demand increases)

In EC2 is where the Python scripts reside, they are all managed by a Flask interface that is waiting for the user input to start the process, once the primary process starts it reads/writes from the CSV files in the EFS (elastic file system).

The EFS contains all the working files in CSV format, the data set used for training and testing of the deep learning model, the results from the searches, the historical headlines, the financial information, and the complete result data set. These files will continue to grow each time a user inputs a new search; the EFS will grow with them and will place them in the correct availability order, as files that do not get much use will have a lower priority than for example the binary data set that works with deep learning model during each run.

Another advantage to EFS is that it can be deployed and mounted to all instances and work seamlessly with them if we need escalation.

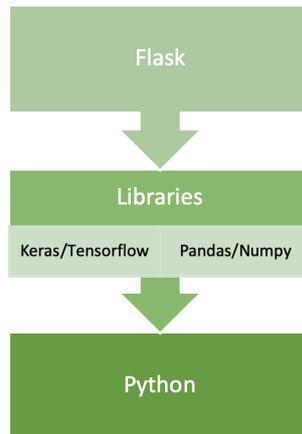


**Fig. 4.** Detailed Cloud Process [1]

### 5.1 Integration in web-based application(interface)

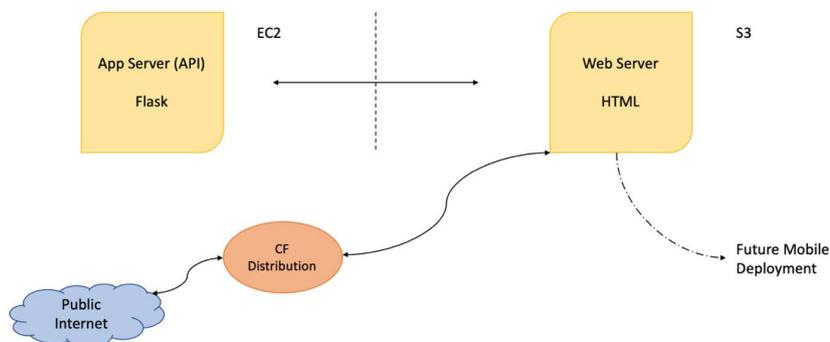
Flask is the microserver used to run the application; we selected it due to its integration with Python and its straightforward interface with our already working code; the stack for the backend is shown on figure 5 Backend Stack on page 14

To have a correct division of responsibilities in our app server, we cannot use Flask as the webserver, we needed to use HTML technology; this way we can do front-end developments independently from the back end. Flask has standard methods that allow communication with the front-end; this way we ensure that our deployment is tech agnostic, meaning that we can have future mobile apps or different web applications without changing our main code.



**Fig. 5.** Backend Stack

For the content distribution, AWS CloudFront is connected to the origin of the data. Then the DNS registries are updated for the domain to work along with CloudFront as shown in figure 6 Backend Distribution.



**Fig. 6.** Backend Distribution

## 6 Conclusion and Discussion

This work was successful in implementing a prediction model with a 64% accuracy in prediction of stock market price movement; we were also successful in creating a feasible blueprint for a deployment entirely on cloud premises services.

The most important part of the prediction model is the customized corpus made from a suitable binary dataset (only 0 and 1 values after normalizing). If we do not have a dataset containing the parameters we are evaluating, the results are not going to be representative or useful for the type of prediction we are after. A small accurate test set brings better predictions than a big messy one. This corpus will be extended with the new data searches provided by users to retrain our model.

The pay-as-you-go services for the cloud provider and API's are required, as the model grows we will need to pay for the services to keep our implementation going; the model was successfully deployed in the free-tier applications, and it is possible to scale all of the applications once the work extends its size and scope.

For future work and further research we might want to explore the growth of the cloud deployment, at the moment, we are using CSV files in an elastic file system; however, if the data keeps growing as intended this might not be the best use of cloud computing resources, for the next phase in the project we might want to add a relational database and convert the CSV files to a SQL format.

Also, we are not saving any user's personal information or manipulating any other sensitive information; however, if the application grows, we might want to add authentication based on user emails or social media.

Another improvement opportunity worth mentioning is the webscrapper development, we might want to add priorities for the websites that were crawled for the custom corpora; as it stands, the webscrapper takes headlines information for different sources (blogs, smaller news outlets, opinion pieces) as it is focusing on volume for the creation of the testing data set. If this first step gets refined and the corpora not only increases in size but in quality of the headline, it will ideally increase its reliability.

Also the deep learning model used was a popular teaching model, we might consider improving the model type to increase reliability. Although at this point it is only using one computer it is set up for future growth within the cloud space

These considerations for future releases might be excellent additions and nice-to-have, the objective of the research was met with a minimum viable product or MVP that gives a prediction based on a selected publication plus a stock name and is deployed entirely in a cloud environment.

## References

1. MicroServices Architectures on AWS.  
<https://www.slideshare.net/AmazonWebServices/microservices-architectures-on-amazon-web-services>

2. SerpWow - Google Search Results API, <https://app.serpwow.com/playground>
3. Abdelaziz, A., Elhoseny, M., Salama, A.S., Riad, A.M.: A machine learning model for improving healthcare services on cloud computing environment. *Measurement: Journal of the International Measurement Confederation* (2018). <https://doi.org/10.1016/j.measurement.2018.01.022>
4. Akshay, K., Shivananda, A.: *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning Using Python*. Apress (2019)
5. Banke K., Slama D., K.D.: *Enterprise SOA: Service-Oriented Architecture Best Practices*. Prentice Hall (2004)
6. Bird, S.: *Natural Language Processing with Python*. O'Reilly Media (2016)
7. Feyzkanov, R.: *Hands-On Serverless Deep Learning with TensorFlow and AWS Lambda*. Packt Publishing (2019)
8. Joshi, K.: *Stock trend prediction using news sentiment analysis*. Tech. rep. (2013)
9. Lamons, M., Kumar, R., Nagaraja, A.: *Python Deep Learning Projects*. Packt Publishing (2018)
10. Linthicum, D.S.: *Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide*. Addison-Wesley Professional (2009)
11. Mathirajan, A.I.S.K.N.K.M.: *Management Research Methodology: Integration of Principles, Methods and Techniques*. Pearson India (2006)
12. Menshawy, A.: *Deep Learning by Example*. Packt Publishing (2018)
13. Naren, J.: *News analytics and dual sentiment analysis for stock market prediction*. *IEEE International Conference on Big Data Analysis(ICBDA)* (December) (2017)
14. Piard, F.: *The Lazy Fundamental Analyst: Applying Quantitative Techniques to Fundamental Stock Analysis*. Harriman House Ltd, (2014)
15. Yaojun, W., Wan, Y.: *Using social media mining technology to assist in price prediction of stock market*. *IEEE International Conference on Big Data Analysis(ICBDA)* (2016). <https://doi.org/10.1109/icbda.2016.7509794>.

# Common sense validation and reasoning using Natural Language Processing

Shrenik Doshi, Praveen Joshi, and Haithem Afli

ADAPT Centre,  
Cork Institute of Technology, Cork, Ireland  
doshishrenik0120@gmail.com  
haithem.afli@cit.ie

**Abstract.** Natural Language Processing is an emerging field of Artificial Intelligence. Various real-world problems are now solved using the advancements of NLP. This paper is focused on how NLP can be used to validate whether a sentence is making any sense or not. Also, generating reasons for the sentence which does not make sense can be addressed. The validation of sentence is being done using the n-gram language models. On the other hand, the problem of generating a sentence is solved using the pre-trained models.

**Keywords:** Common sense reasoning(CSR) · Common sense validation · Natural Language Processing (NLP) · Natural Language Understanding (NLU) · Natural Language Generation (NLG) · Neural Networks (NN)

## 1 Introduction

Natural language understanding (NLU) and Natural Language Generation (NLG) are evolving rapidly. Computers are becoming more and more intelligent. Nowadays, computers can understand a language and generate another sentence using reference sentences. One of the applications of NLU is common sense validation. On the other hand, NLG helps generate reasons commonly termed as common sense reasoning (CSR). For developing an intelligent system, a system must have some common sense which can differentiate between sentences making any sense or not. There has been much work done in the field of CSR. Commonsense reasoning is a very complex problem that is not possible to solve using any single technique of Artificial Intelligence. So to develop a common sense understanding capable system, one needs to incorporate multiple AI techniques. Commonsense reasoning is expected to be solved easily by kids but it is difficult for computers to do appropriate reasoning. The human capacity to comprehend language is general, adaptable, and powerful. Conversely, most NLU models over the word level are intended for a particular assignment and battle with out-of-area data.

## 2 Related Work

As the field of Natural Language Reasoning has evolved over the years, there has been remarkable work done before. In the year 2002, a paper was written

[8] which stated how NLU can help in story understanding and answering questions about such stories. This paper focused on understanding texts written by progressively harder children. The RTE challenges [1] started in the year 2005. These tasks were mainly for recognizing from two given text fragments whether the meaning of one text can be inferred from the other text. It captures major inferences about the variability of semantic expression which are commonly needed across multiple applications [1].

The problem with RTE challenges of the year 2005 are that there are no natural distributions of Text-Hypothesis examples. For example, T-H pairs may be collected directly from the data processed by actual systems, considering their inputs and candidate outputs[1]. The latest RTE challenge was in the year 2011 which is known as Seventh Pascal RTE. There is another similar challenge which is known as Winograd Schema Challenge (WSC) which was held in 2011. Here the system is presented with questions about sentences known as winograd schemas. To answer a question, a system must disambiguate a pronoun whose coreferent may be one of two entities, and can be changed by replacing a single word in the sentence[14]. CoPA (Choice of Plausible Alternatives) is another such system which uses a forced choice format. Each question in CoPA gives a premise and two plausible clauses or effects, where the correct choice is the alternative that is more plausible than the other.[11]

In recent years, the field of Natural Language Processing (NLP) has evolved very rapidly. With the field of NLP growing fast, this field is further divided into many subdomains. Some of these subdomains are Natural Language Understanding (NLU) and Natural Language Generation (NLG). Computers are becoming more and more intelligent. One of the applications of NLU is common sense reasoning (CSR). For developing a smart system, a system must have some common sense which can differentiate between sentences, which makes sense or does not make sense. There has been much work done in the field of CSR. Commonsense reasoning is too hard a problem to solve using any single artificial intelligence technique [8].

On the other hand, NLG is also a very vast field of research and application. One of the applications of NLG is common sense reasoning. This application of NLG is related to the generation of sentences naturally by the machine. The research done in this paper is a combination of NLU and NLG. Common-sense validation and reasoning is too hard a problem to solve using any single artificial intelligence technique[8]. So to develop a common sense understanding capable system, one needs to incorporate multiple AI techniques. Commonsense reasoning-the sort of reasoning we would expect a child to do easily-is difficult for computers to do[8]. The human capacity to comprehend language is general, adaptable, and powerful. Conversely, most NLU models over the word level are intended for a particular assignment and battle with out-of-area data.

### 3 Material

#### 3.1 Dataset and Acquisition

Common sense validation and reasoning both uses different datasets. The dataset specified in this paper is used[13]. The dataset is manually labelled by 7 annotators. Human performance on the benchmark is 99.1% for the common sense validation dataset and 97.3% for the common sense reasoning dataset[13].

For common sense validation, the dataset consists of two similar sentences which are in the same syntactic structure and differ by only a few words. Only one of them makes sense while the other does not.[13]. Below are some statistics of training and test dataset for common sense validation part: For common sense

Commonsense Sentences Against Commonsense Sentences		
Training Data	10,000	10,000
Trial Data	2021	2021
Test Data	1000	1000
Dev Data	997	997

**Table 1.** Common sense validation dataset statistics

reasoning, the dataset consists of three reference sentences for each "against common sense" sentence. Using these reference sentences, a new sentence needs to be generated.

Reference Sentences	
Training Data	30,000
Trial Data	6063
Test Data	NA

**Table 2.** Common sense reasoning dataset statistics

Other than this, an external corpus has also been used to train the language model. This is a collection of the proceedings of the European Parliament, dating back to 1996[5].

Below is the brief discussion on these datasets

1. Validation Task dataset[13]: The dataset is in simple CSV format. This dataset contains 2 data files for all the trains and trials. There is only 1 file for the test data. Train and trial data has 3 columns. The first file, consists of 3 columns. The first column is the sentence id, and the next two columns contain sentences. One of the sentences is making sense, and others do not make sense. The other file is used for classification. There are 2 columns in this CSV file. The first column is the sentence id, and the 2nd

column is column id for each sentence in the file, which does not make sense. The test data simply has only one file with 3 columns. 1st column is the sentence id, and the other 2 columns are sentences. One of them making sense, while others do not make any sense.

2. Reasoning Task dataset[13]: In this dataset, there are 2 different CSV files. The first CSV file has 2 columns. First column in this file is the sentence id, and the other column is the sentence, which does not make any sense. The second data file has 3 reference sentences for each sentence in the first data file. These reference sentences are the reasons that state why the sentence in the first file does not make any sense.
3. Europarl dataset [5]: This is a massive corpus of data. This corpus is extracted from parliament sessions in the European Parliament. It includes data of 21 different languages, but we will only use the one which English. This dataset consists of document files, and these files are read and passed through the pre-processing layer to get the actual trained model.

## 4 Methodology

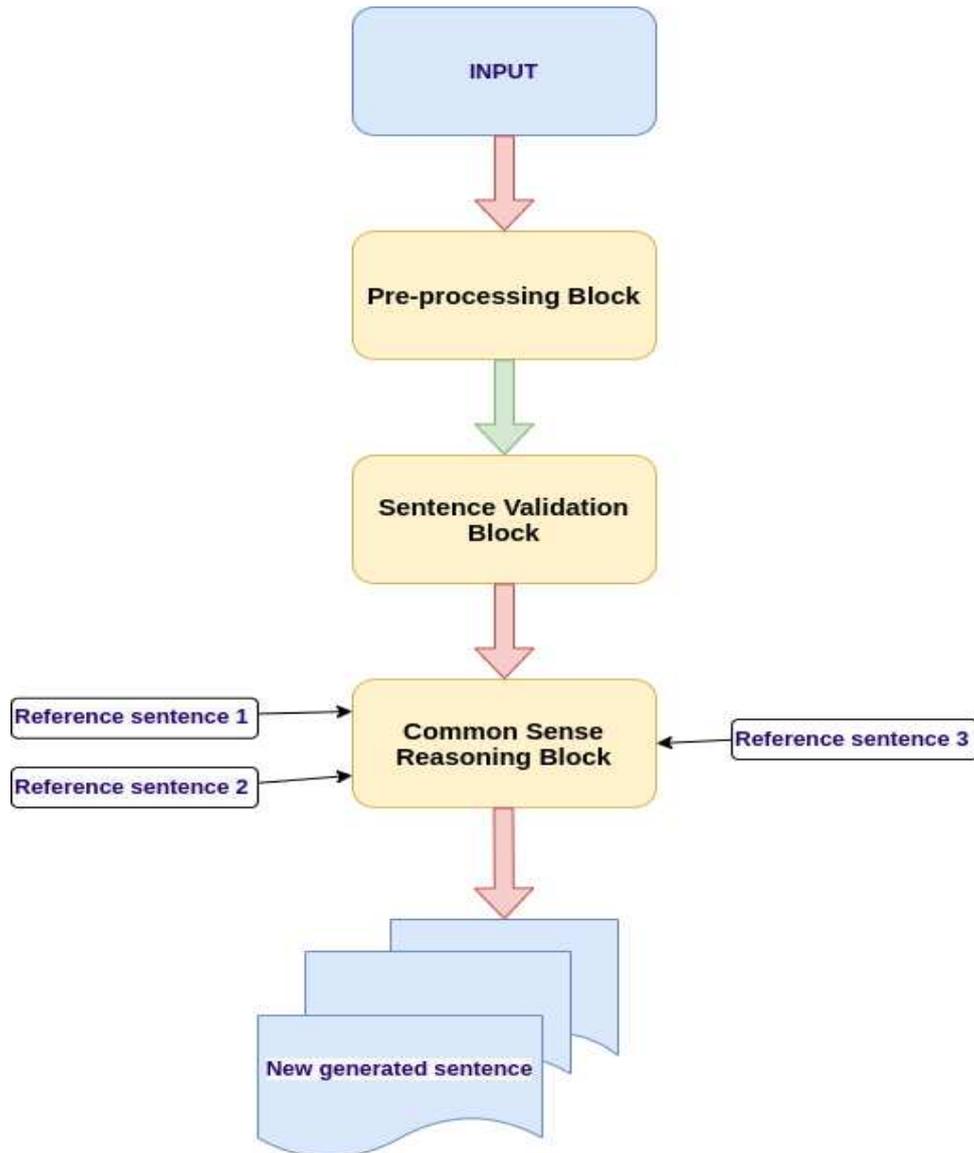
In this section, the detailed architecture will be discussed. This chapter includes all the implementation details, brief description about different parts of system like sentence validation block and reasoning block, and also some important code snippets. A brief description of each main functionality of the system is discussed in detail.

### 4.1 Architecture

Figure 1 shows detailed information about each block and the different functionalities used in each block. As discussed earlier each block is dedicated to a specific functionality. These blocks, for the purpose of this system, are divided into phases. The input to each phase is the output of it's previous phase. Following are different phases the system is divided into:

**Pre-processing Block** This block is responsible for removing the unwanted and useless things from the corpus. The dataset consists of many sentences, and it is a huge dataset. The data have to be brought into some standardized form. The pre-processing step is further divided into 2 phases:

Removing noise: There is a large number of stop-words (like, and, the, a, in, an, etc.). The occurrences of these words are very high. Due to these words, the accuracy of the language may decrease. The language model works on the concept of probability. It predicts the next possible word by looking at the probability. Other than stop-words, there is even punctuation. These punctuation needs to be removed; otherwise, they may also hamper the accuracy of the model. Last but not least, the word's case. The language is case sensitive.



**Fig. 1.** Detailed Architecture

For example, the words "Mango", "mango", and "mAngO" have the same meaning, but if kept like this, they will be treated as different words. This scenario may also affect the accuracy of the model. Now all the words are lowercased. The sentences may also URLs, and these URLs are not important for this system. So we will remove them too.

Text normalization: In this phase of pre-processing, the sentence is first divided into tokens. As the system creates language models, the data fed to these models are in the form of tokens. It is easy to create a language model around tokens. These tokens are nothing but words. Once the sentence is tokenized, the next step is lemmatizing each token. Lemmatization is a process of convert a word into its base word. For example, the words run, running, runs, ran, etc. means

the same. Just the tense of each word is different. When we lemmatize these words, they convert to the base words, which is 'run'.

### **Sentence Validation Block**

This block is responsible for predicting whether the sentence is making sense or not. This block focuses on the n-gram language models:

n-gram language models:

One is a traditional n-gram language model. Ngram is a type of Markovian model. The model is based on counting the occurrences of words in the corpus. Once the frequency of words is found, the probability for that word given some condition can be calculated. While this method of estimating probabilities directly from counts works fine in many cases, it turns out that even the web isn't big enough to give us good estimates in most cases[7]. This is because language is creative; new sentences are created all the time, and we won't always be able to count entire sentences[7]. For the purpose and requirement of this project, 3-grams and 4-grams models are built. These n-gram models are trained on an external corpus of Europarl data[5].

BERT:

Second is the more advanced pre-trained model called "BERT". BERT provides the pre-trained vectors representation of the words, which can be used further with the various AI models[4]. BERT architecture is a frame to provide representations by joint conditional probabilities both from the left and right context for all the processing layers[2]. BERT vectors are used in the experiment to utilize the shallow transfer learning models to enhance the current predictive models. BERT is used as a service to convert processed text to its corresponding vector[4]. Amongst all the models of BERT- BaseUncased has the capability to a single word in as many as 768 different dimensions. BERT is a powerful model that predicts the next words and gives a score for each sentence based on its occurrence in the dataset. Each of these models gives the prediction score for both the sentences. The sentence with the lowest prediction is considered as the sentence, which is against common sense.

### **Common Sense Reasoning Block**

This is the next phase of the system. This block is responsible for explaining the incorrect sentence. As discussed in the previous chapter, the system is not useful until and unless it explains why the sentence does not make sense. Initially, this phase was developed from scratch. LSTM [6] was used to create a sequence to sequence encoder-decoder model. Recurrent Neural Network (RNN) was the heart of this phase. In this approach, we used to generate a sentence using the present reference sentences. This is different from the extractive approach. In this approach, we will generate sentences by abstractively summarizing the references. Original text might not have the generated sentences in them.

This approach of abstractive summarization uses Sequence to Sequence Modelling.

### Sequence-to-Sequence(Seq2Seq) Modelling:

Seq2Seq modeling can be applied to a wide variety of problems like text classification, sentimental analysis, machine translation, or text generation. The only constraint of using this model is that the information must be sequential like a chain of words, sentences, phrases, etc. Our goal is to create a text summarizer where the input is a long list of words (in a body of text), and the output is a short description (which is also a list of words). The approach to solving our is to use Many-to-Many Seq2Seq models. Encoders and decoders are the building blocks of the Seq2Seq model. The entire input data is fed to the LSTM's encoder model, and this encoder reads the entire input sequence. At each timestep, a word or a token from the input sequence is fed to the encoder. The information is processed at each timestep, and only the contextual information present in the input sequence is captured. All the other information is discarded. The encoder step can be illustrated from the below diagram. The process of creating LSTM Seq2Seq is referred from here[10].

Before sending the input sequence into the decoder, some special tokens like <start >and <end>are added to each input target sequence. These tokens indicate the start and end of the target sequence. It is challenging to decode the target sequence with the test sequence. To avoid this, the prediction of the output sequence starts by sending a <start>token as the first word to the decoder. The <end>token indicates that the word is the last in the sequence.

After training process is complete, the model is tested on the new test data which are new source sequences and are not been used in the training phase. We need an inference phase which takes care of this. The entire LSTM network is created, but it may not work for a long sequence of sentences. The LSTM encoder networks manage to convert the whole sequence of input into a vector of a fixed length, and prediction of the out sequence is made by the decoder. But when there is a long sequence, then encoder fails to memorize this sequence and is not able to convert it into vectors. To solve with have added an 'Attention mechanism' to it. This attention layer is responsible for giving an importance level to each word in the sequence. Depending on the importance level, the encoder network will remember only the essential parts in the input sequence, which results in the output sequence. There are no such things as attention layer in Keras. This attention layer is referred from this GitHub repository[12].

### GPT-2

The other approach used is the pre-trained GPT-2 model. This model is one of the most influential models that is used for text summarization and text generation. The model is trained on an extensive corpus. Additionally, the model has been further trained on the dataset of this project. This model takes care

of the context of the sentences and generates a meaningful sentence using the reference sentences. This block takes in additional inputs in the form of sentences. At least 2 reference sentences are needed in this step. These reference sentences are those which users will input and are some reasons about why the sentence is against common sense. As it is still a prototype, it asks the user to enter the reference sentences. Once the model is trained strong enough, then it won't require any additional reference sentences.

### Fine-tuning

As GPT-2 is a pre-trained model, it does not have any knowledge about the dataset being used in this project. To train the model with this dataset, we have to train the GPT-2 model again. But, GPT-2 does not allow direct training on their model. To train the GPT-2 model, fine-tuning has to be performed. To perform fine-tuning, **gpt-2-simple** is used. This is a simple Python package that wraps existing model fine-tuning and generation scripts for OpenAI's GPT-2 text generation model[9].

## 5 Experiments and Evaluation

While training the model, the model was evaluated on many different parameters. Hyper-parameter tuning was performed, and the best possible combinations were used. When the model was evaluated on the trial data, some basic variations were performed.

### 5.1 Evaluation Matrix

The model was evaluated with the pre-processing block and without the pre-processing block. Table 3 depicts the accuracy achieved for the Common sense validation phase with and without pre-processing block. We can observe that the accuracy of the n-gram models (3-gram, 4-gram) and pre-trained BERT is not that good, but the results are acceptable.

Language Model	Without pre-processing	With Pre-processing
Trigram	0.50	0.55
Fourgram	0.45	0.51
BERT	0.49	0.49

**Table 3.** Common sense validation accuracies

Table 4 show the confusion matrix for trigram model where as table 5, shows the confusion matrix for the fourgram model.

Total Samples	Actual Positive	Actual Negative
Predicted Positive	4106	873
Predicted Negative	3671	1350

Table 4. Trigram model confusion matrix

Total Samples	Actual Positive	Actual Negative
Predicted Positive	4379	600
Predicted Negative	4327	694

Table 5. Fourgram model confusion matrix

## 6 Results and Discussions

### 6.1 Results

BLEU is a score for comparing a candidate translation of the text to one or more reference translations [3]. Figure 2 shows the generate sentences and the BLEU score of those sentences. The generated sentences follow the BLEU score.

Generated Sentence: Now it could be the elephant is an evolutionary specialist so he wants to do it the hard way in  
The BLEU score 4.209305712440485e-155

Generated Sentence: dream can't be seen while you are dreaming ,it happens all the time during dream. your  
The BLEU score 0.18207052811092134

Generated Sentence: friendship makes them feel good to know how much others care how they feel. friendship is healthy. the  
The BLEU score 1.1640469867513693e-231

Generated Sentence: I think if you play well, you could make it on the team. But the  
The BLEU score 9.257324954728539e-232

Generated Sentence: Some people make up words for their feelings. For example, they have names like p  
The BLEU score 9.257324954728539e-232

Generated Sentence: Coloured water lollipopscoloured water lollipops are a cheap knock  
The BLEU score 9.412234823955334e-232

Generated Sentence: So what? In order that you might take a better approach with your life, you must realize that  
The BLEU score 8.844844403089351e-232

Generated Sentence: And there's no reason for you to have to change. So stop changing all the time.  
The BLEU score 1.0832677820940877e-231

Generated Sentence: It's not just another woodchip in a box.The first and foremost task we should  
The BLEU score 5.74340400309589e-155

Fig. 2. Blue score for generated sentences

### 6.2 Discussion

This paper was focused on one of the most challenging tasks in the field of NLP, which is Common sense validation and reasoning. There is much research done and still going on in this field. The proposed system is successfully meeting the needs of the problem statement and is challenging the evaluation of the baseline

model. The system clearly shows how language models can validate any given sentences. Additionally, it also showcased how the pre-trained models were useful for the validation task. On the other hand, an attempt was made to create an RNN model using LSTM to create a deep neural network to handle the task of common sense reasoning. This part was mainly focused on how NLG can help to generate reasons to support the validation task. Also, GPT-2 pre-trained model was to handle the task of reason generation.

## 7 Conclusion

The models created for validation tasks were not successful in giving better accuracy than the present pre-trained models. It is observed from the output that due to less training data and also time constraints, the model wasn't trained up to the mark. Though, the models gave accuracies, which are still acceptable, as this is a very new problem.

## 8 Acknowledgments

This research was conducted with the financial support of ADVANCE CRT PHD Cohort under Grant Agreement No. 18/CRT/6222 and at the ADAPT SFI Research Centre at Cork Institute Of Technology. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant 13/RC/2106.

## References

1. Ido D., Bill Dolan, B.M., Danroth: Recognizing textual entailment: Rational, evaluation and approaches **62**, 1–17 (November 2009), <https://tinyurl.com/yc7qlhzt>
2. J. Devlin, M.-W. Chang, K.L., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2018)
3. Jason Brownlee: A gentle introduction to calculating the bleu score for text in python. <https://machinelearningmastery.com/calculate-bleu-score-for-text-python/> (2019)
4. Joshi, P.: Ensemble event driven stock market prediction declaration of authorship (10 2019)
5. Koehn, P.: Europarl: A parallel corpus for statistical machine translation (2001), <http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf>, [Online; accessed 11-May-2020]
6. M. Galetzka, L.S., Weber, C.: Intelligent predictions: an empirical study of the cortical learning algorithm. University of Applied Sciences Mannheim (2014)
7. Martin, D.J..J.H.: Chapter 3. n-gram language models (2019)
8. McCarthy, J; Minsky, M.S.A.G.L.e.a.: An architecture of diversity for common-sense reasoning. IBM Systems Journal; Armonk **41**, 530 – 539 (2002)
9. minimaxir: gpt-2-simple, <https://github.com/minimaxir/gpt-2-simple>

10. Pai, A.: Comprehensive guide to text summarization using deep learning in python (2019), <https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/>
11. Roemmele, Melissa, B.C.G.A.: Choice of plausible alternatives: An evaluation of commonsense causal reasoning. AAAI Spring Symposium - Technical Report (2011)
12. thushv89: Keras attention layer (2019), <https://tinyurl.com/y6vcjkmz>
13. Wang, C., Liang, S., Zhang, Y., Li, X., Gao, T.: Does it make sense? and why? a pilot study for sense making and explanation (2019)
14. Zhang, Liu, L.G.D.V.D.: Record: Bridging the gap between human and machine commonsense reading comprehension (2018)

# ZKPs: Does This Make The Cut?

## Recent Advances and Success of Zero-Knowledge Security Protocols

Stavros Kassaras<sup>1</sup>, Leandros Maglaras<sup>1,2</sup>

<sup>1</sup>Department of Computer Science and Computational Biomedicine, School of Science, University of Thessaly, Lamia, Greece

<sup>2</sup>School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, UK  
kassaras.stavros@gmail.com

**Abstract.** How someone can get health insurance without sharing his health information? How you can get a loan without disclosing your credit score? There is a method to certify certain attributes of various data, either this is health metrics or finance information, without revealing the data itself or any other kind of personal data. This method is known as “zero-knowledge proofs”. Zero-Knowledge techniques are mathematical methods used to verify things without sharing or revealing underlying data. Zero-Knowledge protocols have vast applications from simple identity schemes and blockchains to defense research programs and nuclear arms control. In this article we present the basic principles behind ZKP technology, possible applications and the threats and vulnerabilities that it is subject to and we review proposed security solutions.

**Keywords:** zero-knowledge, interactive proofs, zk-SNARKs, blockchain.

## 1 Introduction

Zero-Knowledge proofs (also commonly referred to as ZKPs) are used as security protocols through which a digital authentication process can be facilitated without the use of any passwords or other sensitive data. As a result of this, no information, either from the sender’s or receiver’s end can be compromised in any way. The idea underlying zero-knowledge proofs first came to the fore back in 1985, when developers Shafi Goldwasser, Charles Rackoff and Silvio Micali [1] presented to the world the notion of “knowledge complexity” — a concept that served as a precursor to ZKPs. Simply put, ZKP is a probabilistic based verification method, which means it provides “fact-like statements” and “statements about personal knowledge” that can accumulate to show that the validity of an assertion is overwhelmingly probable. In other words, they don’t prove something that simply revealing it would be sufficient, but rather verify the knowledge of it. Moreover, the assertion cannot be verified by a third party.

This is especially handy in various situations, e.g. when we want to spend money without revealing how much we used or in which currency it was spent. Besides money, private information such as date of birth, bank statements, transaction histories,

education credentials are vitally important. Companies like Facebook and Google leverage from this data by using it to optimize their services to better appeal to you, and to re-sell it to other companies. It is prevalent now than ever before the need to maintain privacy in a data-based world.

The paper is organized as follows. We provide an overview of the methodical approach used, important literature review and relevant references for this paper in Section 2. In Section 3 we will dive deeper to the nuts and bolts of ZKPs and we will mention important theorems that drove its evolution. In Section 4 we present notable and the most recent use cases of zero-knowledge protocols. In Section 5 and subsequently in Section 6 we are discussing known threats and solutions of ZKP technologies. Finally, we discuss the current state of Zero-knowledge protocols in Section 7 and conclude in Section 8.

## 2 Methodology

Firstly, we approached from a theoretical point of view the rudiments of the ZKPs field. This served not only as a concise presentation of the principles underlining ZKPs but also gave us a historical aspect on how these protocols evolved. In this first step we consulted Rubinstein-Salzedo's "Cryptography" as a primer. This book is less sophisticated from others but manages to cover substantial topics in cryptography in an informal view. Silverman's "An Introduction to Mathematical Cryptography" offered a mathematical approach for ZKPs and notions surrounding their functionality such as Complexity Theory and Random Oracle Models. A detailed and descriptive approach would require a strong background both in cryptology and in abstract algebra.

Secondly, we searched mostly online articles and blogs on the current trends and technologies on ZKPs protocols since internet is the first place that breakthrough methods make headlines. We made sure to accompany all the new implementations appearing in this paper with the relevant citations. A contemporary approach is not only mandated because the field of ZK applications is still in early stage but also it serves as a guide for the aspiring cryptographer or mathematician.

## 3 The Principles of Zero-Knowledge Proofs

We will start this Chapter by presenting two simple examples that demonstrate the concept of zero-knowledge proof.

### 3.1 Ali Babaa's Cave (Example #1).

The first and most important example is inspired by a paper titled "How to Explain Zero-Knowledge Protocols to Your Children" [2].

Consider, for the sake of example, a cave consisting of a circular tunnel. Diametrically opposite to the entrance of this cave, there is a door which can only be opened by a password. Although this situation is probably not a real life scenario, it is quite useful

in the display of the basic properties of ZKP. Now Peggy (also known as the prover of the statement) knows the password to this door, and she wants to prove this to Victor (also known as the verifier of the statement) without actually disclosing it to him. They set off to complete the task as follows:

Peggy goes into a random branch of the cave (that is, left or right). She does this without Victor knowing which branch she chose. Standing at the entrance of the cave, Victor calls out a random branch (again, either left or right), where he wants Peggy to come out from. Providing she really does know the password, she can obey Victor every time, using the door if necessary.

However, if Peggy did not know the password, then she would only be able to return by the named path if Victor were to give the name of the same path by which she had entered. Since Victor would choose left or right at random, she would have a 50% chance fooling Victor. If both (prover and verifier) were to repeat the above process several times, say 20 times, Peggy's chance of successfully anticipating all of Victor's requests would become vanishingly small.

Thus, if Peggy repeatedly appears at the exit Victor names, he can conclude that it is extremely probable that Peggy does, in fact, know the secret word.

### 3.2 Two Balls and the Color-Blind Friend (Example #2).

Another classic example used to demonstrate ZKP is the following [3].

Imagine your friend is color-blind and you have two balls: one red and one green, but otherwise identical in their shape and size. To your friend they seem completely identical and he is skeptical that they are actually distinguishable. You want to prove to him they are in fact differently colored, but nothing else, thus you do not reveal which one is the red and which is the green.

You give the two balls to your friend and he puts them behind his back. Next, he takes one of the balls and brings it out from behind his back and displays it. This ball is then placed behind his back again and then he chooses to reveal just one of the two balls, switching to the other ball with probability 50%. He will ask you, "Did I switch the ball?" This whole procedure is then repeated as often as necessary. He knows if he switched the ball because he did it himself, and you know if he did (because you can see the color) without revealing to him the actual color of the ball.

The above examples demonstrate an important subtle feature, that of zero-knowledge. Victor cannot convince anybody else of Peggy's knowledge about the password. If Victor were to create a transcript (e.g. videotape the whole process) that could potentially be that of the communication between him and Peggy, it would be useless. This transcript would be indistinguishable from a transcript that is entirely fabricated by a cheating verifier. An outsider, watching the recording, could argue that Peggy and Victor agreed in advance about the sequence of chosen branches. Thus, such a recording will certainly never be convincing to anyone but the original participants.

### 3.3 Characteristics of ZKPs

**Interactive ZKPs:** The examples above are forms of interactive proofs since the prover, performed a series of actions to convince the verifier, of a certain fact. The problem with interactive proofs is their limited transferability: to prove an ability attribute, or possession of secret data to someone else, or to the verifier several times, the prover will have to repeat the entire process.

Interactive ZKPs have further properties, namely Completeness and Soundness.

*Completeness Property:* An interactive proof protocol is complete if, given an honest prover and an honest verifier, the protocol succeeds with overwhelming probability (i.e., the verifier accepts the prover's claim).

*Soundness Property*<sup>1</sup>: If the prover is lying, then he cannot convince the verifier that he is telling the truth, except with some very small probability.

*Zero Knowledge Property:* Let's consider the cave example once again and the videotape (transcript) Victor made. We mentioned that if the recording were to be seen by a third party, it would not convince this party for Peggy's knowledge. Thus, if Victor wants to convince a third party, he can ask Peggy to demonstrate the transcript once again, but this time it will be the third party who will pick his own sequence of challenges for Peggy (not Victor's challenges). If there is a way to forge a proof that is indistinguishable from a genuine one (as in the case of the videotape), we say that there is a simulator for the proof in question. A proof of knowledge has the zero knowledge property if there exists a simulator for the proof.

**Non-interactive<sup>2</sup> ZKPs (NIZK):** In a non-interactive proof, the prover can deliver a proof that anyone can verify for themselves. This relies on the verifier picking a random challenge for the prover to solve.

Cryptographers Fiat and Shamir [4-5] found that an interactive protocol can be converted into a non-interactive<sup>3</sup> one using a hash function to pick the challenge (without any interaction with the verifier). Repeated interaction between the prover and verifier becomes unnecessary, since the proof exists in a single message sent from prover to verifier.

---

<sup>1</sup> Soundness can be described mathematically as an expected polynomial time algorithm  $M$  with the following property: if a dishonest prover can with non-negligible probability execute the protocol with the verifier, then  $M$  can be used to extract from this prover the knowledge which with overwhelming probability allows subsequent protocol executions.

<sup>2</sup> A Non-interactive, zero-knowledge proof example can be found here <https://blog.goodaudience.com/understanding-zero-knowledge-proofs-through-simple-examples-df673f796d99>

<sup>3</sup> Only languages in BPP have NIZK proof systems, under suitable hardness assumptions, NIZKs exist for all languages in NP.

## 4 Applications

### 4.1 Blockchain Use Cases

So far ZKPs may seem like a conundrum and you might wonder if there are real world consequences. In this chapter we will present some important applications.

Probably, the most prominent use case of Zero-knowledge proofs is within the context of a blockchain ecosystem. It offers a lot of benefits in regard to validating cryptocurrency transactions without disclosing any data related to it - such as where the transactions originated from, where it went or how much money was transferred.

A real-world use case of this technology is Zcash, a crypto platform that employs a special iteration of zero-knowledge proofs - called zk-SNARKs - that allow native transactions to remain fully encrypted on the blockchain while still being verified under the network's consensus rules.

The possibilities of zk-SNARKs are impressive, you can verify the correctness of computations without having to execute them and you will not even learn what was executed - just that it was done correctly. zk-SNARK stands for Zero-Knowledge Succinct Non-interactive ARguments of Knowledge [6], “Zero-knowledge” because they don’t reveal any knowledge to the verifier apart from ensuring that the transaction is valid, “Succinct” because the size of the proof is small enough to be verified in a few milliseconds, “Non-interactive” because the proof consists of a single message sent from prover to verifier and “Arguments” because the Soundness Property holds true.

The usual cryptocurrencies, like Bitcoin, validate their transactions by linking the sender and receiver addresses, and input and output values on the public blockchain. Instead of exposing the above components, Zcash make use of zk-SNARKs to “obfuscate” them - it diminishes any meaningful connection between sender, receiver and amount.

In a nutshell, if a sender wants to create a shielded transaction, he constructs a proof to show that with high probability, the input values sum to the output values for each shielded transfer, the sender proves that they have the private spending keys of the input notes, giving them the authority to spend and the private spending keys of the input notes are cryptographically linked to a signature over the whole transaction, in such a way that the transaction cannot be modified by a party who did not know these private keys.

Users of cryptocurrencies often couple them with network-layer privacy enhancements like Tor to level up their anonymity (we should better say pseudonymity in this case) with unpleasant results despite their efforts [7], unlikely Zcash does not suffer from the same threats.

Zcash protocol has been found to be vulnerable to a number of attacks [21]:

- **ITM Attack**, that is a linkability attack against shielded transactions
- **Danaan-Gift Attack**, where the attacker transfers a very small but carefully chosen amount of Zatoshis in order to identify the spending of a public address
- **Dust Attack**, that is used to describe any very small amount that does not meaningfully cost much to the attacker

## 4.2 From zk-SNARKs to zk-STARKs

As this wasn't enough there is a more developed version of zk-SNARKs, it is called zk-STARKs - Zero-Knowledge Scalable Transparent Argument of Knowledge and it was introduced in 2018 (very recently) by Eli Ben [8].

Prior to the creation of zk-STARKs, zk-SNARKs required a trusted party to initially setup the ZK proof system which introduced the vulnerability of those trusted parties compromising the privacy of the entire system (read more about vulnerabilities in Section 5). zk-STARKs improve upon this technology by removing the need for a trusted setup. In other terms, zk-STARK proofs present a simpler structure in terms of cryptographic assumptions.

The great advantage of zk-STARK is its scalability, meaning it can move computations and storage off-chain. Off-chain services will be able to generate STARK proofs that attest the integrity of off-chain computations and then can be integrated back on chain for any interested party to validate the computation. Also, while zk-SNARK communication complexity - that is the amount of communication needed to solve a problem distributed among two or more parties - increases in a linear fashion, on the other side zk-STARK develops in the opposite way, and increase only slightly as the computation size grows. The same applies to the verifier complexity. zk-STARKs are about 10 times faster than zk-SNARKs as computation size increases.

## 4.3 Quantum Resistant

Quantum computing has become a topic of interest and despite the fact they are characterized with many novel attributes, unfortunately the truth is far from reality. Quantum computers can achieve only special kinds of calculations and some of them could exploit today's cryptographic techniques. Encryption schemes based on RSA and Elliptic Curve Cryptography could prove obsolete in the near future. Notice that these algorithms rely on private and public key pairings, something that doesn't apply in the case not only of zk-STARKS but also other ZKP methods in general.

## 4.4 Zero-knowledge proofs in Banking

In October 15 in 2018, ING published a report and subsequently an article [9], announcing its own addition of ZKP to the blockchain technology. ING Bank is continuing further down the path of advanced blockchain privacy with the release of its Zero-Knowledge Range Proofs (ZKRP) solution.

The ZKRP scheme proposed can be used to prove a number is within a specific range. For example, a mortgage or loan applicant could prove that their salary or credit score sits within a certain range without revealing the exact figure. As such range proofs are computationally lighter than regular zero-knowledge proofs and run faster on a blockchain.

Not long after this, ING took the solution a step further and introduced Zero-Knowledge Set Membership (ZKSM), described in [10], going beyond numerical data to include other types of information, like locations and names. For instance, banks

could validate that a new client lives in a country that belongs to the European Union, without revealing the country. Simply put, this technology allows information to be shared without revealing contextual details.

#### 4.5 Nuclear Disarmament Applications

Zero-knowledge methods have been devised originally for computational tasks but recently translated into use for a physical system. At the Department of Energy's (DOE) Princeton Plasma Physics Laboratory (PPPL) researchers have developed an experimental verification protocol for weapon dismantlement agreements [11]. Their method includes a system that can compare physical objects while potentially protecting sensitive information about the objects themselves.

The process to prove two objects are identical - potentially nuclear warheads – is as follows: the prover provides two radiographic films already exposed with the inverse image of one test object and place them in two individual sealed boxes. The objects are placed in front of the boxes and getting radiated. This operation is essentially equivalent to adding a positive image on top of a negative image. The verifiers accepts the proof if both images after radiated are flat gray – meaning there is a 50 percent probability that the objects are indeed identical - otherwise he/she rejects it. If multiple tests are run simultaneously and the inverse images are randomized to the transmission pattern of the test object, the probability that they are not identical falls even more. The proof is zero-knowledge because the verifier does not learn anything beyond the result of the proof.

#### 4.6 Other Use Cases of ZKPs

- Ethereum: Ethereum is also working with zk-SNARK proofs since its Byzantium update in 2017 [12].
- PIVX: PIVX is a Proof of Stake (PoS) blockchain-based cryptocurrency created in 2016. At its core, it relies on fungibility, transaction privacy, and community governance. PIVX utilize zero knowledge proofs via ZeroCoin protocols [13].

#### 4.7 Possible Use Cases of ZKPs

There are many areas that can be enhanced by using Zero-Knowledge protocols where trust is required and there are large incentives to cheat, such as:

**Chain Voting Models:** Voting can refer either to political elections or corporate voting, where shareholder participation is a longstanding economical pillar. In any voting procedure, security, anonymity and trust are of paramount importance since these parts are most likely to fail and participation might be lower in the possibility of censorship.

These issues can be resolved with a zero-knowledge method. The whole procedure can move on a public blockchain. Every eligible voter (or shareholder) can cast their

ballot without revealing their identity and even more they can ask for verification of their vote to ensure their ballot was counted.

**Running a computation and verifying its results:** In the last years there is a trend for research centers and enterprises to migrate their data to outside providers. This practice raise concerns about the integrity and confidentiality of computations conducted on this data. We can imagine for example, a research medical center who wishes to have access in a private data center that contains genetic information of individuals, in order to apply a new formula.

This begs the following question, how we can verify the computations of the formula and at the same time achieve it without disclosing patients' identity? A Zero-knowledge method could answer both of these contradictory objects. The data center can apply on behalf of the researchers the formula and prove them in zero-knowledge the result without compromising the individuals' confidentiality [14].

**Data Auditing:** More and more users and enterprises resort in data centers due to storage limitations or for specialized services. Data integrity and availability is of major concern for cloud storage services while users uploading their personal data together with authentication information. This means users have no longer possession of their data that may face risks like loss, corruption or the purchase from a third service or company. The same concerns apply also to distributed ledgers.

Auditing from a third party is critical to prove data centers, financial institutions and exchanges are complying with regulations like GDPR [23]. Profiting from zero-knowledge methods, we can construct protocol schemes to prevent the leakage of verified and private data, a problem that existing auditing protocols suffer from and can have devastating effects. The reader can find recent research in [15-16].

## 5 Threats and Vulnerabilities

Zero-knowledge proofs are by definition methods which satisfy the appropriate security features that interest us. Attacks and vulnerabilities could be found in the designing of a ZKP protocol or in the system resources that support the realization of such features. The former one is most likely to happen since such designing requires a higher level of technical and theoretical sophistication that in the process a mathematical mistake might occur. Such mistakes could go completely unobservable like the “Infinite Counterfeiting Vulnerability<sup>4</sup>”, a mathematical false in a research zk-SNARKs paper that could irreparably damage the market since other cryptocurrencies employ the same algorithms.

---

<sup>4</sup> The story behind it is very interesting and can be found here <https://dci.mit.edu/research/2019/2/6/dci-director-was-interviewed-for-fortunes-latest-article-zcash-discloses-vulnerability-that-could-have-allowed-infinite-counterfeit-cryptocurrency>

## 5.1 Parameter Problem

zk-SNARKs have undergone significant tinkering and exploration into their real-world application and efficiency improvements, yet is not a perfect cryptographic tool.

For the zk-SNARKs to work, an initial setup phase is required where the so called “system parameters” are generated who act as a common reference string shared between prover and verifier and need to be built-in in advanced in every zk-SNARK implementation. This process is known as a “trusted setup” which is a highly polarizing event. If the parameters get compromised, a malicious user could use it to generate fake proofs and theoretically create infinite amounts of counterfeit coins of the native token (not only for Zcash but for any other crypto currency that adopts the zk-SNARK technology) without anyone knowing.

We can imagine a hypothetical scenario where elections take place on a blockchain using zk-SNARKs for proving votes. A powerful entity – such as a politician in the case of a Presidential vote – would have strong incentives to involve the parties with the setup of the zk-SNARK system to share the setup parameters. The entity could tilt the election in their intended direction by creating false proofs for votes taking place.

The biggest problem with the zk-SNARK approach is that users need to implicitly trust in the setup phase and the parties involved to setup the system honestly. Users of the system will never actually know if the setup phase was compromised at the point of setup, or at some point in the future. In other words we could say that the system is as much secure as the incentives to circumvent the system are low. So, if this is the case, the door remains open for a system where users do not implicitly need to trust the parties involved in the system’s setup to be honest.

## 5.2 Possible Attacks

While cryptocurrencies increase their use as an actual currency and payment method so the interest for Simple Payment Verification (SPV) increases to support users who cannot hold the full blockchain ledger in their mobile devices. As cryptocurrencies gear towards portability, software designers have not given much attention to system integrity issues, thus system vulnerabilities are often unobservable. One of these dangers are fault attacks<sup>5</sup>, which potentially could extract data from the CPU or memory of a device at the time a ZKP protocol is taking place.

In [17] the authors demonstrate the first (at the time) fault attack which is initiated purely from software – it removes entirely the physical access. A malicious user could unleash various vector attacks or a combination of them to induce faults in a CPU core while a ZK protocol is under verification. An attacker could fault the steps - and the respective data - which a verifier and a prover follow in a ZK protocol, and intercept (modify or destroy) the messages from one another.

Researchers have identified a group of attacks based on the discovery of subliminal channels in cryptographic primitives that are used for hidden transactions. These

---

<sup>5</sup> A fault attack is an attack on a physical electronic device which consists in stressing the device by an external mean in order to generate errors in such a way that these errors lead to a security failure of the system.

channels can allow malicious prover to embed tagging information about the user into each transaction, thus attacking the main purpose of the zk-SNARK information hiding [22]

## 6 Solutions

One way the Zcash team got around the parameter problem was to create a multi-party computation ceremony [18-19]. During the parameter creation a set of random numbers are produced, a process similar to the setup phase of a public-key cryptosystem. Zcash team refers to these random numbers as “toxic waste” to emphasize the need to get disposed with extreme caution. Multiple independent parties involved collaboratively in the construction of the parameters. It is apparent that it takes all the participants of the “ceremony” to be compromised or be dishonest in order to give away the parameters.

Another way to remedy the parameter problem is the choice users have to send tokens privately or publicly. Zcash concerning, the privacy features are not obligated, but are rather customizable.

More powerful constructions of zk-SNARKS (and generally non-interactive ZKPs) can rise to the challenge of the parameter setup. One of these, already mention above, is zk-STARKS. Another one is zk-ConSNARK which is developed by Suterusu<sup>6</sup>. Suterusu integrates a state of the art zero-knowledge schemes which are scalable and free of complex multi-party setup protocols. ConSNARKs can produce a sound blockchain ecosystem, absence of manipulation for the users and massively improved efficiency.

## 7 Discussion

We start our discussion by noticing first some technical details. The maximum rate, at which a blockchain protocol/technology is processed, is determined by the size of it, the size of the transaction and the intricacy of the underlying computations, consequently this is determined by the ZKP protocol/technology the blockchain adopts.

For example the complexity in terms of size of zk-STARKs rises much slower than the zk-SNARKs for one-time setup, after this phase SNARKs have much less size needed for computation in verifying the proof in faster times. On the other hand zk-ConSNARKs allows for very small and constant size proof computations, which leads to much faster generation and verification times. We could argue that as zero-knowledge technologies evolve, they push for smaller and smaller and/or constant size of proofs that can succeed better verification times [20].

In respect to privacy and confidentiality, most zero-knowledge technologies adopt zk-SNARKS implementations but all require a trusted setup event, meaning it is a one-time event, if a vulnerability or mathematical mistake is to be found, a whole new multi-party ceremony needs to be deployed which is an extremely complex procedure. There

---

<sup>6</sup> <https://www.suterusu.io/>

are few zero-knowledge breakthroughs that have got rid of the obligation of an initial setup.

We can argue, in the context of a zero-knowledge protocol embedded in a system, that one rule is applied; the system is only as good as the secret it is trying to conceal.

## 8 Does It Make The Cut? (Conclusions)

So much of our world is dominated by services that gobble up our personal information, they abuse these data, they sell it to the highest bidder, they do not protect in on their own servers and they essentially leaving it out for ransom. It is vital now more than ever before the need for privacy. Zero-Knowledge innovations, like zk-SNARKs, are up to the task of preserving the confidentiality and the security of users' data. They have the potential to enable trust levels that have never been achieved before. On the other side, since ZKPs have been theorized in 1985, it is the last six or so years we begun to use it in practice. We are still experimenting and try to understand how most effectively to apply it. It might take years until we manage to harness its true potential.

We still have to overcome many challenges and to observe a broader range of applications, but there is no doubt we hold in our hands a novel class of technology, one that sparks further development and innovation.

## References

1. S. Goldwasser, S. Micali and C. Rackoff, "The knowledge complexity of interactive proof-systems," in Proc. of the seventeenth annual ACM symposium on Theory of computing, 1985, pp. 291-304, doi: <https://doi.org/10.1145/22145.22178>
2. Quisquater JJ. et al., "How to Explain Zero-Knowledge Protocols to Your Children", in Advances in Cryptology — CRYPTO' 89 Proceedings, Lecture Notes in Computer Science, vol. 435, pp. 628-631, doi: [https://doi.org/10.1007/0-387-34805-0\\_60](https://doi.org/10.1007/0-387-34805-0_60)
3. N. Zhu, Understanding Zero-knowledge proofs through illustrated examples, Apr. 9, 2019. [Online]. Available: <https://blog.goodaudience.com/understanding-zero-knowledge-proofs-through-simple-examples-df673f796d99>
4. U. Fiege, A. Fiat, and A. Shamir, "Zero knowledge proofs of identity," in Proc. of the nineteenth annual ACM symposium on Theory of computing (STOC '87), Association for Computing Machinery, pp. 210-217. doi: <https://doi.org/10.1145/28395.28419>
5. A. Fiat, A. Shamir, "How To Prove Yourself: Practical Solutions to Identification and Signature Problems," in Odlyzko A.M. (eds) Advances in Cryptology — CRYPTO' 86, vol. 263, pp. 186-194. doi: [https://doi.org/10.1007/3-540-47721-7\\_12](https://doi.org/10.1007/3-540-47721-7_12)
6. Zcash, What are zk-SNARKs?, [Online]. Available: <https://z.cash/technology/zksnarks/>
7. A. Biryukov and I. Pustogarov, "Bitcoin over Tor isn't a Good Idea" 2015 IEEE Symposium on Security and Privacy, San Jose, CA, 2015, pp. 122-134, doi: <https://doi.org/10.1109/SP.2015.15>
8. Ben-Sasson, Eli et al. "Scalable, transparent, and post-quantum secure computational integrity," in IACR Cryptol. ePrint Arch, 2018, p. 46
9. ING, "ING launches Zero-Knowledge Range Proof solution, a major addition to blockchain technology".

10. J. Camenisch, R. Chaabouni and shelat, "Efficient Protocols for Set Membership and Range Proofs," in Pieprzyk J. (eds) *Advances in Cryptology*, Springer, Berlin, Heidelberg, 2008, vol. 5350, pp. 234-252. doi: [https://doi.org/10.1007/978-3-540-89255-7\\_15](https://doi.org/10.1007/978-3-540-89255-7_15)
11. S. Philippe, R. J. Goldston, A. Glaser et al. "A physical zero-knowledge object-comparison system for nuclear warhead verification," *Nature Communications* 7, no. 12890, Sep. 2016. Available doi: <https://doi.org/10.1038/ncomms12890>
12. Ethereum, Byzantium HF Announcement, Ethereum Team, Oct. 12, 2017. Available: <https://blog.ethereum.org/2017/10/12/byzantium-hf-announcement/>
13. PIVX Announcement, Road to PIVX 5.0 with zk-SNARKs, pivx.org Available: <https://pivx.org/road-to-pivx-5-0-with-zk-snarks/>
14. E. Ben-Sasson, A. Chiesa, M. Green, E. Tromer and M. Virza, "Secure Sampling of Public Parameters for Succinct Zero Knowledge Proofs," 2015 IEEE Symposium on Security and Privacy, San Jose, CA, 2015, pp. 287-304, doi: <https://doi.org/10.1109/SP.2015.25>
15. Z. Yan, H. Hongxin, A. Gail-Joon and S. Y. Stephen, "Efficient audit service outsourcing for data integrity in clouds," *Journal of Systems and Software*, Vol. 85, Issue 5, 2012, pp. 1083-1095. doi: <https://doi.org/10.1016/j.jss.2011.12.024>
16. N. Narula, W. Vasquez and M. Virza, "zkLedger: Privacy-Preserving Auditing for Distributed Ledgers," in *Proceedings of the 15th USENIX Symposium on Networked Systems Design and Implementation*, USENIX Association, 2018, pp. 65-80. Available: <https://www.usenix.org/conference/nsdi18/presentation/narula>
17. A. Tang, S. Sethumadhavan and S. Stolfo, "CLKSCREW: Exposing the Perils of Security-Oblivious Energy Management," in *Proceedings of the 26th USENIX Security Symposium*, USENIX Association, 2017, pp. 1057-1074. Available: <https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/tang>
18. Zcash, Parameter Generation, [Online]. Available: <https://z.cash/technology/paramgen/>
19. S. Bowe, A. Gabizon and M. D. Green, "A multi-party protocol for constructing the public parameters of the Pinocchio zk-SNARK," in *Cryptology ePrint Archive*, Report 2017/602, 2017. Available: <https://eprint.iacr.org/2017/602>
20. R. Mannak, Comparing General Purpose zk-SNARKs, medium.com, Nov. 11, 2019. Available: <https://medium.com/coinmonks/comparing-general-purpose-zk-snarks-51ce124c60bd>
21. Ruffing, T., Thyagarajan, S. A., Ronge, V., & Schroder, D. (2018, June). (Short Paper) Burning Zerocoins for Fun and for Profit-A Cryptographic Denial-of-Spending Attack on the Zerocoin Protocol. In 2018 Crypto Valley Conference on Blockchain Technology (CVCBT) (pp. 116-119). IEEE.
22. Biryukov, A., & Feher, D. (2019, June). Privacy and linkability of mining in zcash. In 2019 IEEE Conference on Communications and Network Security (CNS) (pp. 118-123). IEEE.
23. Aliyu, A., Maglaras, L., He, Y., Yevseyeva, I., Boiten, E., Cook, A., & Janicke, H. (2020). A Holistic Cybersecurity Maturity Assessment Framework for Higher Education Institutions in the United Kingdom. *Applied Sciences*, 10(10), 3660.

# Application of statistical functions for microstructure characterization and determination of elastic properties of ceramic foam

Vinit Vijay Deshpande<sup>1</sup>, Romana Piat<sup>1</sup> and Kay André Weidenmann<sup>2</sup>

<sup>1</sup> Department of Mathematics and Natural Sciences, University of Applied Sciences Darmstadt, Schöfferstraße 3, Darmstadt 64295, Germany

<sup>2</sup> Institute of Materials Resource Management (MRM), University of Augsburg, Universitätsstraße 1, Augsburg 86159, Germany

vinit-vijay.deshpande@h-da.de, romana.piat@h-da.de,  
kay.weidenmann@mrm.uni-augsburg.de

**Abstract.** Microstructural features of ceramic foam are numerically evaluated by employing statistical functions. X-ray computed tomography (CT) scans of ceramic foam are utilized to compute statistical functions like two-point correlation function, lineal path function, etc. These functions describe the microstructure of the foam. Segmentation algorithms are applied to isolate the voids and study their shape distribution within the sample. The statistical functions are further utilized to determine the correct size of the stochastic volume elements (SVEs) that can represent the entire foam sample. Ensemble averaging and size enlargement effects on different SVE sizes are evaluated. By comparing the statistical functions of the entire sample with that of the ensemble of correctly sized SVEs, a ranking method is developed to determine the SVE positions inside the sample that resemble the most with the entire sample. These SVEs are then adopted to determine the effective elastic properties of the foam sample. Finite element models of the selected SVEs are constructed and mixed uniform boundary conditions are applied to numerically determine the coefficients of effective stiffness tensor. Lastly, the obtained properties are compared with experimental results available in the literature.

**Keywords:** Ceramic foam, elastic properties, microstructure characterization, statistical functions, numerical modelling, X-ray computed tomography

## 1 Introduction

Nowadays, composite materials are extensively used in a variety of industries including but not limited to aerospace, automotive, medical devices, oil and gas, electronics, etc. Generally, a composite material consists of discontinuous reinforcement phase distributed within a continuous medium called matrix [1]. This reinforcement phase can be particles of various shapes and sizes, fibers or whiskers. An exception to this definition

is a type of composite called interpenetrating phase composite. In these composites, both the phases are continuous and hence cannot be differentiated into matrix and reinforcement. These composites are especially beneficial when there is a need for a material to possess contradictory properties [2]. Their manufacturing involves producing an open porous preform which is then infiltrated by for example a molten metal [3]. The manufacturing method and microstructure of this preform are highly important parameters as they govern the final distribution of phases in the composite.

In this work, an alumina preform manufactured via a slurry base route [3] is studied numerically with an objective to quantify its microstructural features and to use them in determining its effective elastic properties. The preform microstructure is highly porous with voids resembling spheres that are connected to each other, hence the name 'foam'. These voids are characterized by isolating them using image processing algorithms. They are then studied to check their resemblance to spheres and their orientation distribution within the sample space.

Torquato et al [4, 5] has given detailed description of various statistical functions that are used to describe the microstructure of heterogeneous materials. These functions are  $n$ -point correlation functions each uniquely describing the distribution of phases inside the material. They are used in the present work to describe the distribution of porosity within the foam sample. They are calculated for the entire sample and are used to derive important observations on material symmetry and homogeneity of the sample.

In order to evaluate the effective elastic properties of the foam, it is important to determine the appropriate size of a volume element (VE) within the sample that can act as a stochastic volume element (SVE). Statistical functions are used to determine this size by dividing the foam sample into smaller volume elements of different sizes. These functions are then calculated for each VE of each size and the size effects are studied in terms of ensemble averaging and sample enlargement. This study results in determination of a VE size that can act as SVE to the foam sample.

Ensemble average of the effective elastic properties of this SVE can give a good estimate of the effective properties of the entire sample in question. But this will lead to considerable computational expenses as number of VE realizations for selected SVE size would be large. To reduce this number, a ranking method is developed in which VEs are ranked based upon how close their statistical functions are to that of the entire sample. Based upon this, a relatively small number of VEs are selected for averaging.

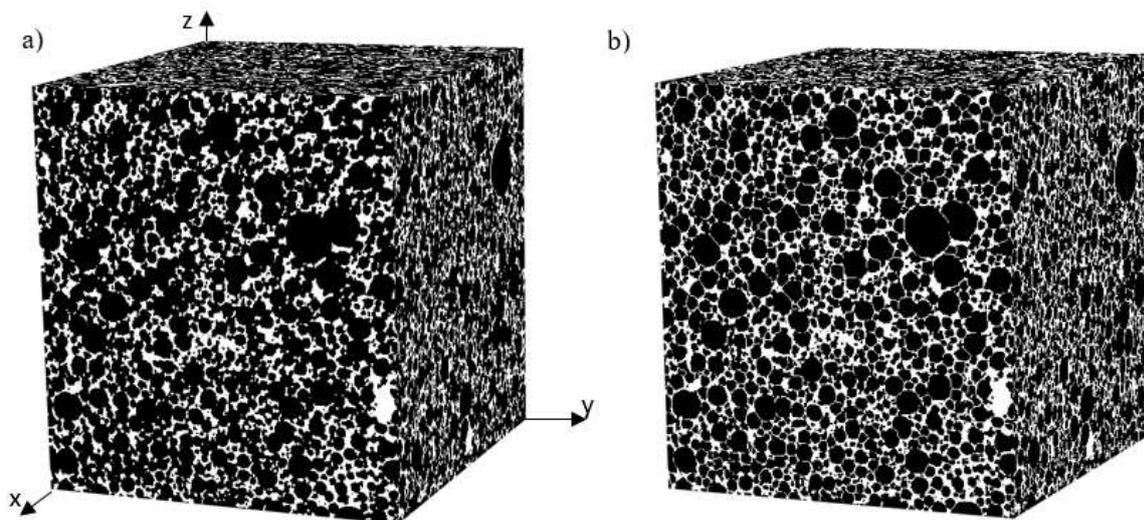
Finite element methods are predominantly used to solve structural mechanics problems numerically. These methods involve discretizing the continuous media into smaller defined shapes called elements within which field variables (stress, strain, displacement, etc) are interpolated using shape functions. They are widely used for evaluating effective material properties of heterogeneous media. The ensemble of VEs selected from the ranking method are meshed and mixed uniform boundary conditions [6] are applied to determine effective linear elastic stiffness coefficients of each VE. Following ensemble averaging, we get effective properties of the entire foam sample.

Lastly, the procedure described above is validated by comparing the calculated effective properties with experimental measurements [3]. The comparison shows that statistical functions can be used to define appropriate size of SVEs and the ranking method helps to reduce the computational effort which would have otherwise required.

The article is organized into following sections: Section 2 describes the image processing steps performed on microcomputed tomography ( $\mu$ CT) scans of the foam sample, segmentation of pores and their shape distribution within the sample. Section 3 includes definition of statistical functions and their evaluation for the entire sample. Section 4 describes the procedure to select appropriate size of SVE. Section 5 describes the ranking procedure. Section 6 describes finite element calculations to determine effective elastic properties followed by comparison with experimental results. The discussion on the entire procedure is in section 7 followed by conclusions in section 8.

## 2 Image Processing of $\mu$ CT Data

A cubic sample of dimensions  $V \approx 5 \times 5 \times 5 \text{ mm}^3$  was scanned using X-ray computed tomography. To avoid artifacts like beam hardening on the boundaries of the cube, a smaller region in the interior of the sample with dimensions  $V_i \approx 1.89 \times 1.76 \times 1.94 \text{ mm}^3$  was selected. Details regarding the image processing can be found in [3]. The  $\mu$ CT data is available in the form of 2D cross sections of the foam sample with each cross section being a grayscale image. These images are stacked one on top of other to form a 3D volumetric map. This is followed by removal of any noise by using median filter. Next, a global threshold in terms of a scalar luminance value is determined which is used to binarize the 3D image. This is done iteratively by altering the threshold value such that after binarization, the final porosity obtained by counting pore voxels matches that obtained by experimental measurements. The porosity of the foam sample obtained by density measurements was 74.5% [3]. Hence the global threshold value is decided such that the porosity in binarized 3D image matches this value. After binarization, a 3D image with pore regions marked as '0' (black color) and alumina region marked as '1' (white color) is obtained. Lastly, an area opening operation is performed in which all the connected regions of alumina phase having volume less than 10 voxels are removed. This removes any hovering alumina regions lying within the pore phase (Fig. 1a).

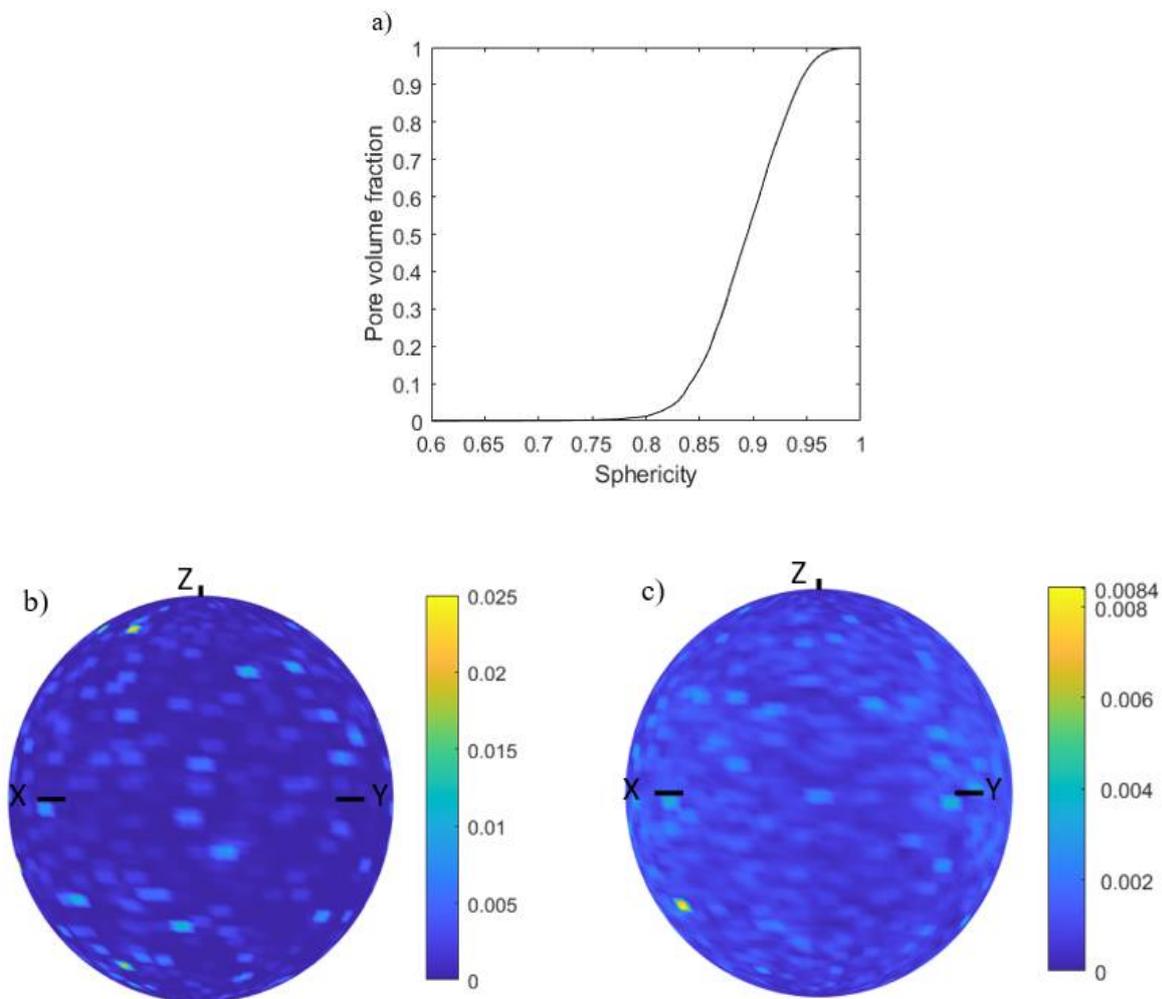


**Fig. 1.** a) Binarized 3D image before segmentation and b) after segmentation

It can be seen from Fig.1a that the pores are interconnected to each other. In order to isolate them, a segmentation method called watershed algorithm [7] is applied. The resulting image is shown in Fig.1b. Thereafter, the individual pores are studied separately to calculate their volume, surface area and orientation. Sphericity of each pore is also calculated which quantifies how close the shape of the pore is with respect to a sphere. It is defined as the ratio of surface area of an equal-volume sphere to the actual surface area of the pore [8]. Its expression is given as:

$$\psi = \frac{\pi^{1/3} (6V_p)^{2/3}}{A_p}. \quad (1)$$

Here,  $V_p$  is volume of pore and  $A_p$  is surface area of pore. Fig. 2a shows cumulative distribution of pore sphericity with pore volume fraction. Pore volume fraction is defined as volume fraction of a particular pore with respect to all the pores present in the sample. Note that the pores that lie at the boundary of the foam sample are not considered in this study because of lack of information about the entire pore. It can be seen that all the pores have sphericity greater than 0.75 and 90% of the pores have sphericity greater than 0.86. The maximum sphericity is 0.97.



**Fig. 2.** a) Cumulative distribution of pore sphericity, b) orientation distribution of pores having sphericity  $0.75 < \psi < 0.85$  and c) orientation distribution of pores having sphericity  $0.85 < \psi < 0.95$

Orientation of each pore is calculated by approximating the pore as an equivalent ellipsoid using principal component analysis [9]. This is done using the ‘regionprops3’ function in MATLAB R2019b [10]. Spherical angles are then calculated for each pore from the eigenvectors obtained from the above MATLAB function. An orientation distribution function is plotted in the form of a spherical plot so that the distribution of pore orientations with respect to coordinate axes can be visualized. The pores are segregated into two sets, one which have sphericity in the range of 0.75 to 0.85 and other which have sphericity in the range of 0.85 to 0.95. Figs.2b and 2c show the orientation distribution of these pores respectively with the colormap showing pore volume fraction for each orientation.

### 3 Microstructure Studies Using Statistical Functions

In the theory of modelling random media, a wide variety of statistical functions have been used to define the distribution of phases within the heterogeneous medium [5]. This section describes five such functions that are employed in this article.

#### 3.1 Two-point Correlation Function

Consider a n-dimensional 2 phase microstructure in which phase  $i$  occupies volume fraction  $v_i$ . An indicator function is defined such that:

$$I^{(i)}(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in V_i \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where,  $V_i$  is the region occupied by the phase  $i$ . The two-point correlation function is defined as:

$$S_2^{(i)}(\mathbf{x}_1, \mathbf{x}_2) = \langle I^{(i)}(\mathbf{x}_1) I^{(i)}(\mathbf{x}_2) \rangle. \quad (3)$$

It is defined as the probability of finding two points at positions  $\mathbf{x}_1$  and  $\mathbf{x}_2$  in the same phase in the medium. The brackets indicate ensemble average. For statistically homogeneous and isotropic medium, the two-point correlation function depends only on the magnitude of the distance between the two positions  $r = |\mathbf{x}_1 - \mathbf{x}_2|$ . Hence it can be expressed in the form  $S_2(r)$ . This function gives an indication of the distribution of the phase within the medium.

#### 3.2 Two-point Cluster Correlation Function

It is defined as the probability of finding two points at positions  $\mathbf{x}_1$  and  $\mathbf{x}_2$  in the same cluster (region of connected voxels of the same color) of the phase of interest in the medium. For statistically homogeneous and isotropic medium, this function depends only on the magnitude of the distance between the two positions  $r = |\mathbf{x}_1 - \mathbf{x}_2|$ . Hence it can be expressed in the form  $C_2(r)$ . This function is a superior descriptor in the sense that it gives an idea of the connectedness of the phase of interest [11].

### 3.3 Lineal Path Function

It is defined as the probability of finding a line segment spanning from  $\mathbf{x}_1$  to  $\mathbf{x}_2$  that lies entirely in the phase of interest. The function contains some connectedness information along the lineal path (length of the segment) and hence contains certain long-range information about the medium. For statistically homogeneous and isotropic medium, the lineal path function depends only on the magnitude of the distance between the two positions  $r = |\mathbf{x}_1 - \mathbf{x}_2|$ . Hence it can be expressed in the form  $L(r)$ .

### 3.4 Pore Size Distribution Function

The pore size distribution function,  $P(\delta)$  is defined in such a way that  $P(\delta)d\delta$  is the probability that a randomly chosen point in the pore phase (any phase of interest) lies at a distance between  $\delta$  and  $\delta + d\delta$  of the nearest point on the pore-solid interface. This function contains connectedness information about spherical regions in the pore phase. It can only be obtained from 3D images of the medium.

### 3.5 Cumulative Pore Size Distribution Function

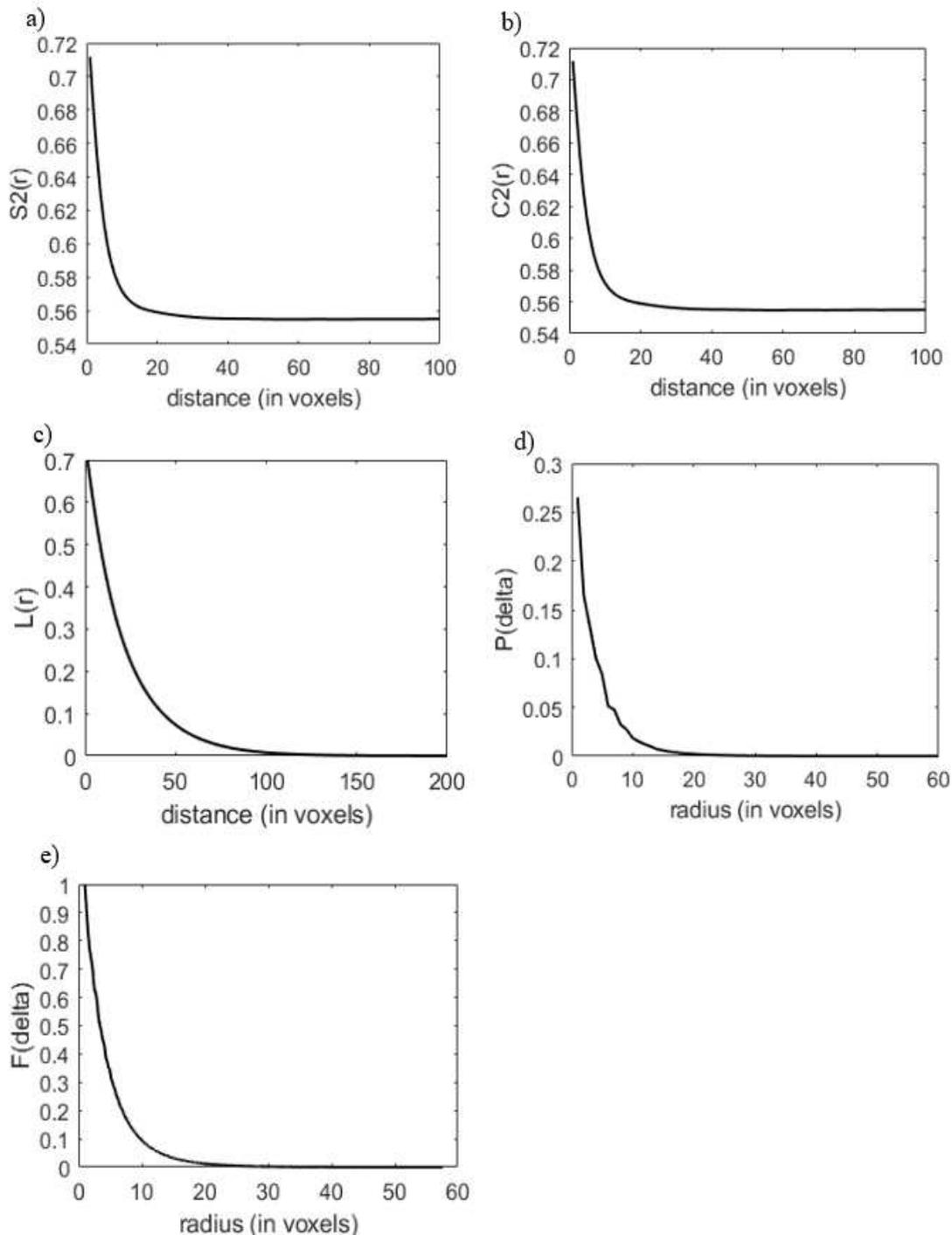
It is defined as the probability  $F(\delta)$  of the sphere of radius  $\delta$  having its centre in the pore phase lie entirely in the pore phase. It is the fraction of the pore space that has pore diameter greater than  $\delta$ .

Along with these functions, two one-point correlation functions called volume fraction and percolation volume fraction are also used in this article. Pore regions can either be connected or disconnected to each other. The fraction of the pore phase that percolates (connects) over the total volume of the phase in the medium is termed as percolation volume fraction. It is defined as ratio of the volume of largest cluster of connected pores to that of entire pore volume present in the medium.

For isotropic medium, it is sufficient to calculate  $S_2(r)$ ,  $C_2(r)$  and  $L(r)$ , only in the orthogonal directions [4]. This reduces the computational costs drastically as compared to brute force method in which these functions are calculated at all voxel positions and in all directions within the sample. Pore size distribution function can be directly evaluated by calculating the Euclidean distance transform of a binary 3D image. The result gives a 3D matrix in which each element is the distance between that voxel and the nearest nonzero voxel (or voxel from different phase) in the binary image. This matrix can be used to bin the voxels based upon the value they possess. The number of voxels in each bin is normalized by the total number of voxels in the phase of interest. A graph of these values with respect to corresponding bin value of  $\delta$  is plotted. For calculating cumulative pore size distribution function, the transformed matrix obtained while determining the pore size distribution function is used and the unique values of  $\delta$  are determined and stored in a column vector. Then at each position of the transformed matrix, all the values lower than its value in that position are selected. Then counters corresponding to each of those values are increased. This process is repeated at each voxel element of the transformed matrix. Finally, the values of each counter are normalized by the total number of voxels in the phase of interest. A graph of these values

with respect to the corresponding value of  $\delta$  is plotted. To calculate percolation volume fraction, the volume of each cluster determined while finding two-point cluster correlation function is calculated and divided by the total volume of the phase of interest. The maximum of these volume fractions is the percolation volume fraction.

For selected foam sample, all these functions are calculated using above described methods to study distribution of pore phase. Note that from here onwards the binary image before segmentation is used. The results are given in Figs. 3a-3e and Table.1.

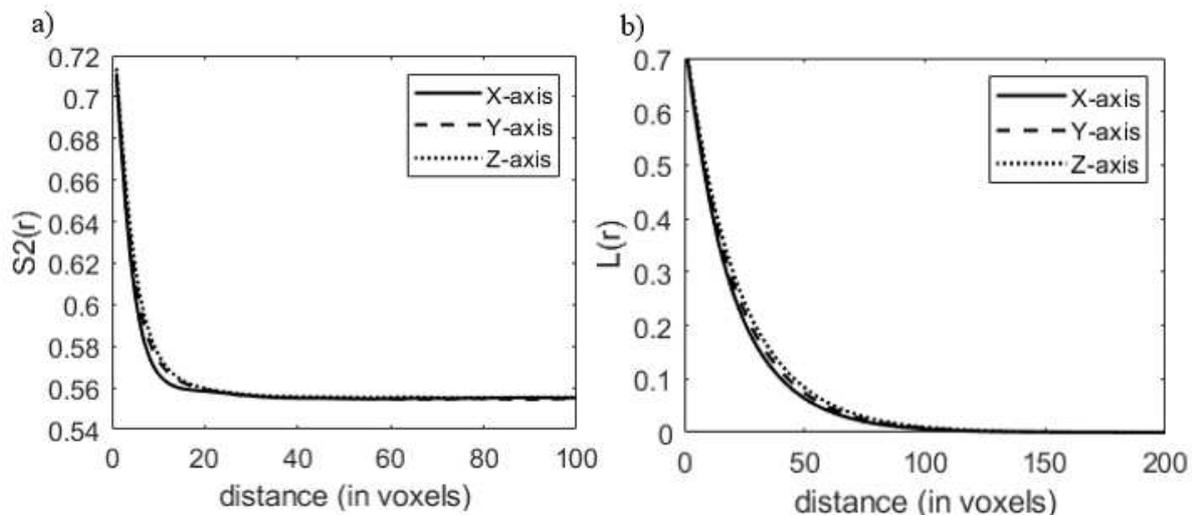


**Fig. 3.** a) Two-point correlation function; b) Two-point cluster correlation function; c) Lineal path function; d) Pore size distribution function; e) Cumulative pore size distribution function of the alumina foam sample.

**Table 1.** One-point correlation functions

Volume fraction $\nu$	0.745
Percolation volume fraction $\nu^*$	0.999

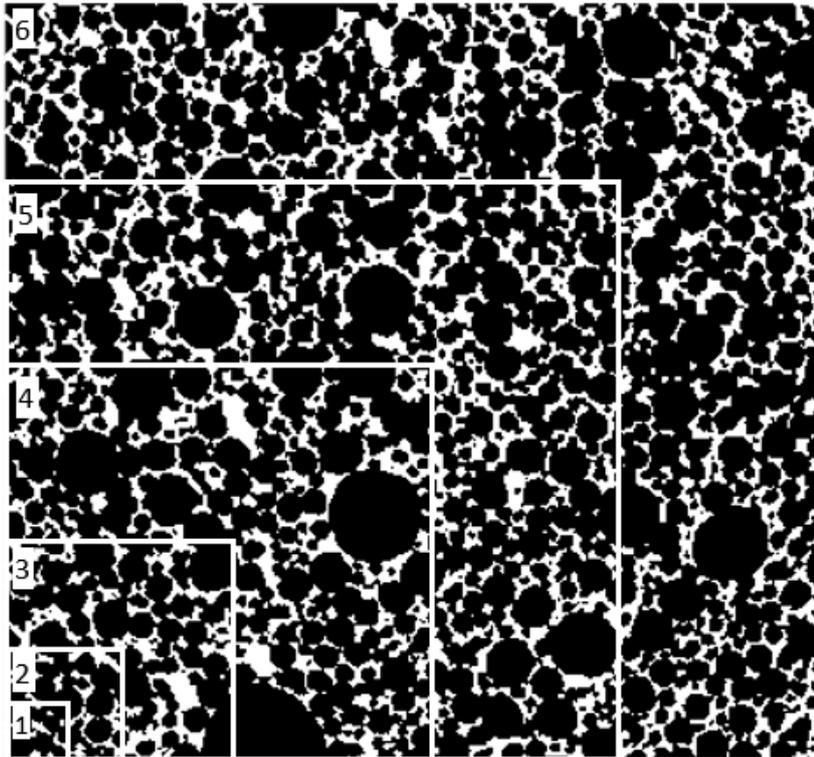
It is observed that for the selected microstructure, since the percolation volume fraction is close to 1, there is no difference between two-point correlation function and two-point cluster correlation function. Hence in the remainder of the article, cluster correlation function will not be used for studying the microstructure. Similarly, it is decided not to use pore size distribution function as cumulative pore size distribution function is smoother and contains all the required information about the pore phase that will be needed further. In order to check the isotropy of the foam sample, the functions  $S_2(r)$  and  $L(r)$  are calculated individually for each orthogonal direction and plotted to check if these functions vary in different directions. The results are given in Figs. 4a-4b.



**Fig. 4.** a) Two-point correlation function and b) Lineal path function in three orthogonal directions

## 4 Selection of SVE Size

The foam sample under question is too big to be used directly for finite element (FE) calculations. Hence appropriate size and position within the sample is to be chosen for further use in FE calculations. Since the material microstructure is random in nature, an ensemble of stochastic volume elements is to be found. For this purpose, the foam sample is divided into smaller regions and statistical functions for realizations of each size are calculated. Fig. 5 shows different sizes considered in this study. Note that each volume element (VE) is cuboidal in shape. For ease of representation it is shown in two dimensions.

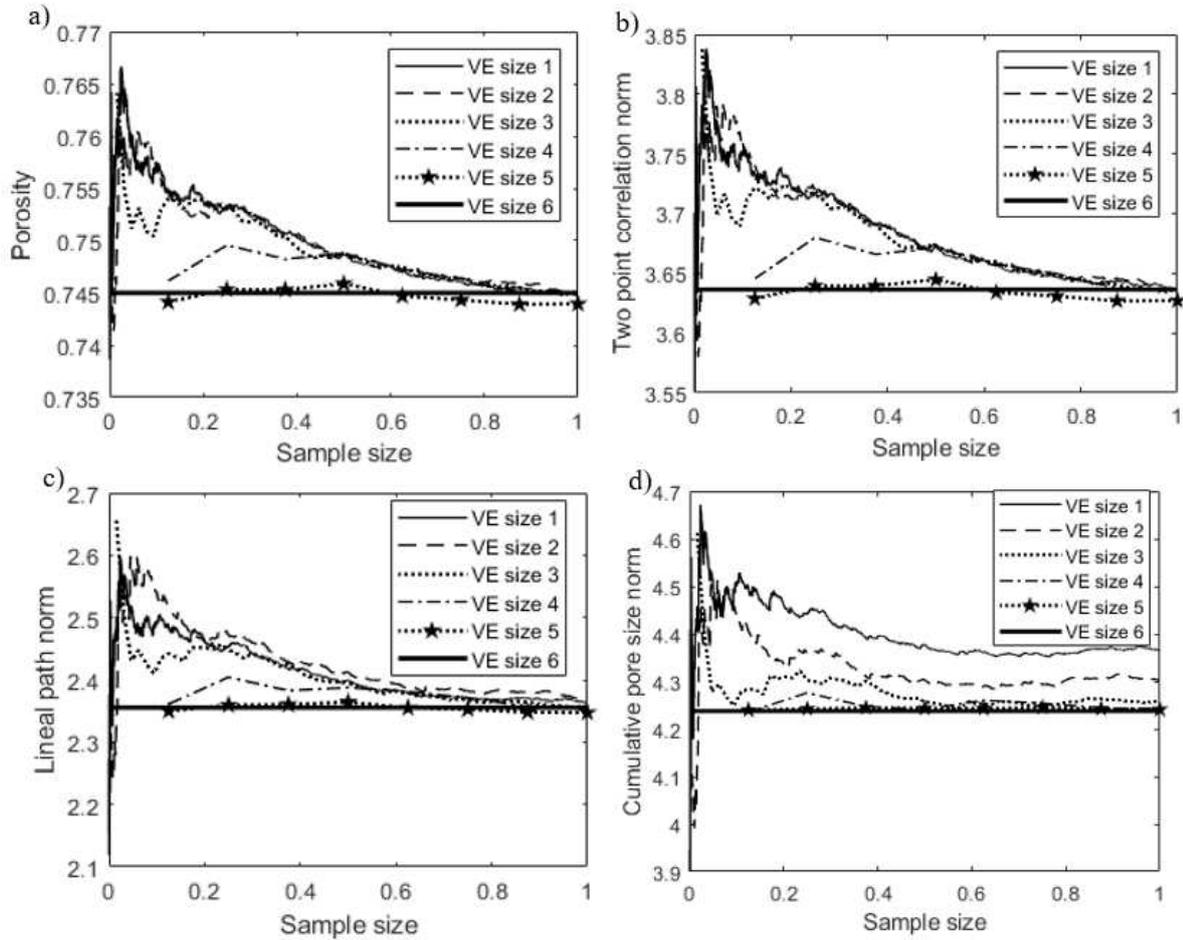


**Fig. 5.** Different sizes of VEs

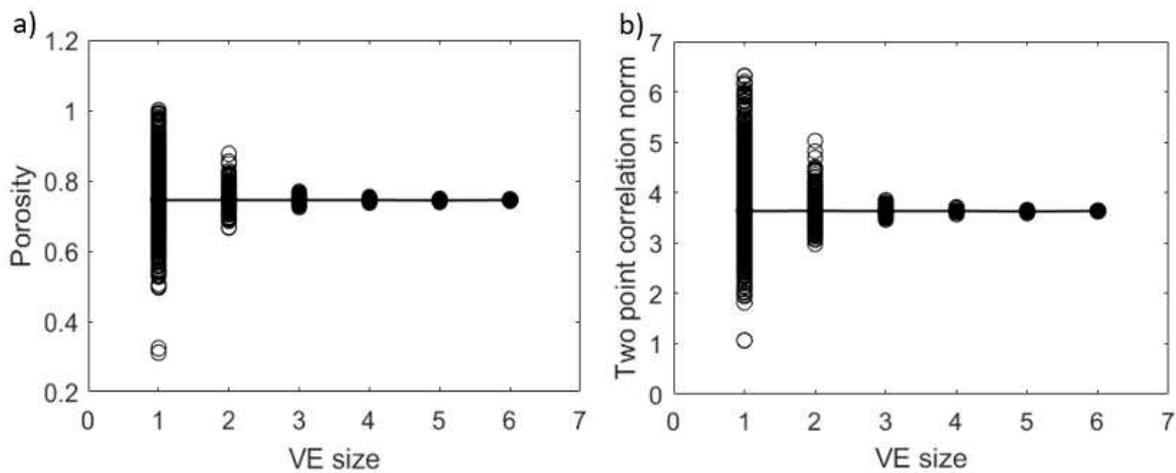
**Table 2.** Dimension and number of realizations of each VE size

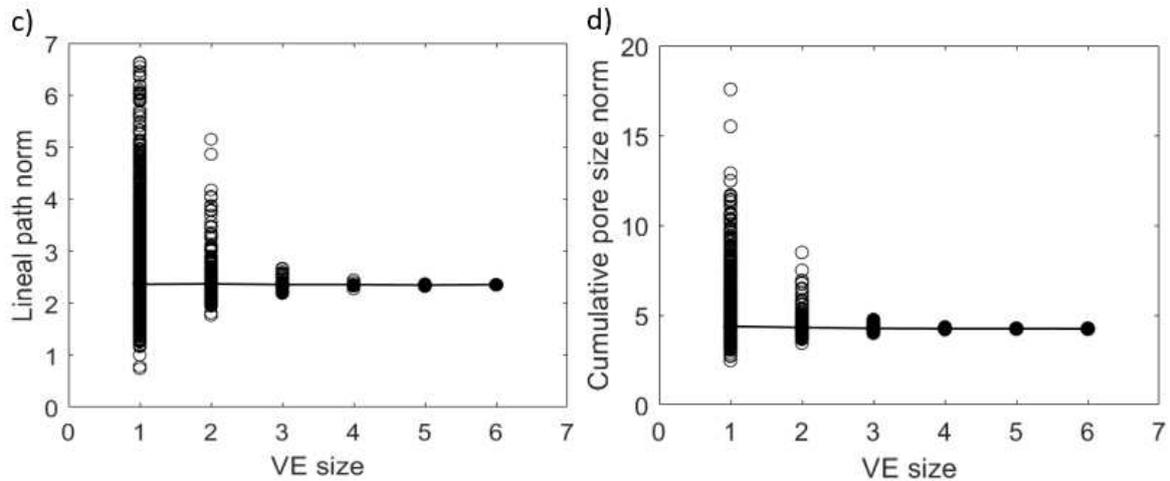
VE Size	Relative edge length	No. of realizations
1	1/16	4096
2	1/8	512
3	1/4	64
4	1/2	8
5	3/4	8
6	1	1

Edge length of each VE size in terms of edge length of the entire sample (size 6) is given in Table.2. Realizations for sizes 1 ,2, 3 and 4 are formed by shifting the VE domain one edge length at a time in all three directions. Realizations for size 5 are formed by considering 2 realizations in each direction such that all the sample space is utilized. Since the VEs are stochastic, the number of VEs in ensemble of each size plays a critical role in determining any useful conclusions from the VEs. The Figs. 6a-6d show the effect of number of samples in the ensemble of each VE size on averaged statistical functions. Here Euclidean norm of each statistical function for each VE is calculated. Since sample size (number of realizations) in ensemble of each VE size is different, the sample size plotted on X-axis is normalized with respect to the total number of realizations for each VE size. Figs.7a-7d show scatter of norms of statistical functions for each VE size. The solid line indicates mean value for each VE size.



**Fig. 6.** Effect of sample size on a) porosity; norm of b) two-point correlation function; c) lineal path function and d) cumulative pore size distribution function for each VE size



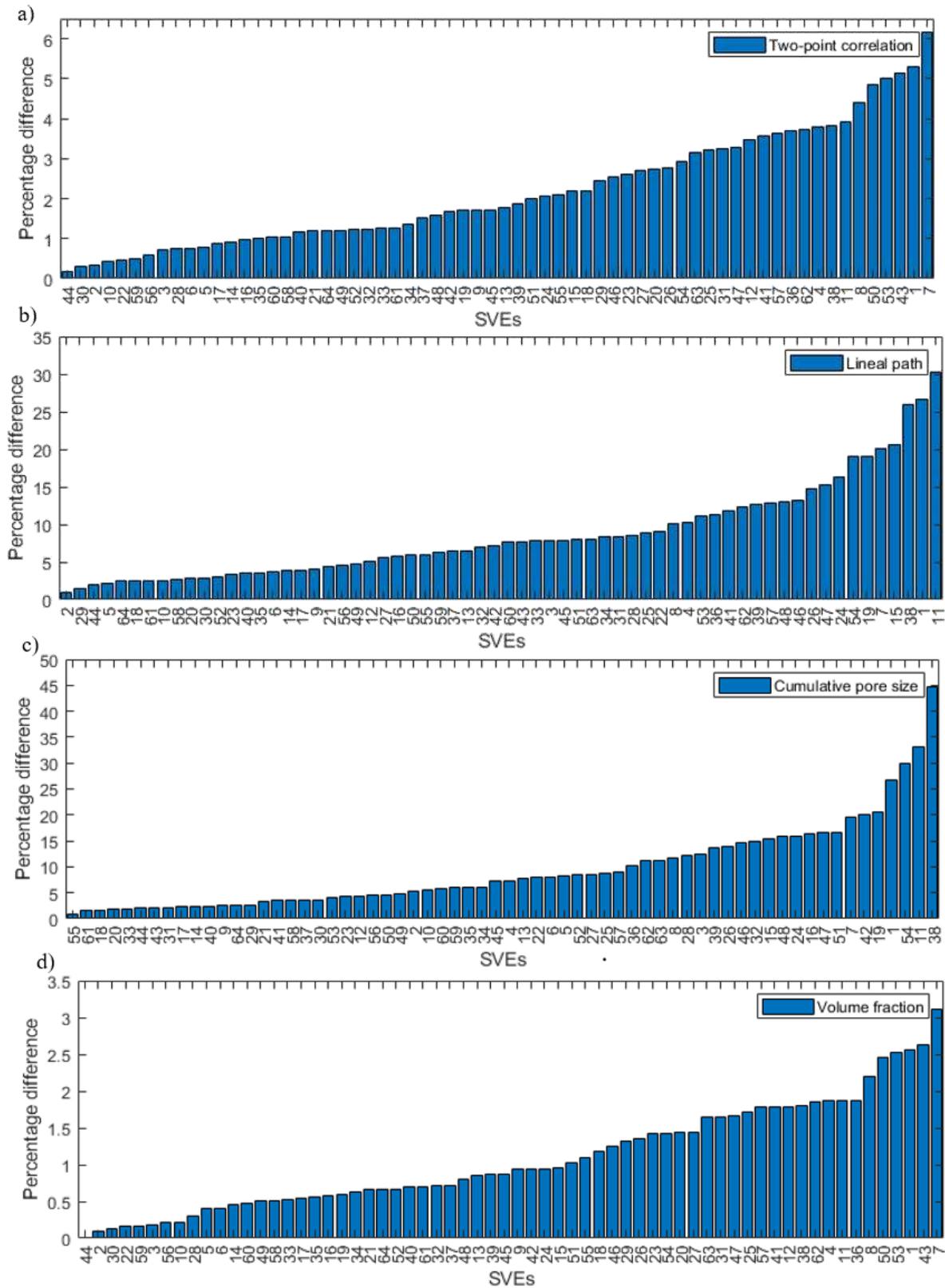


**Fig. 7.** Scatter of a) porosity; norm of b) two-point correlation function; c) lineal path function and d) cumulative pore size distribution function for each VE size along with their average values

It can be seen from Figs.6a-6c that for each VE size, the ensemble average approaches the value of the entire sample (size 6) when all the realizations are taken into account. Since the number of realizations increase as VE size decreases, a greater number of realizations are needed in the ensemble average to reach the value of entire sample as the VE size is decreased. Fig. 6d shows that the ensemble average of cumulative pore size norm approaches that of the entire sample for size 3 and above. The scatter in the results of each VE size is shown in Figs. 7a-7d. Mean value of each VE size equals the value of the entire sample after utilizing all the realizations in each ensemble. Hence choice of appropriate VE size for FE calculations depends upon the size of VE that can be handled by the available computational resources and the number of realizations. For further studies in this article, VEs of size 3 are taken as stochastic volume elements (SVEs). It is because, for this size, the ensemble average of all the four statistical functions converge to that of the entire sample (Figs.6a-6d) and it gave results with acceptable computational expenses.

## 5 Ranking of SVEs

In the previous section, it was decided to use VE size 3 as SVE in further calculations of effective elastic properties. A straight forward way to do this is to calculate effective properties of all the realizations in the ensemble of size 3 and then calculate ensemble average of the effective properties. However, this would need significant computational expenses. Statistical functions can be used here to reduce the number of realizations used in the ensemble averaging. Here, absolute value norm of the percentage difference between statistical functions of each SVE in the ensemble and that of the entire foam sample is calculated. For each SVE, functions  $S_2(r)$ ,  $L(r)$ ,  $F(\delta)$  and volume fraction are evaluated. These results are then rearranged in ascending order of SVEs and plotted in Figs. 8a-8d.



**Fig. 8.** Ranking of SVEs for a) two-point correlation function; b) lineal path function; c) cumulative pore size distribution function and d) porosity (pore volume fraction)

It can be seen that for each statistical function, the ranking of SVEs varies. Hence it is decided to use the SVEs that lie within 5% value for all statistical functions. 5 SVEs are obtained that satisfied this criterion. They are SVE no. 14, 17, 30, 44 and 58. Hence,

instead of using all 64 realizations of the SVEs, these 5 SVEs are selected for ensemble averaging of effective elastic properties.

## 6 Determination of Effective Elastic Properties

The problem of determination of effective properties is based upon the idea that a heterogeneous medium can be converted into a homogeneous medium by utilizing the conservation of energy principle. The criteria given by Hill [12] needs to be satisfied:

$$\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle = \langle \boldsymbol{\sigma} \rangle : \langle \boldsymbol{\varepsilon} \rangle. \quad (4)$$

It says that average of the scalar product of stress  $\boldsymbol{\sigma}$  and strain  $\boldsymbol{\varepsilon}$  tensors over the heterogeneous medium should be equal to the product of their individual averages. Using Gauss theorem, this condition can be generalized for heterogeneous materials [6] as:

$$\int_{\Gamma} (\mathbf{t}(\mathbf{x}) - \langle \boldsymbol{\sigma} \rangle \cdot \mathbf{n}) \cdot (\mathbf{u}(\mathbf{x}) - \langle \boldsymbol{\varepsilon} \rangle \cdot \mathbf{x}) d\Gamma = 0. \quad (5)$$

Where  $\Gamma$  is boundary of a VE and  $\mathbf{t}$ ,  $\mathbf{n}$ ,  $\mathbf{u}$  and  $\mathbf{x}$  are the traction, normal, displacement and position vectors respectively. This condition is satisfied by three types of boundary conditions [13, 14] namely kinematic uniform boundary condition (KUBC), stress uniform boundary condition (SUBC) and mixed uniform boundary condition (MUBC). [15] showed that KUBCs and SUBCs give bounds to the apparent stiffness tensor ( $\mathbf{C}$ ):

$$\mathbf{C}_{SUBC} \leq \mathbf{C}_{MUBC} \leq \mathbf{C}_{KUBC}. \quad (6)$$

Also,

$$\mathbf{C}_{SUBC} \leq \mathbf{C}_{eff} \leq \mathbf{C}_{KUBC}. \quad (7)$$

Where,  $\mathbf{C}_{eff}$  is the exact effective stiffness tensor of the heterogeneous medium. Using the fact that periodic boundary conditions (PBCs) give exact effective stiffness tensor of periodic microstructures, [6] showed that the periodically compatible mixed boundary conditions (PMUBCs) give effective stiffness tensor for non-periodic microstructures that match closely with that obtained by applying PBCs on the same sample by converting it into periodic. This conversion was done by mirroring the non – periodic sample about its three orthogonal planes. These PMUBCs are utilized in this article so as to obtain effective stiffness tensor of the five selected SVEs.

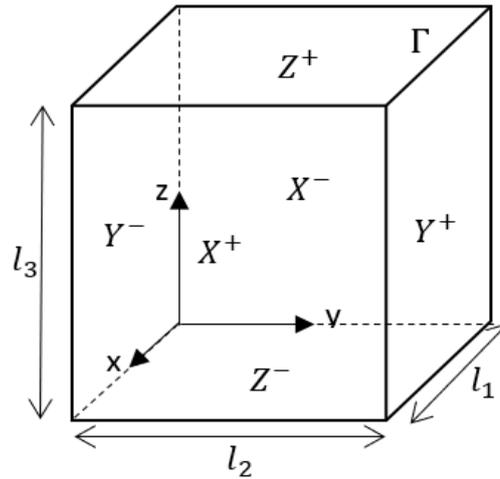
The SVE problem is defined as:

$$div(\boldsymbol{\sigma}) = 0 \text{ in } \Gamma. \quad (8)$$

Such that the boundary conditions satisfy:

$$\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle = \langle \boldsymbol{\sigma} \rangle : \langle \boldsymbol{\varepsilon} \rangle. \quad (9)$$

The coordinate system, dimensions and nomenclature of faces of  $\Gamma$  are given in Fig.9.



**Fig. 9.** Representation of  $\Gamma$

The problem is solved numerically by using finite element method. A commercial software called ABAQUS [16] is used for this purpose. Each SVE sample is meshed using linear tetrahedral elements such that the alumina phase is meshed and the pore phase is kept unmeshed. The elastic material properties of alumina [3] were taken as  $E=360.5\text{GPa}$  and  $\nu = 0.2$ . The effective stiffness tensor  $\mathbf{C}$  is expressed as:

$$\boldsymbol{\sigma} = \mathbf{C} : \boldsymbol{\varepsilon} . \tag{10}$$

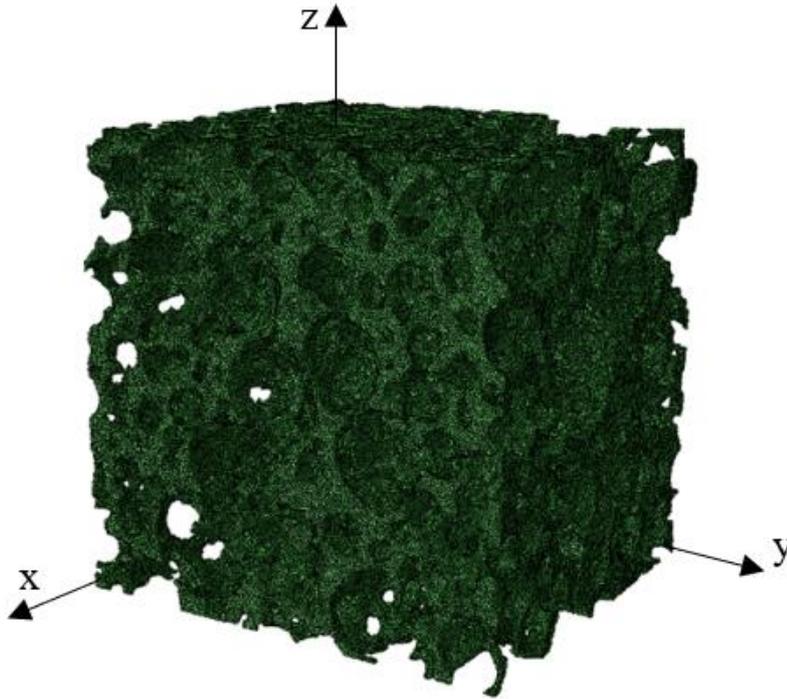
Writing in terms of the respective components, taking into account the symmetries of the tensors and using Voigt notation:

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{12} \\ \sigma_{23} \\ \sigma_{31} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{12} \\ 2\varepsilon_{23} \\ 2\varepsilon_{31} \end{bmatrix} \tag{11}$$

PMUBCs are given in Table.3. Note that  $1 \rightarrow x$ ,  $2 \rightarrow y$  and  $3 \rightarrow z$  in the description of boundary conditions. In the FE simulations, six load cases are defined. In each load case, one strain component with value 0.001 is applied according to Table.3. The corresponding stress tensor is calculated as average stress over the SVE using:

$$\langle \boldsymbol{\sigma} \rangle = \frac{1}{V_{\Gamma}} \int_{\Gamma} \boldsymbol{\sigma}(\mathbf{x}) d\Gamma(\mathbf{x}) . \tag{12}$$

From the results of each load case, each column of stiffness tensor is calculated. FE mesh of SVE 14 is shown in Fig. 10. To apply boundary conditions, all the nodes that lie on each face of the SVE are selected and their degrees of freedom are constrained according to Table.3. The results of the FE simulations are given in Table.4. The unit of stiffness coefficients is GPa. It also contains the results of experimental measurements of the same material referred from [3].



**Fig. 10.** Finite element mesh of SVE 14

**Table 3.** PMUBCs on each face of SVE for six load cases

	$X^\pm$	$Y^\pm$	$Z^\pm$
Tension X	$u_1 = \pm \varepsilon_{11}^0 \frac{l_1}{2}$ $t_2 = t_3 = 0$	$u_2 = 0$ $t_1 = t_3 = 0$	$u_3 = 0$ $t_1 = t_2 = 0$
Tension Y	$u_1 = 0$ $t_2 = t_3 = 0$	$u_2 = \pm \varepsilon_{22}^0 \frac{l_2}{2}$ $t_1 = t_3 = 0$	$u_3 = 0$ $t_1 = t_2 = 0$
Tension Z	$u_1 = 0$ $t_2 = t_3 = 0$	$u_2 = 0$ $t_1 = t_3 = 0$	$u_3 = \pm \varepsilon_{33}^0 \frac{l_3}{2}$ $t_1 = t_2 = 0$
Shear XY	$u_2 = \pm \varepsilon_{21}^0 \frac{l_1}{2}$ $u_3 = t_1 = 0$	$u_1 = \pm \varepsilon_{12}^0 \frac{l_2}{2}$ $u_3 = t_1 = 0$	$u_3 = 0$ $t_1 = t_2 = 0$
Shear YZ	$u_1 = 0$ $t_2 = t_3 = 0$	$u_3 = \pm \varepsilon_{32}^0 \frac{l_2}{2}$ $u_1 = t_2 = 0$	$u_2 = \pm \varepsilon_{23}^0 \frac{l_3}{2}$ $u_1 = t_3 = 0$
Shear ZX	$u_3 = \pm \varepsilon_{31}^0 \frac{l_1}{2}$ $u_2 = t_1 = 0$	$u_2 = 0$ $t_1 = t_3 = 0$	$u_1 = \pm \varepsilon_{13}^0 \frac{l_3}{2}$ $u_2 = t_3 = 0$

**Table 4.** Stiffness coefficients (in GPa) for 5 SVEs and experimental results

	$C_{11}$	$C_{22}$	$C_{33}$	$C_{44}$	$C_{55}$	$C_{66}$	Porosity (%)
SVE 14	19.58	22.87	27.96	6.79	8.63	7.64	74.8
SVE 17	23.71	26.59	29.47	8.01	9.23	8.84	74.9
SVE 30	22.94	28.09	31.53	8.28	9.84	9.06	74.4
SVE 44	25.15	26.11	28.65	8.83	8.92	9.06	74.5
SVE 58	26.37	29.33	29.62	9.25	9.99	9.29	74.1
SVE Average	23.55	26.60	29.45	8.23	9.32	8.78	74.5
Average	26.54			8.78			
Experiments [3]	30.2	29.8	28.5	7.2	6.7	N. A	74.5
Average	29.5			6.95			

## 7 Discussion

The article demonstrates the use of statistical functions in characterizing the microstructure of alumina foam which acts as a preform for manufacturing interpenetrating phase composites. A detailed explanation of image processing steps used to convert the grey-scale CT scans into 3D binary image of foam sample has been given. The image is segmented to isolate the interconnected pores so that each pore could be studied. It is observed in Fig.2a that all the pores have more than 0.75 sphericity which indicates that the pores resemble closely to spheres. This is an indication of isotropy of the microstructure. Figs. 2b-2c show that the pores do not have any preferential orientation. Fig.2b has two bright yellow spots close to Z axis and Fig.2c has one bright yellow spot close to X axis. However, their volume fractions are very less and hence will not impact the effective properties of the sample in any way.

Fig 3 shows statistical functions of the entire foam sample. Only for statistically homogeneous microstructure without long-rang order,  $S_2(r)$  follows limits:

$$S_2(r = 0) = v, \quad (13)$$

$$\lim_{r \rightarrow \infty} S_2(r) = v^2. \quad (14)$$

Here,  $v$  is the volume fraction of pores. In our case,  $\lim_{r \rightarrow 0} S_2(r) = 0.72$  and  $\lim_{r \rightarrow \infty} S_2(r) = 0.55$ . Hence, the limits are satisfied. This proves that the sample is highly homogeneous. Note that the value of  $S_2(r)$  at  $r = 0$  does not exactly match value of volume fraction (refer Table.1) because of the limitations in image resolutions. Improved resolutions can bring this value closer to the volume fraction.  $S_2(r)$  also becomes asymptotic above 35 voxels distance. It means that above this value, there is no observable correlation in the pore voxels. The lineal path function (Fig.3c) becomes asymptotic at around 100 voxels distance. This means that the interconnectedness along lineal path is observable only till the distance of 100 voxels. The cumulative pore size

distribution function (Fig.3e) shows that the maximum radius of the spherical pore that can be fitted into the pore space is around 25 voxels. Fig. 4 shows that  $S_2(r)$  and  $L(r)$  have almost same curves when calculated along three orthogonal directions. This proves that the sample is isotropic as well.

Figs. 6a-6d show the effect of number of realizations on ensemble average values of statistical functions. It can be seen that as the VE size decreases, the number of realizations in the ensemble required to match the value of the entire sample increases. Hence, while choosing the appropriate size of VE, a trade-off is required between VE size and number of realizations. In case of VE size 5, the ensemble average lies very close to that of the entire sample irrespective of number of realizations considered in averaging. This is because the difference between this size and that of the entire sample is very less. The fluctuations in the curves are probably because all the 8 VEs in this ensemble share a lot of common region. Hence not enough independent realizations are available to get converging results. Figs.7a-7d show that if enough number of realizations are considered, the ensemble average of statistical functions matches the value of that of the entire sample. As explained before, the choice of VE size for FE calculations depend upon the available computational resources. The VE size 3 chosen in this study satisfies the requirements to be SVE and also fits the computational resources available.

In order to reduce the computational expense of doing FE calculations on 64 realizations of VE size 3, a ranking method is developed. Here, the difference between the statistical functions of each realization and that of the entire sample is calculated and the SVEs are ranked in ascending order of these values (Figs. 8a-8d). Each statistical function has a different ranking order. Hence, it is decided to find those SVEs that lie within 5 % value for all statistical functions. Since these SVEs have microstructure that resemble the most to that of the entire sample, it is decided to perform ensemble averaging on only these 5 SVEs as against 64 SVEs that would have otherwise required. This has significantly reduced the required computational expenses.

Results of FE calculations are given in Table.4 which shows the diagonal coefficients of effective stiffness tensor for all 5 SVEs. Averaging across all SVEs gives us the SVE averages. We can conclude from these values that the foam sample can be considered as isotropic. The experimental results [3] along the three directions also support this statement. Note that the average porosity of the SVEs matches that of the experimental results. Considering isotropy, average of SVE averages  $C_{11}$ ,  $C_{22}$  and  $C_{33}$  gives value of 26.5 GPa. Similarly, averaging of experimental results [3] of  $C_{11}$ ,  $C_{22}$  and  $C_{33}$  gives value of 29.5 GPa. The simulated value is within 10% deviation of the experimental value. Repeating this for SVE averages  $C_{44}$ ,  $C_{55}$  and  $C_{66}$  gives value of 8.78 GPa and for experimental results [3], value of 6.95 GPa. The simulated value is within 20% deviation of the experimental value. These values prove that the adopted procedure of selecting SVE size, ranking method and the numerical calculations predict the effective elastic properties of ceramic foam with a very high degree of accuracy. Further reduction in this deviation can be achieved by using better resolution of CT scan images.

## 8 Conclusions

This article describes the use of statistical functions to characterize the microstructure of ceramic foam and to use these functions to select appropriate size, number and location of SVEs that are further used for determining effective elastic properties of the foam sample. This has led to reduction in size and number of realizations required to determine effective properties of the material which otherwise would have required significant computational resources. The article also demonstrates the suitability of using PMUBCs to determine the effective material properties of non-periodic microstructures with phases having extreme contrast in material properties (infinity in this case).

The larger goal of this research is to establish structure-property-performance relationship links for the ceramic foam material. The statistical functions act as a tool to quantify the microstructure. An important part of this research is to establish a correlation between the statistical functions and the mechanical properties of this material. This will act as a guide in an inverse problem of identifying appropriate microstructure for target material properties. As per authors knowledge, such method does not exist for a microstructure that is unique to foams. In this paper as a first step, an attempt has been made to determine effective elastic properties of foam by using the statistical functions to reduce the ensemble size.

In existing literature, only the effect of volume fraction on the elastic properties has been studied so far. The next step in this research is to artificially reconstruct the microstructure using target correlation functions and then change each statistical function to study its effect on material elastic properties. This way each function can be controlled precisely. This will be followed by sensitivity analysis of each statistical function w.r.t effective material properties. Correlations between the statistical functions if any will be studied as well. Currently this cannot be done as the microstructure that has been used was derived from X-ray computed tomography and hence there was no control over the statistical functions. Once these steps are done, the research will shift its focus on predicting appropriate microstructures for performance enhancement of the material.

## Acknowledgements

The financial support of the Darmstadt University of Applied Sciences is gratefully acknowledged. The publication partially uses raw and experimental data provided by the project WE 4273/17-1 funded by the DFG. Their financial support is gratefully acknowledged. In this regard, special thanks go to Joél Schukraft for the experimental support. The authors thank Morgan Advanced Materials Haldenwanger GmbH for the friendly supply of complimentary ceramic foam material.

## References

1. Chawla, K. K. *Composite materials*, Springer, New York (2012). DOI:10.1007/978-0-387-74365-3
2. Wegner, L. D., & Gibson, L. J. The mechanical behaviour of interpenetrating phase composites—I: modelling. *International Journal of Mechanical Sciences*, 42(5), 925-942 (2000). DOI:10.1016/S0020-7403(99)00025-9
3. Horny, D. et al. Numerical and experimental characterization of elastic properties of a novel, highly homogeneous interpenetrating metal ceramic composite. *Advanced Engineering Materials*, 1901556 (2020). DOI:10.1002/adem.201901556
4. Torquato, S. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*, Springer, New York (2002). DOI:10.1007/978-1-4757-6355-3
5. Torquato, S. Optimal design of heterogeneous materials. *Annual review of materials research*, 40, 101-129 (2010). DOI:10.1146/annurev-matsci-070909-104517
6. Pahr, D. H., & Zysset, P. K. Influence of boundary conditions on computed apparent elastic properties of cancellous bone. *Biomechanics and modeling in mechanobiology*, 7(6), 463-476 (2008). DOI:10.1007/s10237-007-0109-7
7. Meyer, F. Topographic distance and watershed lines. *Signal processing*, 38(1), 113-125 (1994). DOI:10.1016/0165-1684(94)90060-4
8. Wang, D., & Fan, L. S. Particle characterization and behavior relevant to fluidized bed combustion and gasification systems. In *Fluidized bed technologies for near-zero emission combustion and gasification* (pp. 42-76). Woodhead Publishing (2013). DOI: 10.1533/9780857098801.1.42
9. Drach, B., Tsukrov, I., Gross, T., Dietrich, S., Weidenmann, K., Piat, R., & Böhlke, T. Numerical modeling of carbon/carbon composites with nanotextured matrix and 3D pores of irregular shapes. *International Journal of Solids and Structures*, 48(18), 2447-2457 (2011). DOI:10.1016/j.ijsolstr.2011.04.021
10. Matlab R2019b. [https://de.mathworks.com/products/new\\_products/release2019b.html](https://de.mathworks.com/products/new_products/release2019b.html)
11. Jiao, Y., Stillinger, F. H., & Torquato, S. A superior descriptor of random textures and its predictive capacity. *Proceedings of the National Academy of Sciences*, 106(42), 17634-17639 (2009). DOI:10.1073/pnas.0905919106
12. Hill, R. Elastic properties of reinforced solids: some theoretical principles. *Journal of the Mechanics and Physics of Solids*, 11(5), 357-372 (1963). DOI:10.1016/0022-5096(63)90036-X
13. Hazanov, S., & Amieur, M. On overall properties of elastic heterogeneous bodies smaller than the representative volume. *International Journal of Engineering Science*, 33(9), 1289-1301 (1995). DOI:10.1016/0020-7225(94)00129-8
14. Ostoja-Starzewski, M. Material spatial randomness: From statistical to representative volume element. *Probabilistic engineering mechanics*, 21(2), 112-132 (2006). DOI:10.1016/j.probengmech.2005.07.007
15. Hazanov, S., & Huet, C. Order relationships for boundary conditions effect in heterogeneous bodies smaller than the representative volume. *Journal of the Mechanics and Physics of Solids*, 42(12), 1995-2011 (1994). DOI:10.1016/0022-5096(94)90022-1
16. Abaqus, Simulia. <https://www.3ds.com/products-services/simulia/products/abaqus/>

# Automatic Error Detection For Image To Text Tools

Suzanne Considine<sup>1</sup>[0000-0001-8053-3250], Haithem Affi<sup>1</sup>, and Andy Way<sup>2</sup>

<sup>1</sup> Cork Institute of Technology, Rossa Avenue, Bishopstown, Cork, Ireland  
suzanne.considine@mycit.ie

<sup>2</sup> Adapt Centre, School of Computing, Dublin City University, Dublin, Ireland

**Abstract.** The focus of the work discussed in this paper is to present a breakdown of the most prevalent errors that are present in automatically generated text and present the results of different methods of automatically identifying and correcting these errors. Inspired by analysing the errors present in text generated by image to text generating tools this work is also beneficial to other fields which use automatic text generating, for example bots in help centres and automatic tweeting tools. While bots and tweeting tools will not reproduce the exact errors as the image to text generation tools they will both benefit from being able to identify the errors as this will make them appear more human like in their interaction with clients and tweeters. This work centres on a dataset of captions which were automatically generated by NeuralTalk2 from data provided by NIST. This dataset was manually classified and then compared to other datasets of captions generated by other image to text tools to ensure that the errors identified as being significant were in fact significant across other datasets of automatically generated text

**Keywords:** Errors · Automatically generated text · Image to text.

## 1 Introduction

Error detection and correction has received a lot of attention over the past sixty years. The work so far has mostly focused on errors as a result of human errors, dyslexics, non-first language speakers, and errors as a result of using Optical Character Recognition (OCR) tools (e.g. of an error that can be caused by an OCR tool confusing lr with h).

Methods have been developed for correcting spelling errors or correcting confusion words (peace of cake, their/there/theyre, etc.). The methods developed for these errors include the edit distance [1] for spelling errors (where once a word is identified as a misspelling, depending on the parameters of the tool used, a set of valid words one letter away from the misspelling, and another set two letters away, and so on, are presented as possible corrections). The word from the sets is then usually chosen by using a corpus broken into n-grams of 2 or 3 to see what is the most likely word from the set to fit in. Other methods include rule based systems [2] which tended to be inflexible and needed constant upgrading

to keep up with evolving language and the n-gram method [3] which examines all n words, (usually n is 2 or 3 maximum) which appear consecutively in the text being tested and alerts the system if there is an n-gram in the testing text that does not appear in the training corpus.

From manually classifying 2,387 captions generated by image to text captioning tools we found the following four main types of errors to be prevalent: incomplete sentences, grammatically correct nonsense, redundant repetition and general grammar mistakes. Please see Table 1 for examples of these errors and their prevalence in the dataset.

**Table 1.** Error type and prevalence.

Error Type	Total	% of Total Classified
Correct	1775	74
Incomplete sentences (A woman is sitting on a big)	113	5
Grammatically correct nonsense (A man has a dog in his mouth)	194	8
Redundant repetition (The man is drinking tea and drinking tea)	197	8
Grammar	108	5

Apart from grammar errors, the methods discussed above, which work so well for the human and OCR errors, do not work as well or even at all, for the other three error types present in automatically generated text; grammatically correct nonsense, incomplete sentences and redundant repetition (errors where the system has repeated strings). When using the n-gram approach for error detection the parameter for n is unlikely to be set higher than 3 because once you go beyond 3 there are usually so few examples of the exact same 4 word or more string in the training corpus that it is not fit for purpose. So, the sentence, The woman is holding a baby in her mouth would not be detected as incorrect using 3-grams, as on their own, each set of three concurrent words makes sense. An incomplete sentence, The man was running away from a very big would also escape detection as an error with the n-gram method for similar reasons, and the error , a man is walking a dog and walking a dog, would also go undetected using this method

This study will focus on this area of the detection and correction of errors present in automatically generated text, and in particular, those related to text generated by image to-text tools.

## 2 Related Work

There doesn't appear to be any published effort on the detection of incomplete sentences or redundant/error repeated strings within sentences. While identifying whether sentences make sense or not is gaining increasing attention, a method still hasn't been devised that can satisfactorily detect nonsense. Nor

is there much documentation showing the results of how good systems are at detecting nonsense.

ELMo(Embeddings from Language Models) [4] and BERT (Bidirectional Encoder Representations from Transformers) [5] are two tools which compute vectors for words (tokens). Both tools are similar in that they both work on the premise that the words that come after the token being vectorized, are just as relevant as the words that come before, when determining the context of the token. ELMo first takes into account the words before the token, then takes into account the words after the token, and thereafter, concatenates this information to compute the vector for the token. BERT however is capable of taking in the words before and after the token simultaneously and using this information to predict a masked word.

Both ELMo and BERT differ from and outperform previous vectorization applications such as word-to-vec and GloVe, in that they can take the full sentence into account to get the vector representation for a token, thereby obtaining a more representative vector.

ELMo and BERT were not designed specifically to focus on detecting nonsense, but it has been reported [6] that ELMo is 69.1% accurate and BERT 70.4% accurate at detecting nonsensical statements. These figures were arrived at by building a dataset of 2021 rows, each row comprising of the 5 separate parts (namely s1, s2, r1 ,r2 ,r3; where firstly, s1 and s2 were statements of which one of s1 or s2 made sense while the other was nonsense, and secondly, one of r1, r2, r3 was the reason why either s1 or s2 was nonsense while the other two were illogical statements), and then feeding this labelled database into ELMo and BERT to calculate their relevant accuracies at detecting nonsense.

While their experiments did not focus on detecting nonsense, an earlier paper [7] developed a technique they call verbaphysics that could be used to build a classifier to detect whether statements are plausible or not. Verbaphysics involved providing an application with knowledge of what would be a plausible concept and what would not, by gathering and applying specific information regarding the objects being assessed, namely, size, weight, strength (e.g. diamond is stronger than glass), rigidity (e.g. glass is more rigid than wire) and speed. During their trials, they selected 50 Levin verbs (Levin, 1993) which corresponded to 1100 unique verbs; then picked out the top 100 most used verbs from these. They then obtained approximately 9 frames per verb. Example of a frame: The man threw the ball and from that they gleaned information such as, the man is stronger than the ball, that the man is bigger than the ball, the ball is faster than the man. The end product is a database of objects and how these objects should relate to each other, but only as regards these five qualities. And they used this to determine whether a concept is plausible or not. For example, a man cannot throw a house as it is bigger, a man cannot chase a tree as a tree has no speed.

(Wang et al., 2018) built on the work of Forbes et al., 2017, showing again that world knowledge is necessary for semantic plausibility. They build a dataset of an equal number of plausible and implausible statements while Forbes et al. graded

their 5 variables (viz. size, weight, strength, rigidness and speed) on as a value of either -1, 0 and 1, which meant the statement, the man hugged the ant, would be as plausible as the statement, the man hugged the cat, according to Forbes et al. as the information they have stored is that the man is bigger than both the ant and cat, when in fact, it is not as plausible. To address this problem, Wang et al. developed a more finely grained grading system for the variables thereby allowing distinctions to be made between scenarios that wouldnt have been under the Forbes et al. model. So that under the improvements made by Wang et al., a man hugging a cat would be deemed possible, but a man hugging an ant, not so as the ant is so much smaller than the man. Wang et al. created a dataset of 3062 statements, which for testing to ascertain whether world knowledge improves the accuracy of semantic plausibility classification. They ran (Van de Cruys, 2014) neural net classifier and found that it was only 64% successful at distinguishing between the plausible and implausible statements, but when world knowledge was injected/projected onto the neural net classifier, the accuracy rose to 76%.

### 3 Dataset

The dataset which we manually classified, consists of captions generated by the image captioning tool, NeuralTalk2<sup>1</sup>, using data (viz. videos) provided by National Institute of Standards and Technology (NIST). NeuralTalk2 generates captions by passing the image through a Convolutional Neural Network (CNN) which extracts concepts from the image. Depending on the data used to train the model, the extracted information from the images (referred to as the concepts) can be a noun (e.g. vase, horse, car, etc.) or an observation (e.g. a man on a horse, a frisbee in the air, etc.). These concepts are passed to a Recurrent Neural Network (RNN) which uses the concepts along with a corpus (database of manually generated captions from the same domain) to generate the captions. The RNN uses the concepts as inputs and calculates the probability of the next word in the caption based on the text in the corpus. This method of generating a sentence is unique to this domain, therefore some of the errors in the sentences are unique to these text generating applications.

Our dataset of captions generated by NeuralTalk2 consisted of 8,000 unique captions of which 2,387 were randomly and manually analysed. The errors detected are shown in Table 1 above.

We then examined a number of smaller datasets of captions generated by image to text generating tools, which were provided to us by NIST. The same errors as above were present in each dataset we examined but to varying degrees of severity.

### 4 Experiments

Our experiments revolve around the three error types from our manually classified dataset which cannot be solved by traditional n-grams method. While each of the three error types needs a different technique for detection, our aim is to

automate the necessary pre-processing of the data so that we can then feed the features necessary for detection of each error into a learning model.

#### 4.1 Incomplete Sentence Errors

The maximum length of our captions is 14 words long. In order to use each word as a feature, we need to have an equal number of words for each caption. We do not want to lose information by shortening the captions, so we padded each caption in the dataset, where necessary, at the start with unk to make each caption 14 words long. We then convert each word to its Part Of Speech (POS) tag using the NLTK library. We then hash all the POS giving each POS a unique number. See Table 2 for an example of the various stages a caption of insufficient length goes through to be prepared for the learning model.

The insertion of the word unk at the start of each caption which is less than 14 words long, does not influence the learning model as we drop these unk features and only focus on the pattern of the POS tags at the end of the sentences to detect the incomplete sentences.

We consistently get a success rate of over 80% for identifying the incomplete sentences.

**Table 2.** Preprocessing Example for Learning Model.

Step of Process	State of Data
Caption	a bike is parked next to a wall
Caption Expanded	unk unk unk unk unk unk a bike is parked next to a wall
POS	JJ JJ JJ JJ JJ JJ DT NN VBZ VBG IN DT NN
Hash (POS)	1 1 1 1 1 1 2 3 4 5 6 2 3

#### 4.2 Redundant Repetition (\* and \*)

Our approach to identifying this error type is based on the observation that any caption which has repeated nouns or adjectives, is usually symptomatic of this error type. While there could be occasional exceptions to this rule e.g. a woman is handing a phone to a woman, we have not encountered a caption like this yet. If a caption repeats a noun (e.g. a man is wearing a tie and a tie, a group of people sitting around a table with a table, etc.), or an adjective (e.g. a girl is holding a green and green box, etc.), we will tag this caption as having the \* and \* error type. The learning model needs information specific to these captions, but we will not interfere with the POS tags already prepared for the incomplete sentences.

Instead we extract the nouns and adjectives from each caption and write them to an excel spreadsheet. A Python script goes through each row checking if there are duplicate nouns or duplicate adjectives and tags each caption accordingly.

This tag is an additional feature which we add to the features already prepared for incomplete sentences. We add this feature to the start of the rows being fed into the language model, and now we have 15 features. Note again we have not interfered with the information needed to identify POS tags.

The learning model now detects these duplication errors over 80% of the time.

### 4.3 Nonsense Errors

Nonsense is the most challenging of the three errors to detect as there are so many reasons why a caption can be considered nonsense. The nonsense can revolve around a verb, a subject doing something impossible (e.g. A man is flying through the air, A dog is holding a glass, etc.) or adjective (e.g. a fluffy woman, etc.). An item can be in an implausible location or position (e.g. A bench in a car, A cat on a cats head, etc.). There are many different variations of this theme.

In this paper we will focus on what is plausible regarding verbs. We begin by extracting the verbs from our manually classified dataset, yielding a total number of 22 verbs. We then make a decision on what subject types can precede each verb in the dataset, and we end up with 5 sets of subjects, see Table 3.

**Table 3.** Subject Types and Corresponding Verbs They Can Precede.

Subjects	Verbs
Human	brushing / cutting / smiling / talking / posing / riding
Human/animal/primate	jumping / playing / laying / running / wearing
Human/animal/primate/bird	watching / eating / holding / looking / sitting / sleeping / standing / walking
Human/primate	throwing / spitting
Human	driving
Bird/other objects	flying”

We utilise a variation of the n-gram method to identify nonsense statements which contain these 22 verbs. We gather captions from Wiki Commons, Google Image, and GCC containing these 22 verbs. We then extract: (a) all the nouns preceding the verb in the captions, (b) the verb, and (c) the first noun following the verb. Refer to Table 4 below for examples of captions and corresponding 3-grams. We build a corpus of 3-grams relating to the 22 verbs we extracted from the dataset.

Taking a similar approach to our dataset, we iterate through the captions extracting: (a) The nouns before the verb, (b) The verb, and (c) The first noun after the verb.

We then check the nouns before each verb, and check that one of these nouns belongs to a member of the subject type that we have deemed capable of executing this verb. If this is the case, we discard all the other nouns preceding

the verb so that we are left with first 2 words of our 3 gram namely the subject and the verb.

If this is not the case, we check this caption to see if we need to include a new noun in our subject type for this verb, and if not we label the caption as nonsense as we have a subject preceding a verb where it should not, (e.g. a dog is singing, etc.).

Next, we extract the first noun after the verb, and now we have our 3-grams from our dataset, taking the following form: the subject before the verb+ the verb+ + the first noun after the verb.

We check if these 3-grams exist in our corpus of 3-grams, and if they dont, we label the sentence as nonsense. If it does, we label the sentence as plausible.

Our method correctly identifies the nonsense and plausible sentences with 63% accuracy. Analysing the results further, we have 125 nonsense captions, 97 of which are correctly detected giving us a 78% success rate. We have 670 plausible captions of which 410 are correctly detected, giving us a 61% success rate.

The nonsense captions which go undetected had the following errors which were not picked up by our method, The woman is holding a puppy in her mouth, or the girl is holding a pink cat. Our method produced the following 3-grams woman holding puppy and girl holding cat, these n-grams did exist in our corpus so the tool labelled the captions as plausible, and as we are only working on verbs we do not expect to pick up on these errors in this work.

## 5 Conclusions and Future Work

The main contribution of this paper is to present an ongoing Automatic Error Detection for image to text tools projects that aims at building the first corpus, at the best of our knowledge, to be used for these types of tasks.

We believe that injecting knowledge into applications is necessary for improving common sense in artificial intelligence. In the short term/immediate future we intend to continue our work regarding extracting subject-verb-object relationships for detecting nonsense and in particular, we want to place a special focus on extending the corpus, and developing further techniques that will enable the maintenance of a compact corpus capable of providing extensive knowledge relating to verbs to our application.

## References

1. Levenshtein, L.: Binary codes capable of correcting deletions, insertions and reversals. 1966. Soviet Physics Doklady, 10, 707710.
2. Heidhorn, G., Jensen, K., Miller, L., Bird, R., Chodrow, M.: The EPISTLE text-critiquing system. 1982. IBM Systems Journal 10.1147/sj.213.0305.
3. Mayes, E., Damerau, F.: Context Based Spelling Correction.: 1991: Information Processing and Management 27(5): 517-522.
4. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations.: 2018, In Proc. of NAACL

5. Devlin, J., Chang, M., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding.: 2018, arXiv preprint arXiv:1810.04805
6. Wang, C., Liang, S., Zhang, Y., Li, X., Gao, T.: Does it Make Sense? And Why? A Pilot Study for Sense Makeing and Explanation.: 2019, arXiv preprint arXiv:1906.00363.
7. Forbes, M., Choi, Y.,: Verb physics: Relative physical knowledge of actions and objects.: 2017.: ACL.

## Chapter 3

---

# Visual Computing and Data Analytics



# Determining Elastic Properties of Particle Reinforced Polymer Composites by Numerical Modeling of their Microstructures

Pascal Alexander Happ<sup>1</sup>, Timo Bensing<sup>2</sup>, Simon Terwolbeck<sup>2,3</sup>, Katharina Malek<sup>2</sup>,  
Martin Moneke<sup>2</sup>, Romana Piat<sup>1</sup>

<sup>1</sup> Faculty of Mathematics and Natural Sciences, Darmstadt University of Applied Sciences,  
Germany

Pascal-Alexander.Happ@h-da.de

<sup>2</sup> Institute of Plastics Engineering, Darmstadt University of Applied Sciences, Germany  
Timo.Bensing@h-da.de

<sup>3</sup> Granula Polymer GmbH, Germany

**Abstract.** A numerical model of particle reinforced composites was created to calculate the elastic material properties. Therefore, a thermoplastic polymer filled with different types of precipitated calcium carbonate was chosen as the material. Size and shape of the particles as well as their distribution and orientation within the polymer matrix were examined with a scanning electron microscope. Representative Volume Elements (RVEs) based on different particle shapes and distributions on a micro scale were numerically modeled. Six sets of boundary conditions in displacements were applied to the RVEs to simulate three load cases of uniaxial tension (in the directions of the three global coordinate axes) and three shear load cases. These results were used to calculate the components of the stiffness tensor of the composite. This process was carried out for single particle inclusions to study the contribution of the particle to the overall elastic material properties as well as for RVEs with multiple particles, to study the contributions of the particles interactions. Lastly, a more accurate representation of real microstructures was created, taking agglomerates into consideration. Studies about the influence of the agglomerates to the elastic material properties were carried out.

**Keywords:** Numerical Modeling, Particles Interaction, Micromechanics, Scanning Electron Microscopy, Polymer Matrix Composites, Agglomerates.

## 1 Introduction

Particle reinforced polymer composites are applied in a variety of different fields satisfying the individual requirements of the intended applications in various industries. One objective of filler materials deals with changing the mechanical properties of the polymer [1]. Despite the specific filler material, the manner of modification in the mechanical properties depends on several factors like size, shape and amount of the added particles. Further, distribution and dispersion of the particles within the polymer matrix

have an impact on the mechanical behavior of the composite. Thus, there is a high variation potential in the development process like changing the specific filler type or filler amount in order to reach the predefined properties. As a result, different composites can be produced and need to be analyzed. This experimental approach usually involves a high effort at producing the different composites, testing their mechanical properties and evaluating the results.

Another possible way is predicting the mechanical behavior by using numerical studies. Thereby, the mentioned effort can be reduced and the development process can be shortened. Even though every composite is created to meet specific requirements and therefore is not freely interchangeable with other particle reinforced composites, they do share the mechanics at the microscopic level. Thus, a theory has been developed that describes the mechanics at the microscopic level to predict the material behavior at a macroscopic scale. For this, studies of particle reinforced polymer composites at a micro scale are carried out to increase the insight of the mechanical behavior.

In this work a thermoplastic polymer filled with different types of precipitated calcium carbonate (PCC) serves as research subject. The objective is predicting the stiffness of the composites against tensile and shear loads. As a necessary input parameter for the numerical model both the specific size and shape of the particles as well as their distribution in the compounded composite need to be analyzed.

## 2 Investigation of the Filler Particles and Composites

### 2.1 Studied Materials and Experimental Methods

As mentioned above a thermoplastic polymer is used as the material filled with five different types of PCC. Compared to naturally gained grounded calcium carbonate PCC has several advantages. As a synthesized product it provides the possibility of creating different particle sizes and shapes [2]. As a result, the five added materials exhibit divergent sizes and shapes.

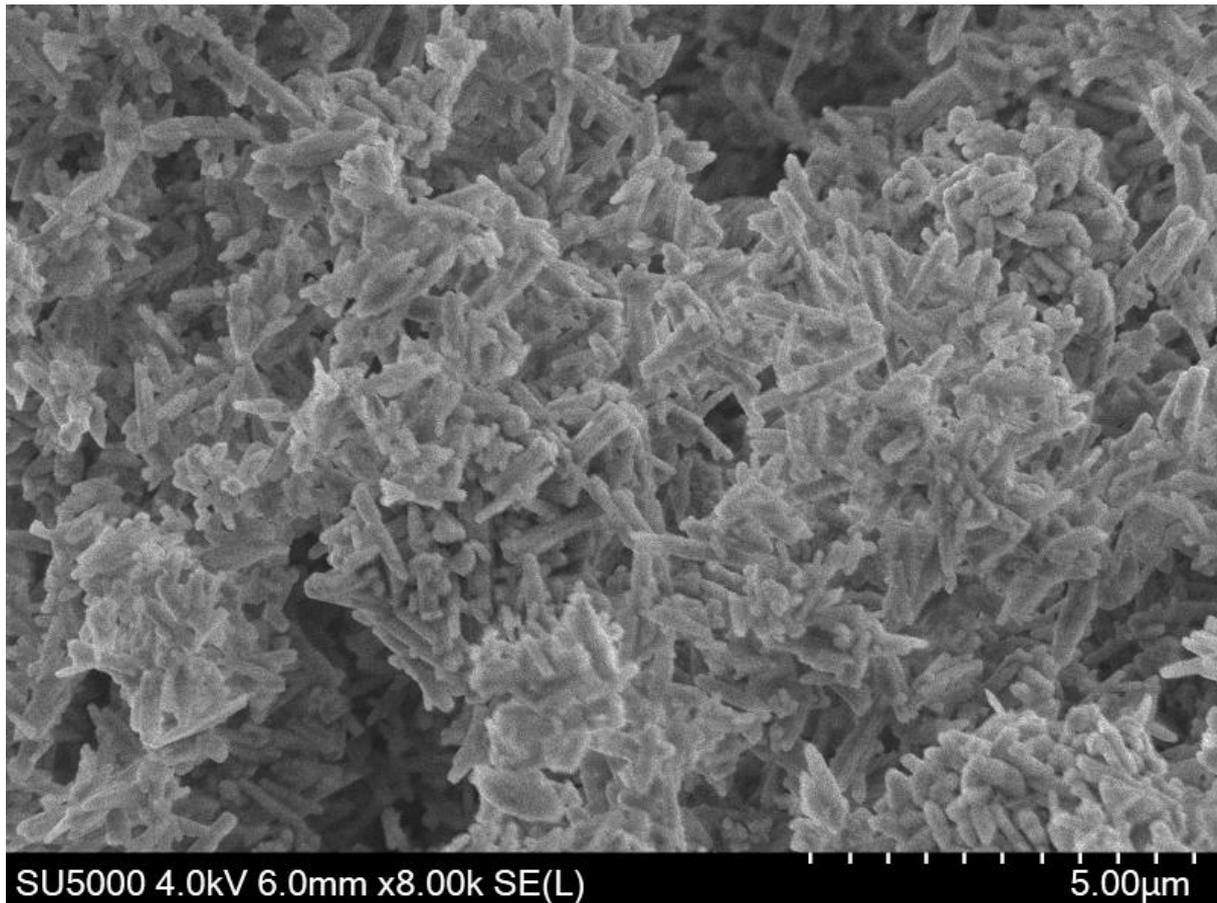
A scanning electron microscope (SEM) was used to gain the necessary information about the added particles, whereby a Hitachi SU5000 served as the device. In a first step, microscopic images were taken of the particles. The particles were placed as a thin layer onto the sample holder and attached with a conductive carbon adhesive. The images were created by detecting the secondary electrons. From the images the size and the shape of the particles were estimated. Furthermore, the aspect ratio as the ratio between the largest and the smallest dimension of the particle was calculated [3].

After investigating size and shape of the particles itself the compounded composites were studied again by using the SEM. The desired information was gained by performing a notch-impact test according to EN ISO 179-1:2010 and investigating the fractured surface. Therefore, multipurpose test specimen as defined in EN ISO 20753:2018 (type A1) were manufactured and provided with a v-notch (depth: 2 mm, angle: 45°, radius notch root: 0.25 mm). The notch was brought into the test specimen in order to ensure that every sample breaks. A Zwick-Roell PSW 5113 served as the testing device and the notch-impact tests were carried out by using a pendulum with a working capacity of 2 J. The fractured surfaces were coated by 4 nm Au layer. The images again were

created by detecting the secondary electrons. On the basis of the microscopic images distribution, dispersion and the orientation of the particles within the polymer matrix were estimated.

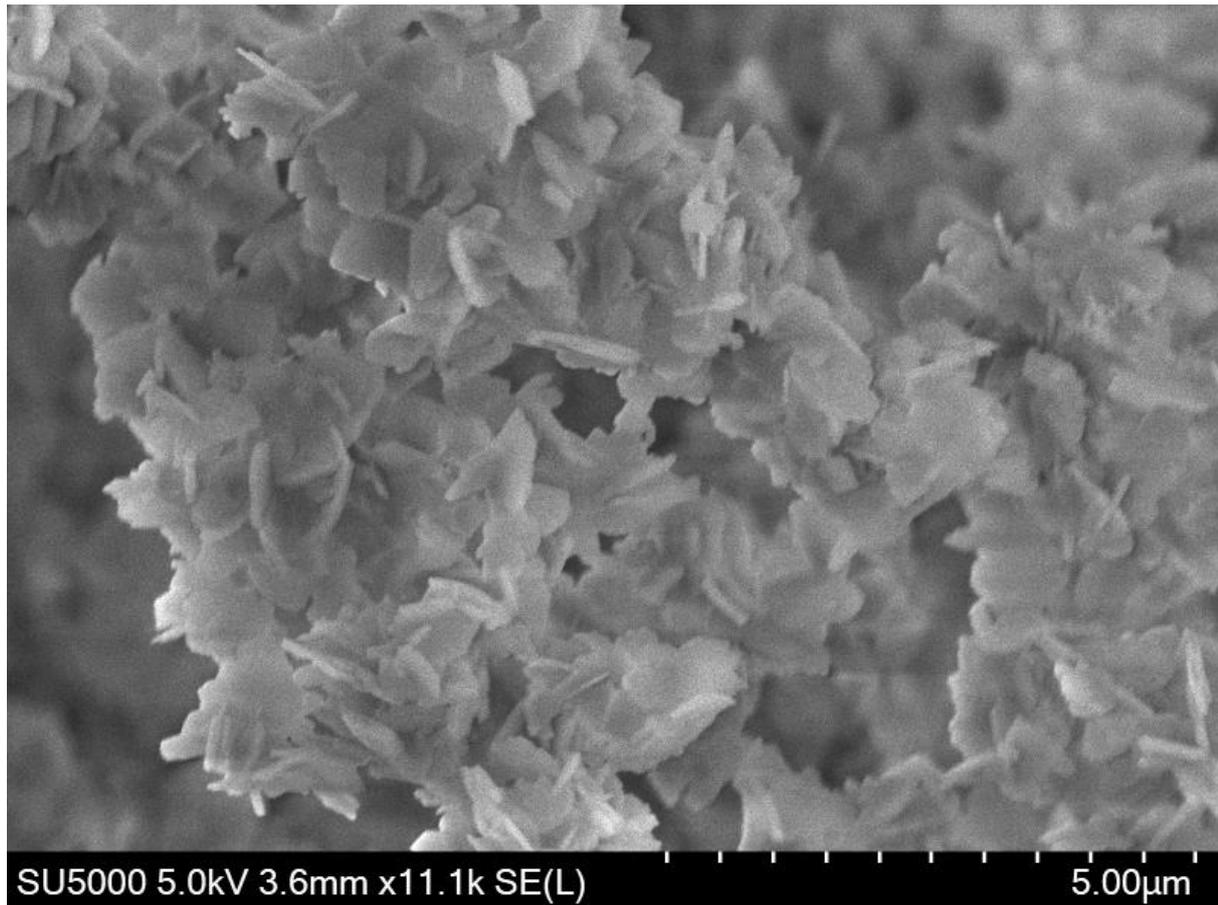
## 2.2 Scanning Electron Microscopy Analysis

At first the microscopic images of the particles are examined. An image of one example taken by the microscope is shown in Fig. 1.



**Fig. 1.** SEM image of needle-shaped PCC particles

The image shows needle-shaped particles stick together in an agglomerate. The particles have an estimated length of around 1-2  $\mu\text{m}$  and a width of around 0.25  $\mu\text{m}$  in the center, yielding an aspect ratio from 4 to 8. All of the other four PCC-types show unique shapes as well. Besides the mentioned needle-shaped particles there are also spherulitic, skalenohedral and platelet-shaped particles. The fifth type is a mix between skalenohedrons and needles. To demonstrate the differences, the platelet-shaped PCC-type is shown in Fig. 2.



**Fig. 2.** SEM image of platelet-shaped PCC particles

The information gained from the investigation of the different PCC particles with the SEM is summed up in Table 1.

**Table 1.** Shapes, sizes and aspect ratios of the five investigated PCC particle types

PCC-type No.	Shape	Size	Aspect Ratio
1	Skalenoedrons	Length: ~ 1 $\mu\text{m}$ Width: ~ 0.3-0.5 $\mu\text{m}$	~ 2-3.33
2	Needles	Length: ~ 1-2 $\mu\text{m}$ Width: ~ 0.25 $\mu\text{m}$	~ 4-8
3	Mix of skalenoedrons and needles	Length: ~ 1-2 $\mu\text{m}$ Width: ~ 0.5 $\mu\text{m}$	~ 2-4
4	Platelets	Length: ~ 1-2 $\mu\text{m}$ Width: ~ 0.01 $\mu\text{m}$	~ 100-200
5	Spherulites	$\text{\O}$ : ~ 0.05-0.1 $\mu\text{m}$	1

Next, the fractured surfaces of the composites are examined aiming to get information about distribution, dispersion and orientation of the particles in the polymer matrix. As a justifiable assumption one would expect the platelet-shaped particles to have a higher

orientation in flow direction of the melt during the injection molding process compared to the other four PCC-types due to their higher aspect ratio.

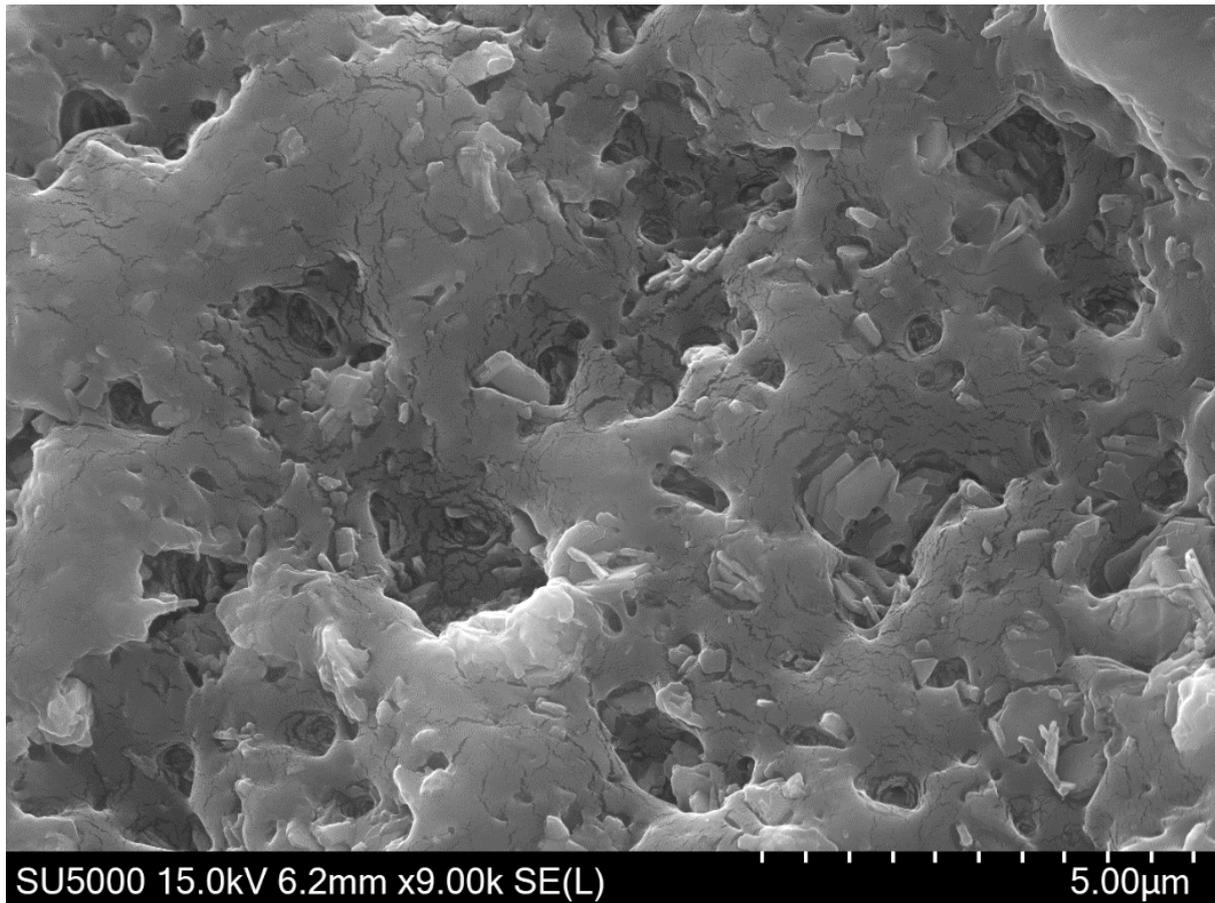
A microscopic image of a composite with the needle-shaped filler material is shown in Fig. 3.



**Fig. 3.** SEM image of the fractured surface of the thermoplastic polymer filled with needle-shaped PCC-particles; the highlighted area marks an agglomerate

In the image one can see a mostly homogenous distribution of the particles within the polymer matrix. However, scattered residual agglomerates can still be located such as highlighted on the middle left side of the image. The compounding process could not break up the agglomerates totally. In conclusion, they must be taken into account in the numerical analysis. Furthermore, the particles show no preferential orientation within the matrix at all, which is traced back to their low aspect ratio.

This description can be transmitted analogous both to the sklenohedral-shaped particles as well as to the mixture of the two types. They also seem to have no preferential orientation in the viewed images and have a low aspect ratio. An exception are the platelet-shaped particles as presented in Fig. 4.



**Fig. 4.** SEM image of the fractured surface of the thermoplastic polymer filled with platelet-shaped PCC-particles

As expected, the particles appear to have a higher orientation whereby the base area tends to align perpendicular to the flow direction (the direction of flow is perpendicular to the image surface). Mainly the tips of the particles protrude from the fractured surface. This confirms the expectations due to the higher aspect ratio of the particles.

Finally, the spherulitic-shaped PCC-type does not have an orientation due to its particular shape.

### 2.3 Limitations of the Investigation

Unfortunately, the information gained from the investigation with the SEM is limited in several aspects. First, the sizes are just estimated by using the scale integrated in the micrograph and not quantitatively measured. In a similar way the conclusions made about the distribution, dispersion as well as the orientation of the particles within the matrix are limited, because they are qualitatively evaluated. Second, the examined fractured surfaces of the composites are a small section and therefore cannot be generalized for the entire specimen. On top of that, the surface was created by deploying a notch-impact test and thus was affected by a large dynamic force. This factor could also have caused an impact onto the particles, e.g. by breaking their linkage to the polymer matrix.

### 3 Numerical Modeling

#### 3.1 Methodology and Assumptions for the Numerical Evaluations of Elastic Properties

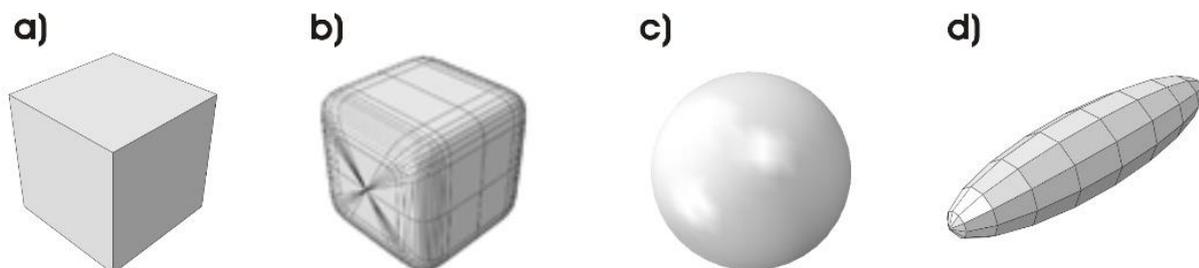
An approximation of the particles by use of ellipsoids allows the analytical evaluation of the elastic properties of the composite, opening up two branches of composite evaluation methods, making it possible to compare results. On the one hand, using the Eshelby solution for ellipsoidal inclusions [4], semi-analytic methods of Mori-Tanaka [5] and Lielens [6], as well as the Dilute inclusions method [7-9] can be used as homogenization methods, to evaluate the elastic properties of the composite. On the other hand, numerical methods such as the FEM analysis can be used to study the effects of various particles onto the composite [10] properties.

In this research we utilized the numerical evaluation. For the numerical calculations Representative Volume Elements were created (RVEs) as discussed by Khisaeva et al. [11] and Gitman et al. [12]. The algorithm to create RVEs consisting of multiple periodic distributed particles is based on the Random Sequential Adsorption (RSA) algorithm proposed by Rintoul et al. [13]. Boundary conditions were placed onto the surfaces and six load cases were evaluated, as described by Drach [14], three of them being of uniaxial tension and three of them being of shear deformation.

An ideal smooth surface of the particles, as well as an ideal bonding between matrix and particles was assumed. Considering the multiple particle evaluations, no overlapping of the particles was allowed. For these calculations only the elastic behavior was considered.

#### 3.2 Numerical Modeling of the Composite with Periodic Distributed Single Particles

In the following studies four particle shapes were taken into consideration: ellipsoidal, spherical, cubic and cubic with smooth edges, as depicted in Fig. 5. For creation of the ellipsoidal, spherical and cube particle shapes for the FE calculations analytic functions were used.

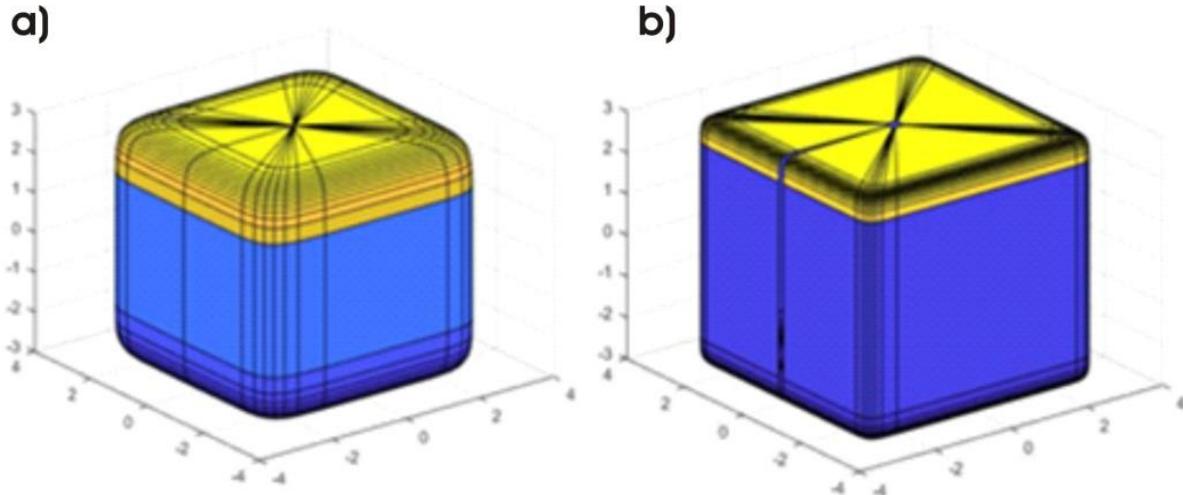


**Fig. 5.** Studied particles: a) Cubic, b) Cubic with smooth edges, c) Spherical, d) Ellipsoidal

Furthermore, the equations for super ellipsoid (Eq. (1)), as described by Jaklič et al. [15], were applied to create cubes with smooth edges:

$$\left(\frac{|x|}{a}\right)^n + \left(\frac{|y|}{b}\right)^m + \left(\frac{|z|}{c}\right)^k = 1, m, n, k \in \mathbb{R}_+ \quad (1)$$

The parameters influencing the radius of the corners of the cubes resulting in particle forms which are depicted in Fig. 6.



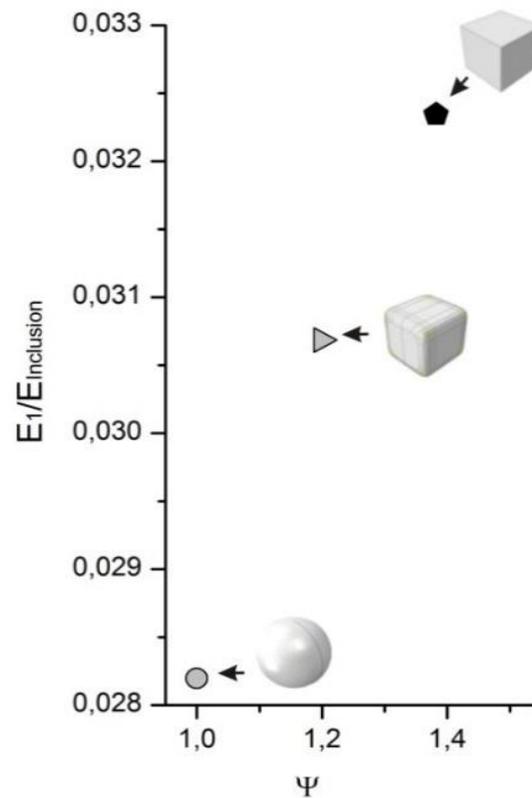
**Fig. 6.** Cubes with different smooth edges created by the use of super ellipsoid a) for  $n=m=k=8$ ,  $a=b=c=3$  and b) for  $n=m=k=20$  and  $a=b=c=3$

The particles were then embedded into a matrix, creating the RVE. Six load cases were deployed to the surface of the RVE and the stresses were calculated using numerical methods. Next, the stress volume averages were calculated. Lastly, the Young's moduli were evaluated as proposed by Drach [14]. The specific numeric calculations were done in the ABAQUS software [16].

The Young's modulus of the composite has been normalized by the Young's modulus of the particle and displayed over  $\psi$ , the surface to volume particle ratio, as proposed by Wadell [17] with slight modifications. The modified equation (Eq. (2)) is written below. Here  $S_P$  and  $V_P$  are the surface and volume of the particle respectively.

$$\Psi = \frac{S_P}{\sqrt[3]{\pi(6V_P)^2}} \quad (2)$$

The results are depicted in Fig. 7.



**Fig. 7.** Normalized young's modulus for specific particle form over surface to volume ratio

The surfaces of the particles play a measurable role for the overall elastic Young's modulus of the particle reinforced composite. The real form of the approximation cannot simply be approximated by a sphere, only taking the enclosed volume (of the particle) into consideration.

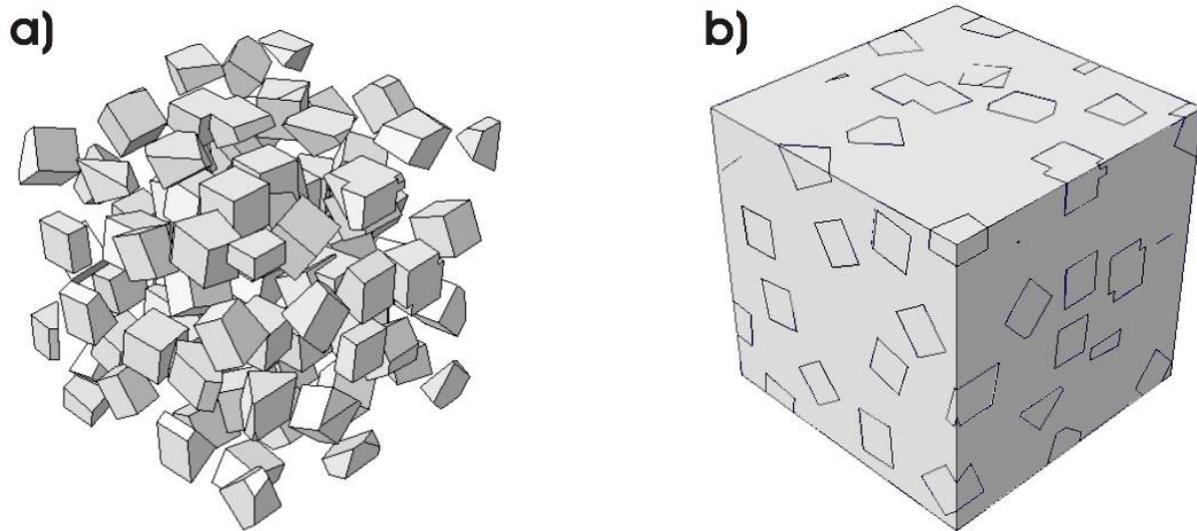
### 3.3 Numerical Modeling for Multiple Particles

As it was shown in the last chapter the form of the particle as well as its surface play a major role in the resulting overall elastic properties.

Up to this point only single inclusion set ups were considered under the premise that the particles are far apart from each other. So that there are no interactions, or the interactions are small to the point of being negligible. This is acceptable if the volume fraction of the particles in the composite is small. Most of the time this is not the case, as can be seen in Fig. 3. Here particles are usually close together, so that the interactions should be taking into account. Furthermore, simulations must include the formation of agglomerates, which are represented by use of particle clusters.

The positioning of the particles in a multiple inclusion set up and as clusters is realized by the use of algorithms. For this process the particles created for the single particle set-up were used. The particles are placed by the algorithm inside a cube, while allowing a certain protruding of the particles. The algorithm places particles inside the cube until the preset volume fraction is reached, as proposed by Seguardo et al. [18]. Studies were provided to estimate the appropriate amount of particles sufficient for obtaining

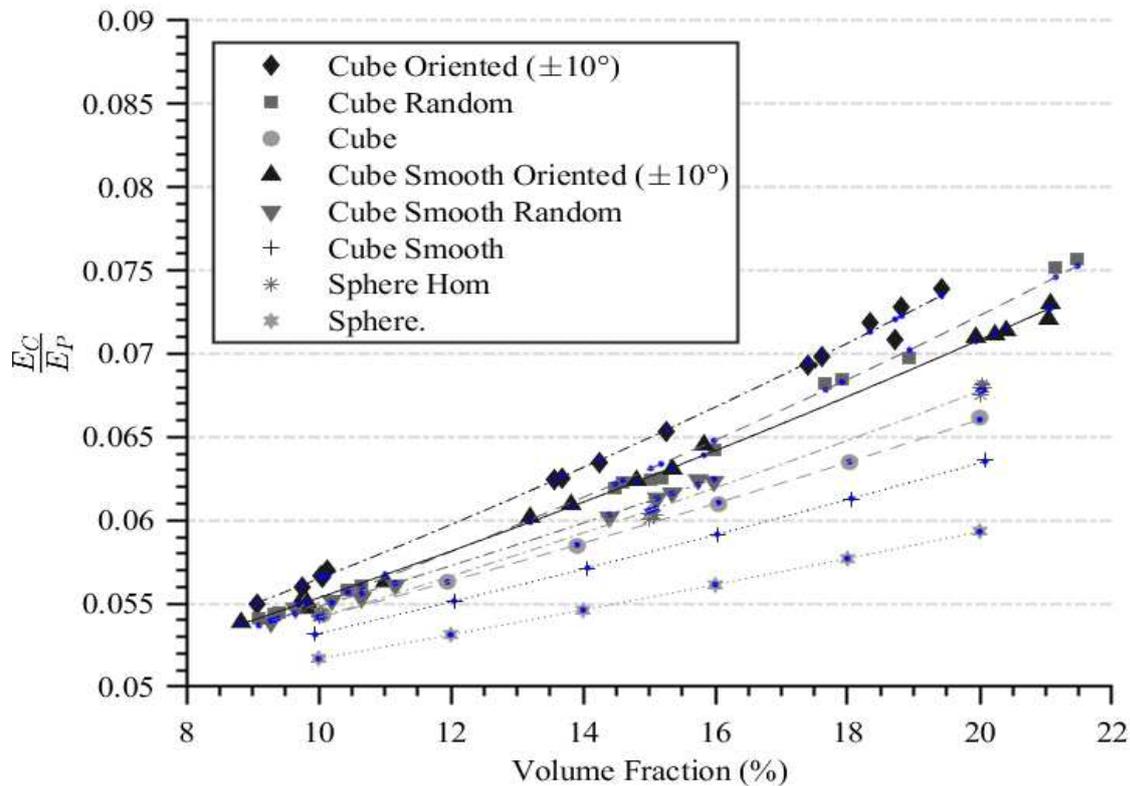
the RVE. For a random homogenous distribution of a cube like particle the obtained RVE is depicted in Fig. 8.



**Fig. 8.** RVE for random homogenous distribution of a) cubic particles, b) The same RVE with particles embedded in the matrix

Calculations were carried out for different particle forms and volume fractions considering the case of the random homogenous distribution. A closer look was taken onto the influence of the orientation of the particle on resulting elastic properties. Using self-written placement algorithms (in MATLAB [19]) the particles were generated, placed randomly and oriented within an axis direction, if needed. The following particles were studied: cubic, cubic with smooth edges, spherical and ellipsoidal (Fig. 5).

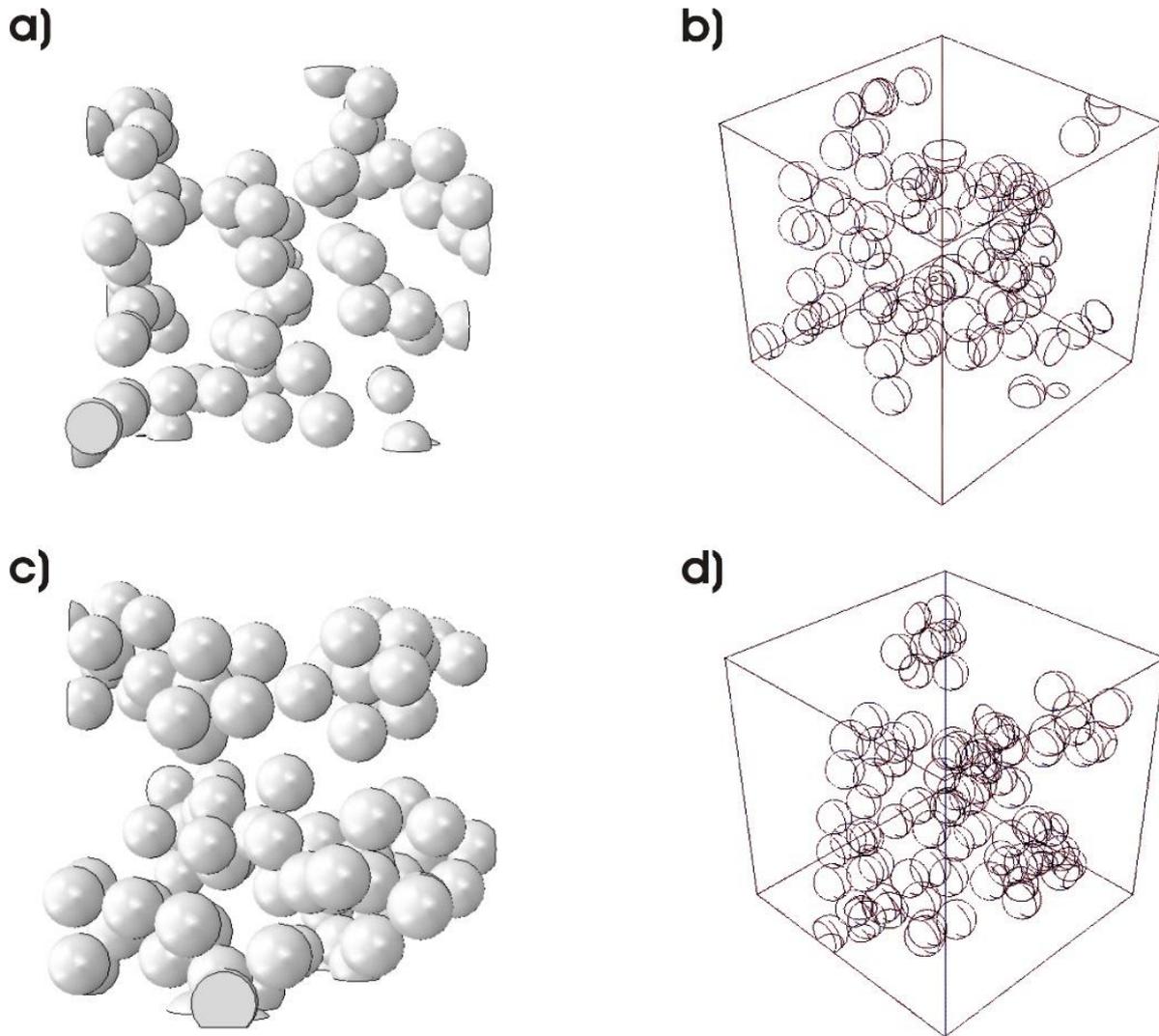
The elastic properties of the composite for each particle shape were evaluated using the ABAQUS software. The values of the normalized Young's modulus of the composite considering different particle shapes can be seen in Fig. 9, with  $E_C$  being the Young's modulus of the composite and  $E_P$  being the Young's modulus of the particle. The multiple inclusion set ups were also compared against their single inclusion counterpart. It can be seen, that the multiple particles set up achieves higher Young's moduli than the periodic distributed single particles in general. This could be attributed to the interactions of the particles having a beneficial contribution.



**Fig. 9.** Normalized effective Young’s modulus of the composite for different particle shapes and distributions for X1 direction.

In general, particle shape and particle orientation play a major role on the overall elastic material properties of the composite. The approximation of the real particle in a multi-particle set-up has an influence on the quality of the predicted elastic properties as well as it has for the single particle set-up. Furthermore, the slight deviation between particles of cube like shape with smooth and sharp edges can be observed.

Finally, a study was carried out on the effects of agglomerates onto the overall elastic material properties. For this the algorithm was edited to place the particle to form a specific cluster. Here two clusters were studied. On the one hand a chain cluster of the particles and on the other hand cloud like cluster. For the chain cluster each particle is placed next to the previously placed one. The specific location for the subsequent particle is randomly selected. The general chain cluster is depicted in Fig. 10 a) b).

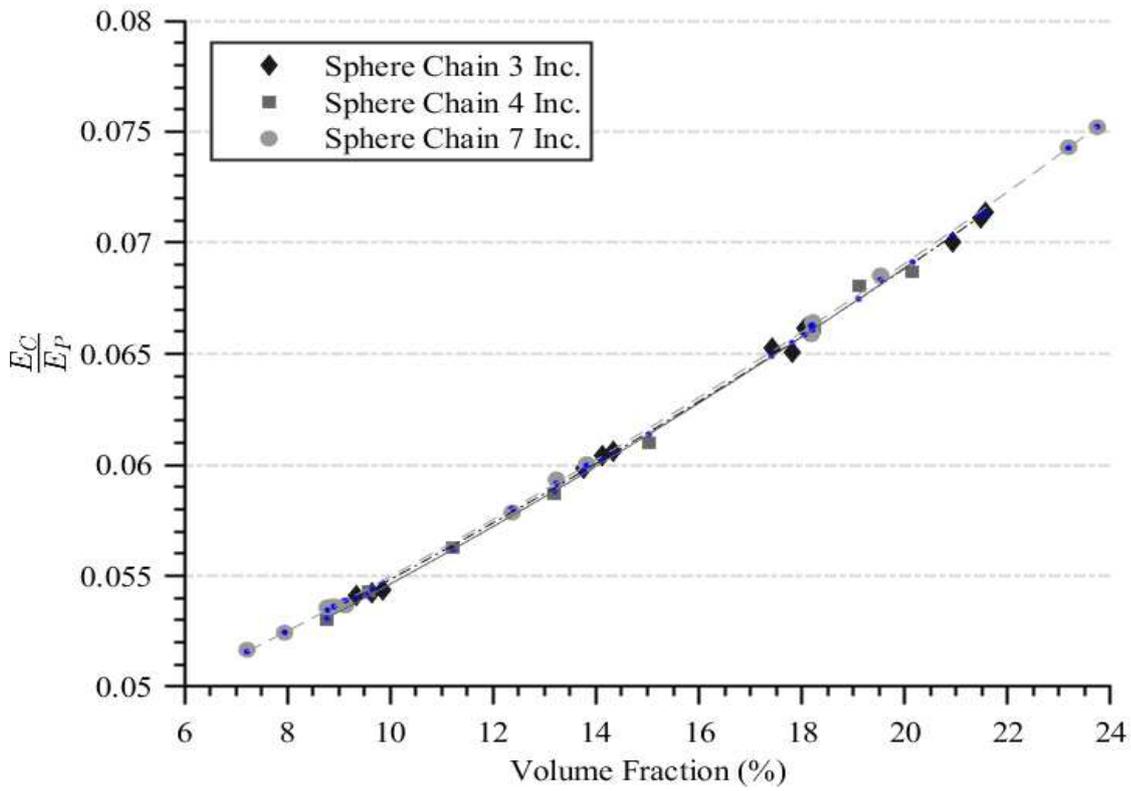


**Fig. 10.** a) & b) Chain cluster, c) & d) Cloud cluster for spherical particles.

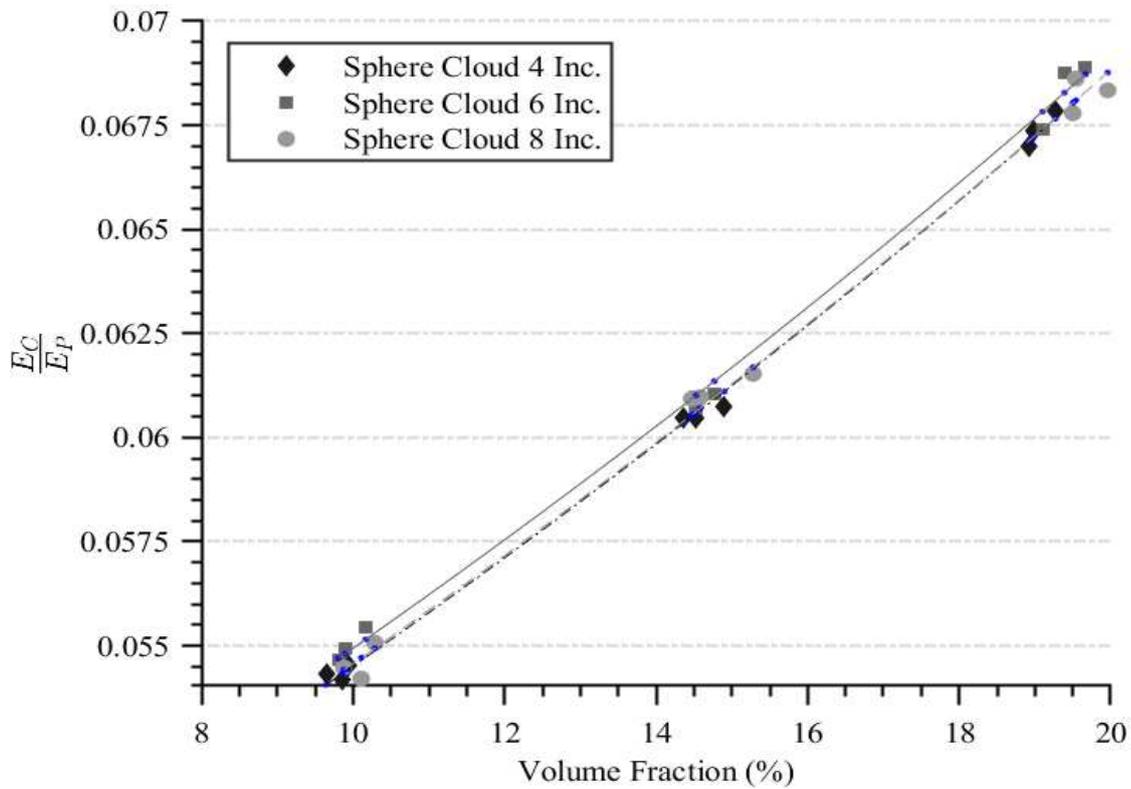
For the cloud cluster all particles are placed next to the initial particle. The specific location of every subsequent particle is randomly selected. This configuration is depicted in Fig. 10 c) d).

Of further interest is the amount of particles and their effects on the overall elastic material properties. For this a study was carried out, which compares the effective Young's moduli of the composite for different chain lengths (Fig. 11) and different sizes (number of added particles) of the cloud cluster (Fig. 12).

The lengths of the chains do not seem to have an effect on the overall elastic material properties. Considering the cloud cluster a slight deviation can be observed.

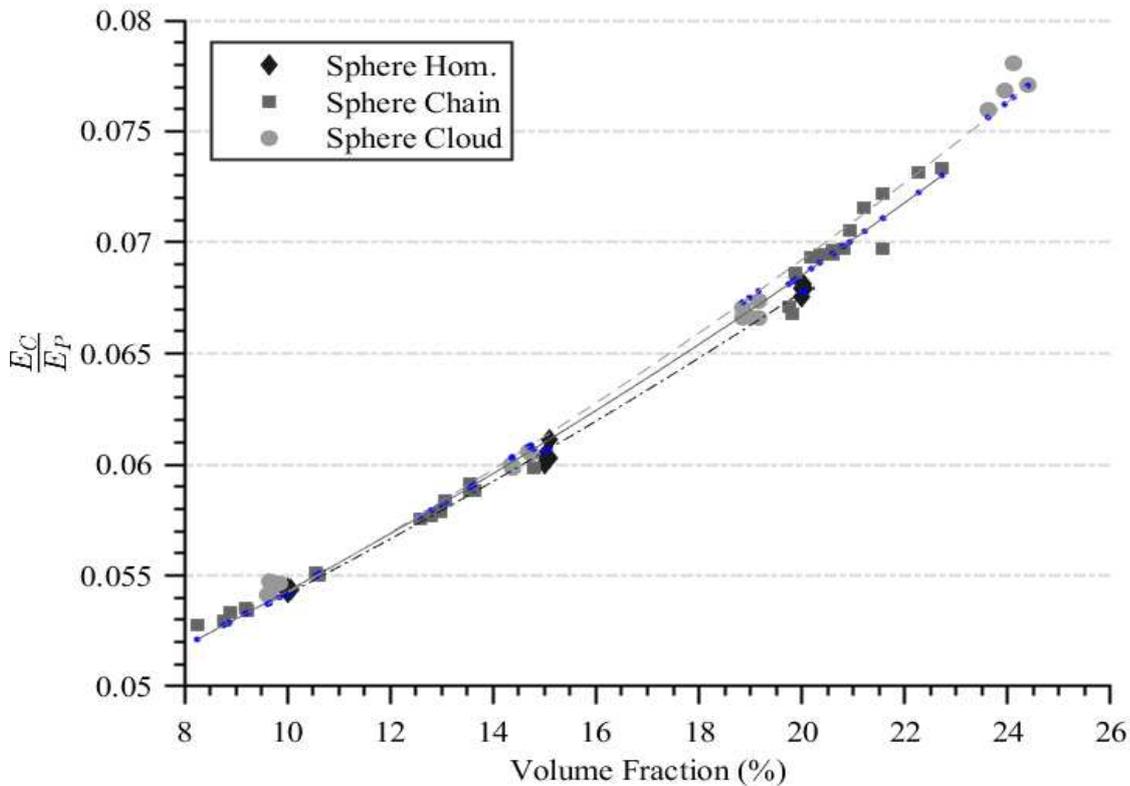


**Fig. 11.** Normalized effective Young's modulus of the composite for different chain lengths for X1 direction.



**Fig. 12.** Normalized effective Young's modulus for different cloud cluster sizes for X1 direction

Finally, the different placement methods, that are periodic distributed single particles, chain and cloud clusters were compared in Fig. 13.

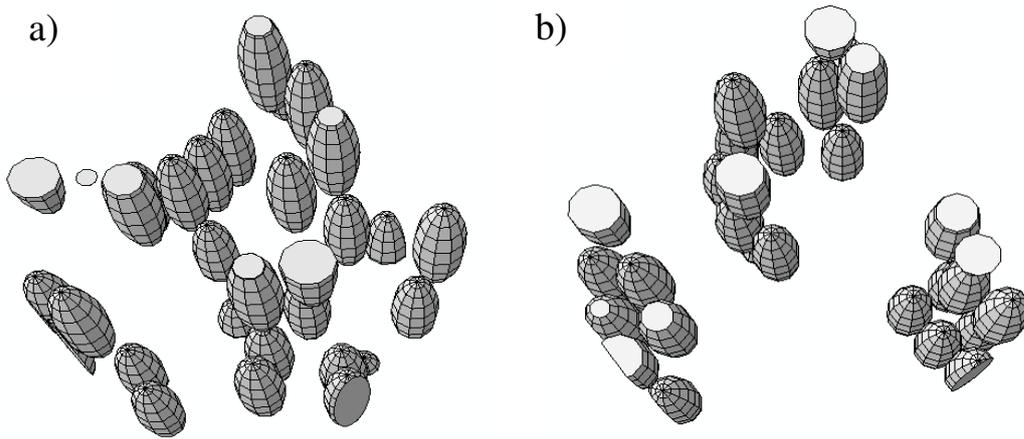


**Fig. 13.** Normalized effective Young's modulus of the composite for different distributions of the spherical particles in X1 direction.

It can be seen here that the cluster configurations achieve greater Young's moduli in general compared to periodically distributed single particles. The difference in Young's modulus even grows with an increase in volume fraction of the particles.

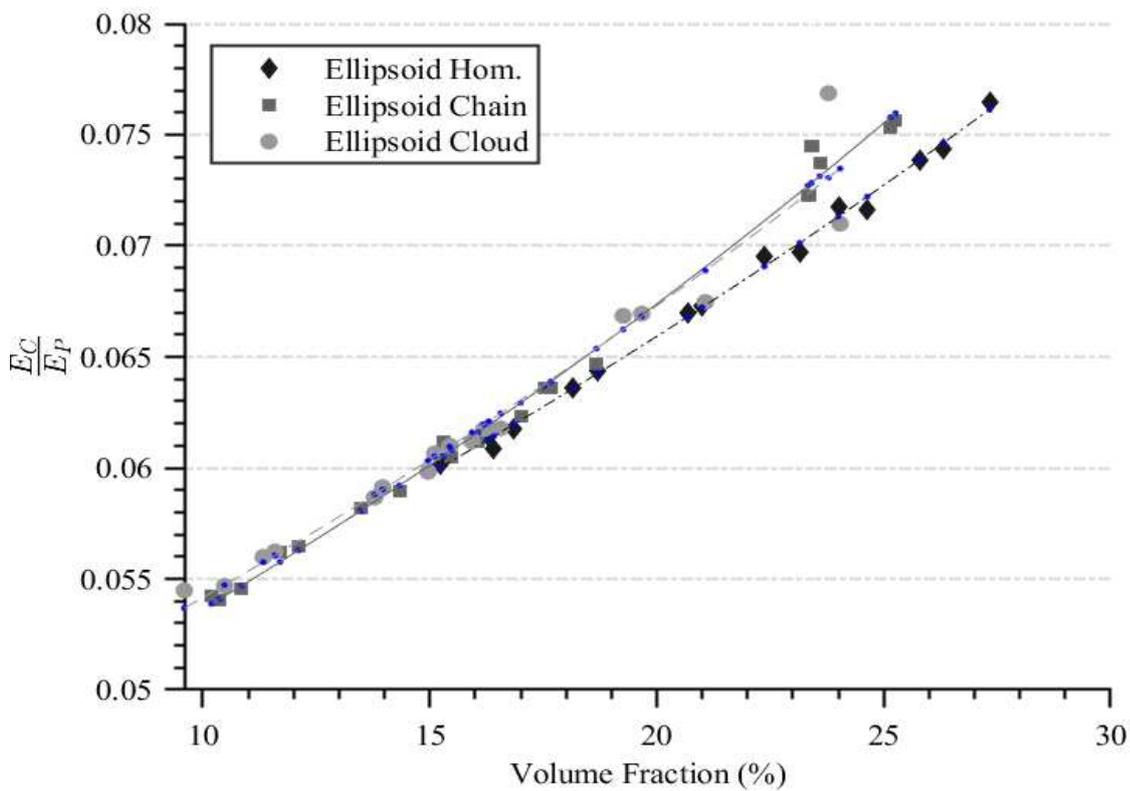
A further interesting result of the evaluation is the difference between the clusters themselves. Different cluster formations do lead to measurably different Young's moduli. Here (Fig. 13) we can see that the cloud clusters do outperform the chain clusters considering the effective Young's modulus of the composite.

In the following steps the influence of the ellipsoidal particles on the effective composite properties was studied. The algorithm used for the placement of the sphere particles beforehand was then altered to calculate the center points of the ellipsoids, which are in contact with each other. The general direction the clusters can grow is random. The difference in the cluster creation methods are which surfaces of the ellipsoids are in contact. Chain like clusters only allow contact between the surfaces of the previously and subsequently placed particles. This way enabling the different forms of particle clusters (chain and cloud). This algorithm was used to create the RVEs, as can be seen in Fig. 14. The ellipsoids are strictly oriented along the X3 axis.

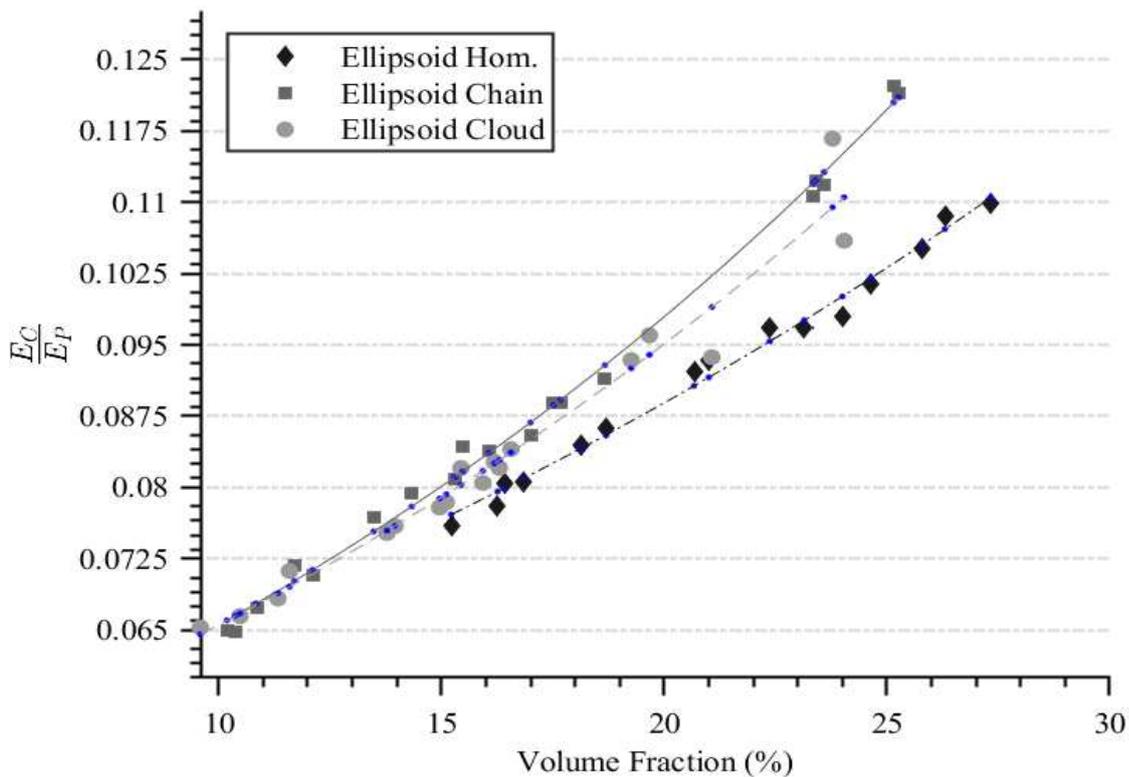


**Fig. 14.** Multiple particle set up for ellipsoidal particles in a) chain cluster and b) cloud cluster

The Young’s moduli for the three different placement methods (periodic single (homogenous), chain and cloud cluster) were evaluated. The results for the X1 direction are depicted in Fig. 15, and those for the X3 direction are shown in Fig. 16.



**Fig. 15.** Normalized effective Young’s modulus of the composite for different distributions of the needle-shaped particles in X1 direction



**Fig. 16.** Normalized effective Young's modulus of the composite for different distributions of the needle-shaped particles in X3 direction

The differences between the cluster and homogenous (periodic singular inclusion) positioning methods of the particles reappear, as was seen in Fig. 13 for sphere particles. Clusters generally outperform the homogenous distribution considering again the effective Young's modulus of the composite.

In stark contrast to the results for the spheres (Fig. 13), this time the chain clusters achieve greater Young's moduli compared to the cloud clusters, which is highlighted for the X3 direction.

The conclusion must be drawn that type of cluster and the particle cannot be studied independently from each other. The root cause of this shift in the clusters influence on the effective elastic properties requires further research.

## 4 Outlook

Both the SEM as well as the numerical analysis are part of a joint project and were carried out in parallel. Thereby, size and shape of the particles and their distribution within the polymer matrix were studied. A numerical model for predicting the elastic properties of the composite was built and successfully validated in numerical calculations.

In the next step, size and shape of the modeled particles as well as their orientation within the matrix will be adjusted to the above-mentioned new findings. Volume elements consisting of homogenous distributions of the particles with scattered clusters

will be studied. In consequence, based on the adjusted design of the microstructure more realistic results are expected.

Furthermore, the investigation of the filler particles will be continued by examining the surface roughness of the particles. In parallel, the mechanical parameters will be also determined by carrying out experimental tests like the standardized tensile test. Thereby, the calculated parameters of the numerical investigation can be verified by comparing them to those of the experiments.

## Acknowledgements

The financial support of the Darmstadt University of Applied Sciences is gratefully acknowledged.

This research was funded by the Federal Ministry for Economic Affairs and Energy based on a decision of the German Bundestag.

## References

1. Baur, E., Harsch, G., Moneke, M.: *Werkstoff-Führer Kunststoffe. Eigenschaften - Prüfungen - Kennwerte*, 11th edn. Carl Hanser Verlag GmbH & Co. KG, München (2019).
2. Rothon, R.: *Fillers for Polymer Applications*. Springer International Publishing, Cham (2017).
3. Alger, M.: *Polymer Science Dictionary*. Springer Netherlands, Dordrecht (2017).
4. Eshelby, J.D.: The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc. Lond. A* (1957).
5. Benveniste, Y.: A new approach to the application of Mori-Tanaka's theory in composite materials. *Mechanics of Materials* (1987).
6. Lielens, G.: *Micro-macro modeling of structured materials* (Ph.D dissertation, UCL/FSA, Louvain-la-Neuve, Belgium, 1999).
7. Kachanov, M., Sevostianov, I.: *Micromechanics of Materials, with Applications*, vol. 249. Springer International Publishing, Cham (2018).
8. Dvorak, G.: *Micromechanics of Composite Materials*, vol. 186. Springer Netherlands, Dordrecht (2013)
9. Quaar, A.: *Micromechanics of rate-independent multi-phase composites. Application to Steel Fiber-Reinforced Concrete*. Dissertation, Université Catholique de Louvain (2006).
10. Trofimov, A., Drach, B., Sevostianov, I.: Effective elastic properties of composites with particles of polyhedral shapes. *International Journal of Solids and Structures* (2017).
11. Khisaeva, Z.F., Ostoja-Starzewski, M.: On the Size of RVE in Finite Elasticity of Random Composites. *J Elasticity* (2006).
12. Gitman, I.M., Gitman, M.B., Askes, H.: Quantification of stochastically stable representative volumes for random heterogeneous materials. *Arch Appl Mech* (2006).
13. Rintoul, Torquato: Reconstruction of the Structure of Dispersions. *Journal of colloid and interface science* (1997).
14. Drach, B.: *Multiscale Numerical Modeling and Characterization of Carbon/Carbon Composites*. Dissertation, University of New Hampshire (2013).
15. Jaklič, A., Leonardis, A., Solina, F.: Segmentation and Recovery of Superquadrics. *Computational Imaging and Vision*, vol. 20. Springer, Dordrecht (2000).

16. ABAQUS, Simulia, Providence, RI, USA, <http://www.simulia.com/>
17. Wadell, H.: Volume, Shape, and Roundness of Quartz Particles. *The Journal of Geology* (1935).
18. Segurado, J., Llorca, J.: A numerical approximation to the elastic properties of sphere-reinforced composites. *Journal of the Mechanics and Physics of Solids* (2002).
19. MathWorks, Natick, MA, USA, <https://de.Mathworks.com/products/matlab.html>

# Integration of Pattern-Based Learning Management Systems with Immersive Learning

## Content

Michael Winterhagen<sup>1</sup>[0000-0002-9487-3544], Dominic Heutelbeck<sup>2</sup>[0000-00003-1395-2775], Benjamin Wallenborn<sup>3</sup>[0000-0001-9234-5310], and Matthias L. Hemmje<sup>1</sup>[0000-0001-8293-2802]

- <sup>1</sup> Chair of Multimedia and Internet Applications (MMIA), Faculty of Mathematics and Computer Science, University of Hagen, Germany {[michael.winterhagen](mailto:michael.winterhagen@fernuni-hagen.de), [matthias.hemmje](mailto:matthias.hemmje@fernuni-hagen.de)}@fernuni-hagen.de
- <sup>2</sup> Research Institute for Telecommunication and Cooperation, Germany [dheutelbeck@ftk.de](mailto:dheutelbeck@ftk.de)
- <sup>3</sup> University of Hagen, Center for Media and IT (ZMI), Germany [benjamin.wallenborn@fernuni-hagen.de](mailto:benjamin.wallenborn@fernuni-hagen.de)

**Abstract.** In our previous e-learning publications we introduced the concepts of *Course Authoring Tools* (CAT), and of *Didactical Structural Templates* (DST) which are a further development of the CAT. DSTs are defined as a possibility to describe the pedagogical structure of a course, a study program or an applied game in an abstract way. The idea of DSTs is based on the structure of IMS Learning Design (IMS-LD), which is a quasi-standard for modelling learning structures. We have shown what DSTs are useful for and that there is a need for an editor for DSTs which we already presented in combination with the *Didactical Structural Template Manager* (DSTM). In this paper we will focus on and go in deeper detail in implementing the same DST as a classic Learning Management (LMS) course, as a LMS course with applied gaming content and as a stand alone applied game. Therefore, these scenarios will show that it will be possible for learners to switch between different implementations of a specific DST and the learners having the same learning progress as if they had used just one of the implementations of this specific DST. One main contribution is to demonstrate the integration between, web-based, immersive Virtual Reality learning activities developed as native applications, and VR authoring tools by a gateway server. In addition, we present an RESTful API for sharing the DST and the used Competence Profiles to external tools.

**Keywords:** IMS Learning Design · IMS-LD · Didactical Structural Templates · Learning Tools Interoperability · LTI.

## 1 Introduction

Producing learning content for *Higher Education Institutes* (HEI) consumers is quite complex in *Learning Management Systems* (LMSs) for producers of

learning content. The *Knowledge-Management Ecosystem Portal* (KM-EP) [6] is an Educational Portal which contains different management systems. A *Course Authoring Tool* (CAT) has been introduced in [16] within the KM-EP to reduce this complexity and make it simple to produce learning content in the way that the producers only have to fill in the necessary information and are able to concentrate in producing the learning content instead of configuring the learning content within the LMS.

### 1.1 Motivation, Problem Statement and Approach

The so-called *Didactical Structural Templates* (DST) have been introduced in [19] and extended in [17]. As described, the DSTs are based on the *IMS Learning Design* (IMS-LD) [8] and represent the pedagogical structure of a course and cannot only be used as pedagogical structure for creating courses. In fact, the DSTs can also be used as a pedagogical structure for a hybrid environment existing of a "classical" course with integrated applied gaming content just like a pedagogical structure for applied game which can be a web-based computer game or a Virtual Reality (VR)/ Augmented Reality (AR) based game. Therefore, one DST can have different implementations.

The advantage of this approach is, that learners will be able to switch between different implementations of one DST whenever they want to and they have got the same learning progress as if they had used only one specific implementation of this DST. This means if learners like gaming, they can use the applied gaming implementation to work on the learning content. If it is easier for the learners to answer the self-tests or the final test – to stay in the exemplary stated pedagogical structure of a course – as e.g. multiple-choice quizzes, they can switch to a course within an LMS to answer the questions.

Within a wide range of domains, realizing learning activities as immersive experiences, e.g., in VR, may significantly improve the learners experience and success [1]. While tools like the WebXR Device API [15] have the potential to close the gap between traditional browser-based learning and content played out via dedicated VR Hardware. In industrial VR learning environments, specialized hardware-based controllers may be necessary for the appropriate simulation of machine operation, and the virtual environments are of high fidelity, derived from engineering CAD systems. Complex environments, such as models of complete factories are very demanding with regards to graphics capabilities of the hardware. Thus, learning activities and self tests in VR often require the use of native application running on a capable machine, connected to specialized controllers. While CATs are often also able to support the authoring of web-based content, immersive applications, both native and WebXR-based, require specialized authoring tools which are able to handle the spatial and graphical content and complex interfaces. In practice, such authoring tools are built on top of game development tools and platforms, such as the Unity [14] or Unreal [5] engine.

Such immersive activities are potentially embedded within a DST alongside traditional web-based content. The integration of this type of content into the

authoring process, deployment and execution of a course instance implementing a DST rises a number of technical and architectural challenges.

During web-based learning in a traditional LMS, the use of external web-based tools is well supported by the *Learning Tools Interoperability* (LTI) [10] Standard by a number of modular APIs and protocols. This allows to start activities implemented using WebGL or WebXR directly. However the reliance of LTI on OAuth [13], *Transport Layer Security* (TLS) [7], and forwarding between tool server and LMS server for launch requests does prohibit directly launching into a native application installed on the users computer, while retaining authentication and course context information. E.g., launch URLs have to be public TLS secured web-resources. A locally running tool cannot fulfill the TLS requirements running on 'localhost'.

To summarize the remainder of this paper addressing our five research questions, which we will require to work on shown below:

1. How can a DST representation be applied to a “classical” course production in an LMS?
2. How is the responsibility of content creation distributed between the different content authoring tools.
3. How can we share and cross-reference content between the different tools and execution environments.
4. Can LTI-style launch requests be realized in order to launch into native applications while retaining the learners authentication and course context?
5. How can native external tools send back results to the LMS?

## 1.2 Methodology

As the basis of our research methodology, the multi-methodological framework of Nunamaker and Chen [12] is used for the structured research and development of information systems. The framework is divided into four phases supporting different methodological strategies: Observation, Theory Building, System Development, and Experimentation. To achieve our research goal to answer our research questions, the methodological phases can be executed repeatedly in any order. It is also possible to return to previous phases.

## 2 Concepts and Technologies

LTI is introduced in [11]. LTI supports a connection between LMSs and external applications like gaming platforms. An applied game prototype based on the *Unity Game Engine* (Unity) [14] has been introduced in [18]. This prototype is later used as an example for implementing an LTI connection. However, this only works for web-applications and not for desktop-applications. To enable this also for desktop applications, the solution presented in [18] has to be extended.

### 3 Conceptual Work

After describing the problems and research questions in chapter 1.1 we want to present our conceptual work in this chapter. To realize the DST as an applied game we have different challenges we have to work on, to make this option work:

1. LTI can only be used for web-based applications, not for stand alone or desktop-applications. How can LTI be used for our scenario?
2. How can the applied game access the DST?

#### 3.1 LTI Launch Requests and Services for Immersive Native Applications

In this section we present our architecture (see figure 1) and implementation of a technical solution for the above described challenges arising from the need for *Immersive Native Applications* (INAs) in order to support immersive VR learning content. Overall, the idea is to use a so-called *LTI Gateway* to mediate between a) the LMS implementing LTI, b) the VR-specific authoring tools, and c) the VR learning tool installed on the users computer. The goal is to support standard LTI 1.3 [10] launch requests to trigger learning activities using the local INA and to allow it to use LTI Advantage [11] services offered by the platform (LMS), e.g. Assignment and Grade Services 2.0 [9].

In LTI 1.3 a platform and an external tool have to be carefully paired and the tool and platforms have to be authorized by the respective administrators. In this process, public keys and several URLs have to be exchanged and unique IDs for tool and platform have to be provisioned. Generally the URLs have to be mutually reachable by the systems hosting either component and it is mandatory for both sides to have all connections secured by TLS supported by the matching certificates.

In order to use the INA, it has to be installed locally on the machine used by the learner. While in special circumstances, this may be a dedicated machine where it is possible to assign a DNS entry and to obtain a TLS certificate. Then, the INA could directly support the LTI protocols by running matching background services. However, in practice learners will not use a single machine shared among them. Thus, even in a learning environment with multiple machines that could fulfill these conditions, in the LTI domain model each machine would constitute a different external tool which implies that in a course the same learning activity on different machines would be listed separately. Furthermore, it is likely that a course is not consumed by a predetermined number of local learners, but by a variable number of remote learners with machines in uncontrolled environments and behind NAT routers. Thus, a direct provisioning of local installations of the INA as LTI tools is not feasible.

In order to use other LTI Advantage services, external tools acquire OAuth 2.0 [13] access tokens. While the OAuth 2.0 standard technically allows for INAs to obtain such tokens directly, it is not guaranteed, that the matching OAuth

flows are supported by all LMS. However the flows for web-based applications to obtain the tokens is mandatory for the LTI standards.

To support both launch request and LTI Advantage services, a dedicated web-based service, the LTI Gateway is introduced. The LTI Gateway itself acts as a single point of entry and an LTI tool for the platform and is able to relay the launch request to an INA and service calls from the INA to the platform. Thus, pairing between tool and platform only happens once for the gateway and not for each local installation of the INA.

The gateway can be referenced in a course like any other external LTI tool. When a user decides to access the learning activity and clicks on the matching link provided by the platform, a standard LTI Launch Request to the gateway is triggered. In the process of the launch request protocol, the gateway obtains information about the user and the context of the launch via different claims stored in the *JSON Web Token* (JWT) [4] used in the launch. In addition, the gateway obtains an access token to use the platforms LTI Advantage services. The course author may have decided to add additional custom parameters (string key-value pairs) to the launch, which are also made available to the tool in the matching JWT claim.

Given, that users may have different sections and activities of a course open at the same time or that users may, often unintentionally, click a launch link multiple times the gateway has both to disambiguate launch requests based on the user and context information.

Thus, the gateway is a stateful service which upon receiving a launch request and obtaining the information creates a gateway session identified by the tuple of user identifier and LTI context. Each session has a predefined, decided by the gateway administrator, *time to live* (TTL) and a randomly generated Base64 encoded 128-bit *Universally Unique Identifier* (UUID), acting as a session key. For this session, the triple, session key, LTI Advantage access token, and JWT token are persistent in RAM for the TTL. Before creating a new session for a launch request, it is checked if another session for the given triple already exists, which is used if present.

After creating, or loading the session, the learner is presented with a TLS secured web page containing information on the origin of the launch and instructions on how to obtain the INA, if the user still needs to install it. The user is presented with a launch link to start the INA. The launch link is structured as follows: [CUSTOM PROTOCOL]:[TOOL\_NAME]?sessionKey=[SESSION\_KEY], e.g., `i21:virtualFactory?sessionKey=jhSLqIopUf_ZiP86n6VdaA`. At install time, the INA registers itself as the handler for the custom protocol (e.g., adding the respective registry entries for Windows). Thus, when clicking the link, the web browser will launch the INA and provide the URL as a command line argument. In this way, the INA is able to obtain the session key. Compared to a direct authorization of the INA and its user via OAuth this opens an attack vector, where on a compromised machine the launch may be intercepted and the session key may be obtained by a malicious application. This however is only possible as a local attack and the key is always encrypted in transit over the network.

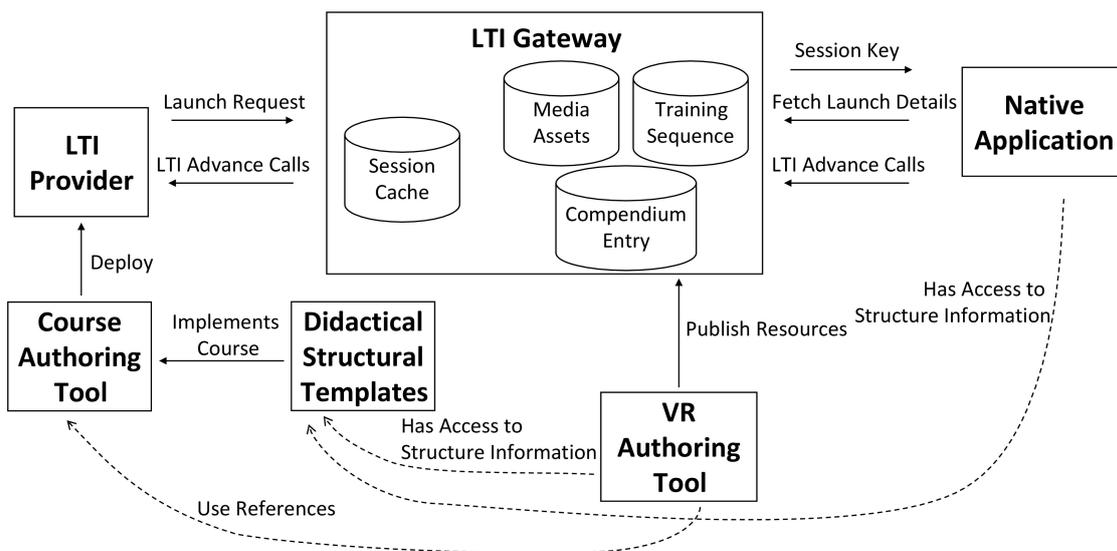
We consider this to be an acceptable remaining risk. This requires local code execution privileges which imply further extensive attack potential.

Given the session key, the INA can retrieve the session data from the gateway by accessing GET: <URL>/launchRequest/{sessionKey}, returning a JSON object containing user information, LTI resource link, line item, and context and custom data provided by the course author.

The INA can now start the learning activity according to this information. For each LTI Advance service the gateway also provides matching proxy services which accept the session key as authentication, such as results or scoring services. Whenever the INA accesses the proxy services with the session key, internally the request is mapped to a matching LTI request to the platform, using the stored access token, which is never shared with the INA.

### 3.2 VR Authoring and Resource Server

Figure 1 gives a rough overview of the interaction of the different parts described in the previous sections.



**Fig. 1.** LTI Gateway and VR Authoring Tools

Within the VR Authoring Tool, the producers can define the Training Sequences, Compendium Entries and Media Assets which are published on a server which holds this content. The meaning of this elements are as follows:

The *Media Assets* can be texts or pictures and can be used by Compendium Entries.

The *Compendium Entries* can be understood as elements of knowledge transfer. They consist of contents in a specific order. Each content is composed of e.g.

texts and pictures which are stored as Media Assets. The *Compendium Entries* e.g. deliver information about the environment of the game or the Training Sequences. To access the Media Assets, the Compendium Entry has to connect to the content server and download it from there.

At last, the *Training Sequences* can be understood as learning elements of a learning path. This Training Sequences consist of so-called *Tasks* which are in a specific order and have to be done in the defined order. Each Task contains so-called *Objectives* which have no specific order. Each Objective will be displayed in a menu in the game. To complete a task, each Objective has to be done completely.

Having described the *Training Sequences* and its structure, we can define a mapping between the IMS-LD and an applied game using the *Training Sequences* as shown in table 1.

**Table 1.** Mapping IMS-LD - applied game

IMS-LD element	applied game element
method	applied game
play	levels
act	Training Sequence
activity structure	Task
activity	Objective

### 3.3 Delivering the Structure of a Didactical Structural Template to an External Application

In this section we present our technical solution for the above described challenge 2. We support two kinds of delivering the structure of an DST to an external application:

1. The *Didactical Structural Template Manager* (DSTM) provides the possibility to export an DST as XML-file which meets the IMS-LD specification. Therefore, the DST could be exported in this format and can be used to design the structure of the game.
2. The KM-EP provides a Learning Design API which will be described in short in the following. The external application can access this Learning Design API to get all information about the needed DST.

The Learning Design API is defined as a *Representational State Transfer* (REST) [3] with the following endpoints:

- GET: <URL>/webservice/learningdesign  
This endpoint returns a list of all available DSTs.
- GET: <URL>/webservice/learningdesign/{id}  
This endpoint returns the structure of a specific DST.

The return value of each endpoint is made in *JavaScript Object Notation* (JSON) [2]. The specific return value's format will not be explained in detail here. With this both kinds of delivering the DST to an external application, it is possible for e.g. an applied game to access the needed DST.

## 4 Prototypical Implementation

After describing the technical requirements and the technical prerequisites, we are now ready to show, how it is possible to implement a DST as

1. a normal Moodle course,
2. a gamified Moodle course, and
3. an applied game.

This will be done within the following subsections.

### 4.1 Implementing a Didactical Structural Template as a Moodle Course

After defining an DST at first we want to implement this into a Moodle course. To do so, we have to go to the DSTM within the KM-EP and press on the *install* button next to the DST we want to implement.

After pressing the *install* button, the DSTM will implement the selected DST as Moodle course according to the mapping we presented in [17] with default values for each learning element. Therefore, we now have an empty Moodle course created which we have to fill with the needed learning content.

To fill this empty Moodle course with learning content, we can now use the CAT within the KM-EP to edit every learning element of the created Moodle course.

After saving the changed learning element, the changes will directly be transferred to Moodle. After the complete course has been filled with the needed learning content, the course is ready to be published and used by learners.

### 4.2 Implementing a Didactical Structural Template as a Gamified Moodle Course

As second option we want to implement the same DST of 4.1 as a Moodle course but with an applied game.

Before we implement the course, there has to be implemented an applied game with an LTI interface. For our example, we use a web-based applied game implemented with the *Unity Engine* (Unity) [14].

We already described in detail in [18] how the LTI connection between Moodle and the applied game works. We will show exemplary, how the implemented applied game can be included within an Moodle course.

At first the URL of the created applied game has to be added as software into the KM-EP. After the applied game is known within the KM-EP, it can be

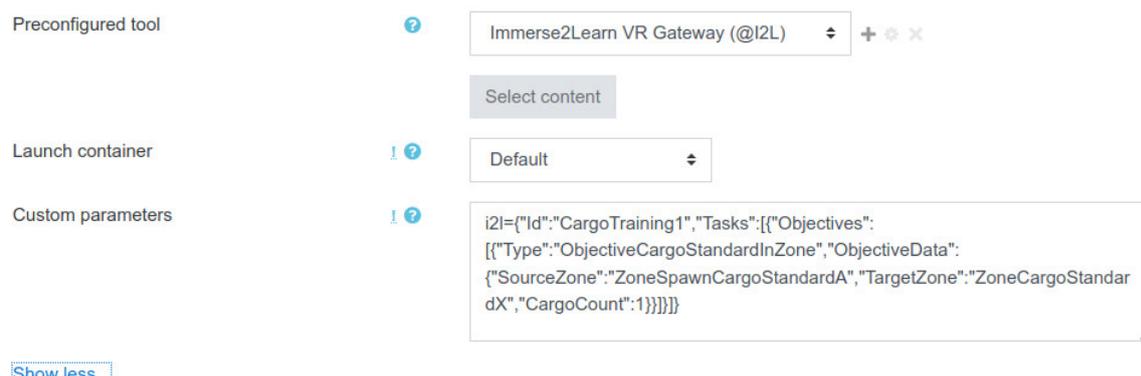
referenced by a Moodle course. Therefore, an learning element has to be defined as asset which references the applied game with help of the CAT.

When a learning element with an underlying applied game is selected by the learners, the applied game will start and the learning result will be transferred back to the LMS via LTI.

### 4.3 Implementing a Didactical Structural Template as an Applied Game

As third option we want to implement the same DST of 4.1 as a stand alone applied game. In our scenario, this game will be an VR/AR game.

**Prototypical Implementation of LTI connections for Stand Alone Applications** To make it possible to call the LTI Gateway mentioned in 3.1, we have to define an LTI connection in Moodle (see figure 2).



The screenshot shows the Moodle LTI connection configuration interface. It has three main sections:

- Preconfigured tool:** A dropdown menu showing "Immerse2Learn VR Gateway (@I2L)" with a plus, star, and close icon to its right.
- Launch container:** A dropdown menu showing "Default".
- Custom parameters:** A text area containing the following JSON:
 

```
{
    "id": "CargoTraining1",
    "tasks": {
      "objectives": [
        {
          "type": "ObjectiveCargoStandardInZone",
          "objective_data": {
            "source_zone": "ZoneSpawnCargoStandardA",
            "target_zone": "ZoneCargoStandardX",
            "cargo_count": 1
          }
        }
      ]
    }
  }
```

At the bottom left, there is a "Show less" button.

**Fig. 2.** Define LTI connection in Moodle.

For our purpose we need at least version 1.3 of the LTI specification. After the LTI connection for the LTI Gateway has been defined within Moodle for a specific learning element, a click on the corresponding learning element will open a website which provides a link to the applied game.

After clicking on the provided link, an applied game will be downloaded and started. This applied game will contain the LTI connection to return learning results to Moodle.

## 5 Conclusions

In this paper we have shown how to implement a given DST

- as a normal Moodle course,
- as a gamified Moodle course, and
- as an applied (VR) game without Moodle or another LMS.

We also have extended the LTI connection, which until now only could be used for web-based applications, for stand alone applications via an LTI Gateway. We used this LTI Gateway for an applied game which implemented an DST.

To give an external application access to the pedagogical structure of an DST we offered two options:

1. exporting the pedagogical structure of an DST via the extended IMD-LD specification and
2. providing a REST-API which provides the pedagogical structure of an DST in JSON format.

If we have two different DSTs, namely *DST A* and *DST B*, the learners learning progress can be further developed, when the learners switch between different implementations of *DST A*. If they switch into an implementation on *DST B*, they will have another, separate learning progress apart from *DST A*.

## 5.1 Future Work

Future work will be a further formal evaluation of the DSTs and the DSTM, especially in the way to use the DST as pedagogical structure for an applied game without connecting it to a Moodle course. Also a formal further evaluation will be based on learners switching between different implementations of an DST having the same learning progress.

The process of creating a Moodle course by installing it from an DST is not as good as it could be. To improve the installation process, it would be desirable to have a wizard, so that the learning content can already be filled in while installing the DST as a course.

We also have to extend the possibility to connect to applied games via LTI for desktop-applications.

## Acknowledgements

This publication has been produced in the context of the RAGE and Immerse2Learn projects. The projects have received funding from the European Union's Horizon 2020 Research and Innovation Action under grant agreement No 644187 and from European funds for regional development (EFRE). However, this paper reflects only the authors' views and the European Commission is not responsible for any use that may be made of the information it contains.

## References

1. Allcoat, D., von Mühlen, A.: Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology* **26** (2018). <https://doi.org/https://doi.org/10.25304/rlt.v26.2140>, <https://journal.alt.ac.uk/index.php/rlt/article/view/2140>

2. Crockford, D.: The javascript object notation (json) data interchange format (2017), <https://tools.ietf.org/pdf/rfc8259.pdf>
3. Fielding, R.T.: Architectural Styles and the Design of Network-based Software Architectures. Ph.D. thesis (2000), [https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding\\_dissertation\\_2up.pdf](https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation_2up.pdf)
4. Force, I.E.T.: Rfc 7519 - json web token (jwt) (2015), <https://tools.ietf.org/html/rfc7519>
5. Games, E.: The most powerful real-time 3d creation platform - unreal engine (2020), <https://www.unrealengine.com/en-US/>
6. Globit-GmbH: Ecosystem portal, <https://www.globit.com/#ep>
7. IETF: The tls protocol (1999), <https://tools.ietf.org/html/rfc2246>
8. I.G.L.-Consortium: Learning design specification (2003), <https://www.imsglobal.org/learningdesign/index.html>
9. I.G.L.-Consortium: Ims learning tools interoperability (lti) assignment and grade services specification (2019), <https://www.imsglobal.org/spec/lti-ags/v2p0/>
10. I.G.L.-Consortium: Learning tools interoperability core specification (2019), <http://www.imsglobal.org/spec/lti/v1p3/>
11. I.G.L.-Consortium: Lti v1.3 and lti advantage (2019), <https://www.imsglobal.org/activity/learning-tools-interoperability>
12. Nunamaker, J.F., Chen, M., Purdin, T.M.D.: System development in information systems researchs (1990), <http://gkmc.utah.edu/7910F/papers/JMIS%20systems%20development%20in%20IS%20research.pdf>
13. OAuth: Oauth 2.0 (2012), <https://oauth.net/2/>
14. Unity: Unity technologies, <https://unity3d.com/de>
15. W3C: Webxr device api (2019), <https://www.w3.org/TR/webxr/>
16. Wallenborn, B.: Entwicklung einer innovativen Autoren Umgebung für die universitäre Fernlehre. Ph.D. thesis, FernUniversität Hagen (2018), [https://ub-deposit.fernuni-hagen.de/receive/mir\\_mods\\_00001428](https://ub-deposit.fernuni-hagen.de/receive/mir_mods_00001428)
17. Winterhagen, M., Hoang, M.D., Lersch, H., Fischman, F., Then, M., Wallenborn, B., Heutelbeck, D., Fuchs, M., Hemmje, M.: Supporting structural templates for multiple learning systems with standardized qualifications. In: EDULEARN20 Proceedings (2020)
18. Winterhagen, M., Salman, M., Then, M., Wallenborn, B., Neuber, T., Heutelbeck, D., Fuchs, M., Hemmje, M.: Lti-connections between learning management systems and gaming platforms. Journal of Information Technology Research **13** (2020)
19. Winterhagen, M., Then, M., Wallenbrn, B., Hemmje, M.: Towards structural templates for learning management systems taking into account standardized qualifications. Tech. rep., FernUniversität Hagen (2019), [https://www.researchgate.net/publication/340601087\\_Towards\\_Structural\\_Templates\\_for\\_Learning\\_Management\\_Systems\\_taking\\_into\\_account\\_Standardized\\_Qualifications](https://www.researchgate.net/publication/340601087_Towards_Structural_Templates_for_Learning_Management_Systems_taking_into_account_Standardized_Qualifications)

# Visual Dashboards in Trend Analytics to Observe Competitors and Leading Domain Experts

Lennart Sina<sup>[0000-0003-0157-1473]</sup>, Dirk Burkhardt<sup>[0000-0002-6507-7899]</sup>,  
Kawa Nazemi<sup>[0000-0002-2907-2740]</sup>

Darmstadt University of Applied Sciences, Human-Computer Interaction &  
Visual Analytics group, Germany

stlesina@stud.h-da.de, {dirk.burkhardt, kawa.nazemi}@h-da.de

**Abstract.** The rapid change due to digitalization challenge a variety of market players and force them to find strategies to be aware of developments in these markets, particularly those that impact their business. The main challenge is what a practical solution could look like and how technology can support market players in these trend observation tasks. The paper outlines therefore a technological solution to observe specific authors e.g. researchers who influence a certain market or engineers of competitors. In many branches both are well-known groups to market players and there is almost always the need of a technology that supports the topical observation. This paper focuses on the concept of how a visual dashboard could enable a market observation and how data must be processed for it and its prototypical implementation which enables an evaluation later. Furthermore, the definition of a principal technological analysis for innovation and technology management is created and is also an important contribution to the scientific community that specifically considers the technology perspective and its corresponding requirements.

**Keywords:** visual trend analysis, visual analytics, business intelligence, information exploration, innovation management

## 1 Introduction

Market trends are a major opportunity and risk in the ongoing digitalization of current markets. A variety of new players that are well known today, for instance Tesla, Uber Pop, Lieferando are still relatively new web giants like Amazon or Google, who have had huge success in establishing a successful digital business. Most of their technologies are used in our daily life. Many of those technologies are replacing or outpacing traditional businesses with the help of information technology and are making their usage easier and more comfortable. These new giants can benefit from the digitalization, as they have been making use of the opening opportunities. At the same time other giants fall because they fail to recognize these opportunities. And even more, they are unable to complete a sustainable digitalization strategy in time. The list of losing players, like AOL, Quelle, or Kodak, is long and it shows, how crucial trend observation in

a fast-changing world has become and how important it is to react to those trends timely and sustainably.

However, trend observation is quite a complex issue that requires experts with significant domain and technological knowledge. It also requires tools that support a periodical trend observation in a variety of data sources. Trends can be found in many data sources. To register them as an important upcoming topic, a considerable amount of data from a variety of sources is needed. In addition to data (sources) and expert users, there is also a third component that is essential to successfully identify trends: A technological solution to perform an appropriate data analysis in perspective of trends.

Therefore, this paper aims on introducing a new technical approach to observe trends in perspective of competitors and leading domain experts to enable an early trend recognition. Instead of starting the trend analysis from the topical point by direct searches or topical listings, what we already investigated [12], we propose a different approach by beginning the analysis with “the inventors” – so to say the persons behind developed new technologies. The idea aims more on monitoring certain groups of inventors instead on investigating principal new developments in general. This kind of monitoring is common in highly innovative fields or market niches with just very few competitors. The approaches oversee research and engineering actions of persons in form of authors and enable analysts to early identify new findings and developments that might be impacting the analyst’s business.

## 2 Foundations to Visual Trend Analytics

*Trend analysis* is a frequently used term in several domains, which makes it essential to define it more clearly to reach a common understanding. To find an access to the topic of visual trend analytics, we must first investigate the current state of the art of several closely related fields. A well-known area in which trend analytics is regularly performed is the banking industry, where market analysts try to predict future stock price movements based on temporal data [1]. However, in this paper we consider (Visual) Trend Analysis always in perspective of economic and market analysis [2].

In general, a *trend* is defined as “a method of identifying and describing specific changes over a long period of time, and the future can thus be predicted using past patterns” [3]. Complementary to this definition, *trend detection* can be defined as “Knowledge of emerging trends is particularly important to individuals and companies who are charged with monitoring a particular field or business. [...] Manual review of all the available data is simply not feasible. Human experts who are tasked with identifying emerging trends need to rely on automated systems as the amount of information available in digital form increases [4].” While trend detection is a relevant task within trend analysis, we follow the definition of trend *analysis* by Sharma et al. [6] where it is “a process of estimating gradual change in future events from past data”. Trend analysis is a quantitative method that requires precise specification of objectives that will be fulfilled from such investigation of data. The following list shows different goals,

presented by Chandler and Gray [6][7], that can be achieved by performing trend analysis:

- To describe the past behavior of a process.
- To try and understand the mechanisms behind observed changes.
- To make assessments of possible future scenarios, by extrapolating past changes into the future.
- To enable analysis of systems where long-term changes serve to obscure the aspects of real interest.
- Detect and estimate the magnitude of a trend.
- Identification of time periods in which there was a substantial trend and times in which there was negligible trend.
- To predict and forecast a trend.

## 2.1 Relevance of Trend Analytics in Innovation and Technology Management

Nowadays, most innovations are not only a random side-effect of daily work. It can be described as a much more seriously organized procedure. Especially international companies have big research and development (R&D) departments, in which they follow a well-organized innovation process including a type of innovation center. In contrast to other departments, the R&D departments have a higher flexibility and less focus on the company's business. However, the intention in all their activities is to extend the company's products and services. For that purpose, the innovation center is building the bridge between research actions and normal departments.

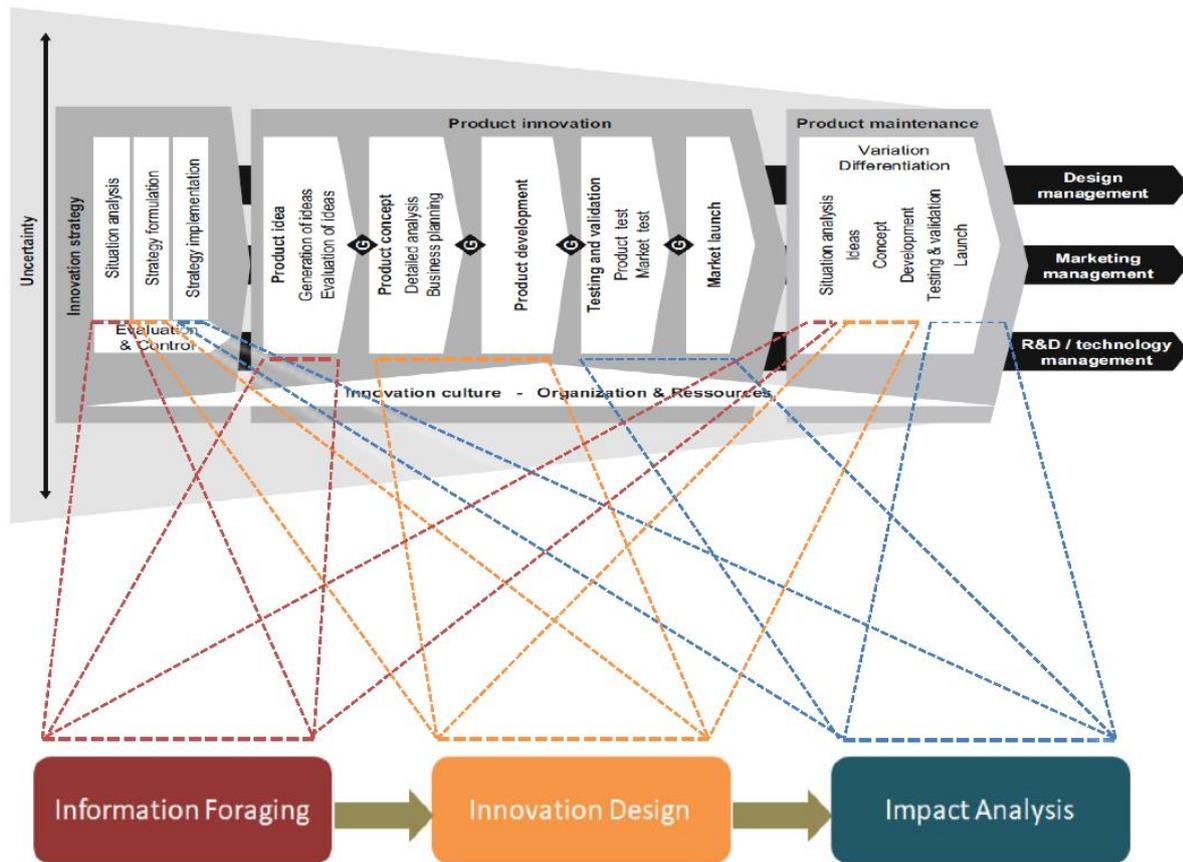
One of the more established innovation process models, which a variety of organizations have implemented, is the one of Gaubinger et al. (top model at Fig. 1). The process describes the innovation process in a tripartite process during which the uncertainty decreases. Each phase is subdivided into separate stages.

The first phase of the process is "innovation strategy" during which the situation is analyzed until a strategy can be formulated and implemented. The process is subject to steady evaluation and control and covers three stages.

In the second phase, called "product innovation", the general uncertainty has already decreased. Furthermore, due to the lower uncertainty the steps are more precise and cover a higher number of stages, namely five. It starts with the first stage in which the product ideas are generated and evaluated. Afterwards, a concrete product is conceived which involves a detailed analysis of the product and general business planning. After the conceptualization stage has come to an end, the product can be developed. Before the product is launched on the market, it should be tested to ensure the quality and the market's acceptance.

The last phase is the maintenance of the product. This phase is mostly a variation differentiation which might involve important security updates, new features, or simple design differences. Every variation goes through different stages before its launch, which are very similar to the second phase. Because of the similarity to the second phase, the number and stage objectives are similar, except for one additional to perform

the situation analysis, in which current product situation is analyzed and ideas for variations or rather advancements are gathered. Those ideas are conceived, developed, and tested before they are launched.



**Fig. 1.** The figure shows on top the innovation process by Gaubinger et al. [8] and on the bottom our simplified analysis model for performing visual trend analysis. The transformation and model reduction are sketched with the lines.

From a technical point of view, the entire innovation process can be supported with a suitable analysis software. Especially the research for existing approaches, algorithms or available technologies and systems can be supported and it is in principle in any phase almost similar. From the perspective of research systems, all three phases of the innovation process can be summarized into a three-stage model, which takes the main research objectives and goals into account (Fig. 1 shows on the bottom the three-stage innovation research model and covers also the transitions to the original innovation model above).

The first step includes the process of analyzing the current situation and developing ideas. As the first step mostly depends on the finding of information, we call it “information foraging”.

The second step involves creativity in choices for the innovative product. This includes creative new ideas and their conceptualization as well as their development. As this step involves creativity and design choices for the innovation, we call it “innovation design”.

The last step that can be identified is what comes after the product is developed, which mostly contains tests and validation phases. As it is about analyzing the impact of the product, we are going to call that phase “impact analysis”.

As the identified three phases are not static and include a lot of trial and error, the process and the identified phases are an iterative process, which may have to start over at any step.

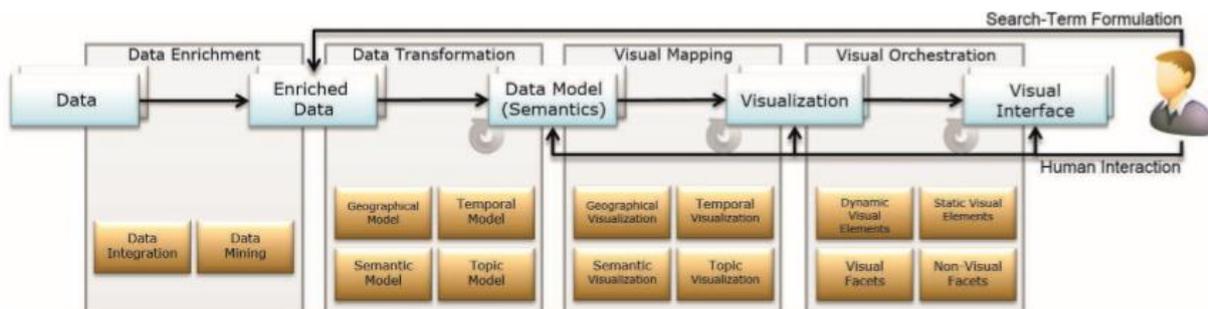
## 2.2 Visual Trend Analytics Methodology

An important aspect of Visual Trend Analytics is Visual Analytics, which is also the main interface functionality with which the user gets in contact. Keim et al. [9] describe visual analytics as: “[...] an iterative process that involves information gathering, data preprocessing, knowledge representation, interaction and decision making. The ultimate goal is to gain insight into the problem at hand which is described by vast amounts of scientific, forensic or business data from heterogeneous sources.”

Visual trend analysis combines trend analysis with the benefits of visual analytics.

Nazemi et al. [10, 12] describe their approach to visual trend analysis as a process which starts with the extraction and identification of key topics and their relevance, over a specific period, from scientific publications. The extracted information can be displayed in visual analysis tools which support users in the process of identifying trends and making decisions.

They developed an approach that utilizes the benefits of visual analytics to gather thus topic information and visually display the data inside several visualizations.



**Fig. 2.** Our Conceptual Model includes the five transformation steps: Data is enriched, structured, mapped to Visualizations, and orchestrated into a Visual Interface [10, p3].

The starting point of their approach lies in the formulation of a search term by a user. Depending on the search query, data is enriched and transformed into different data models. Those data models can be made for geographical, temporal, semantic or topic information. In a process of visual mapping the data is transformed and combined to be displayed in different visualizations which are orchestrated and displayed inside a visual user-interface. The interface enables user interactions on different levels. The user may change the data models used, the data displayed or interact with the visualizations.

With their approach they try to deliver a system to answer several key questions which are relevant when identifying technological trends.

They define these questions as the following [10]:

1. When did technologies or topics emerge and when did they establish?
2. Where are the key-players and key-locations?
3. Who are the key-players?
4. What are the core topics?
5. How will the technologies or topics probably evolve?
6. Which technologies or topics are relevant for a certain enterprise or application area?

### **3 Visual Dashboard Approach to Observe Trends of Researchers and Inventors**

In order to make objective decisions, decision-makers from various industries (e.g. the mobility, transportation and logistics domain [17]) need information about the behavior of the competition.

Abstract data gathered from text mining algorithms can contain such information about the competition, but neither has that data only quantitative measures that can be visualized nor is their dimensionality just two. They have far more different perspectives or aspects on data that could help to achieve the main goal of information visualization. There are different ways to visualize these data, e.g. by super-imposing visual variables or visualizations and by juxtaposing those into visualization dashboards [15, 16]. Commonly superimposing visualizations leads to overstraining users due to the complexity of the visual representations. Therefore, juxtaposing visualization as so-called dashboards or visualization dashboards leads to a more efficient interaction of the visualized data. But there are several ways to create such dashboards and different methods to ensure an easy interaction to meet the users' requirements [11].

#### **3.1 Visual Trend Analytics in Innovation Management**

The process of innovation management starts according to Gaubinger et al. [8] with the development of an innovation strategy. The first step when building such a strategy is an analysis of the current situation.

The information gathered in the process defined by Nazemi et al. [10] can assist in this process, by providing valuable information about technologies and the most relevant authors in different research areas. As the information comes in huge amounts of nested datasets, the data must be transformed into data models which can be visualized and orchestrated.

### 3.2 Data Processing for Dashboard Generation

There are several steps of data transformation applied before the data is ready to be displayed inside the different visualizations. As fundament we consider the already established data processing methodology in theory and practice as described and explained by Nazemi et al. [12].

At first, enriched data is gathered with the text mining algorithm LDA, which uses methods of unsupervised learning to convert collections of discrete data into topic models [14]. The data is indexed, so that it can be requested from a database.

As the data still has not reached its final form in which it can be displayed inside different visualizations, there are several relevant steps to undertake before doing so.

The first step is to convert the data into different data models, according to the transformational steps shown in figure 2. Data models for the topics, authors and the temporal development are needed.

Those models can be used and combined into visual structures. An example for the combination of the data structures is the process of receiving the name of a specific topic. As the topics are represented by numeric values which refer to a specific entry inside the topic data model, many models must be combined with the topic data model.

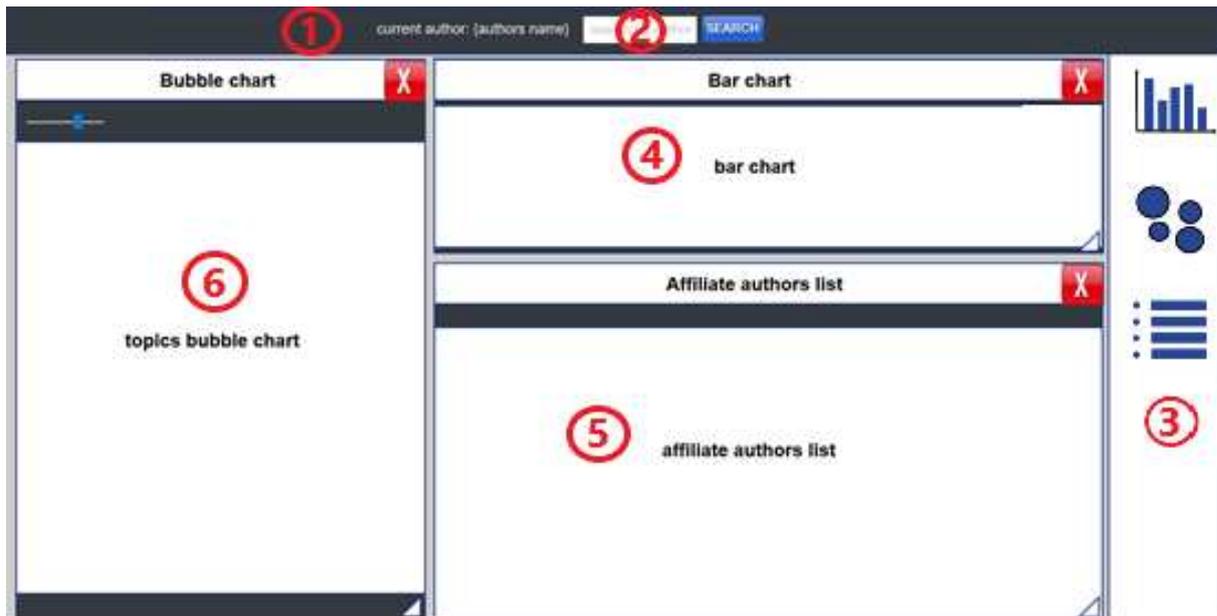
Before the final data can be displayed, data cleansing is necessary. Nonexistent values must be marked and displayed for completeness. For example, if there was a publication of an author without a publishing year, the value is set to 0 and will be displayed as the first bar in the bar chart.

### 3.3 Dashboard Design for Author Analysis

In the next section, we describe the developed analysis dashboard according to the different analysis stages, based on Shneiderman's Visual Information-Seeking Mantra [13].

#### Overview Phase

The overview phase provides basic information on an author's research behavior, the main topics of his publications and his co-authors.



**Fig. 3.** Wireframe initial view on the dashboard application.

Figure 2 shows the initial view on the dashboard application. (1) is the navigation bar at the top of the screen. It displays the current author, that means the author the dashboard is currently created for. It can be set initially to the most interesting author.

The next component to look at is the search field. The user may search for an author, by typing his name into the search field and clicking the search button. This will automatically change the name of the current author and it will also reload the entire dashboard.

Another navigation bar is (2) which is on the right side of the screen. The side navigation is displayed on the right side of the window. It contains several draggable components which the user must drag to the screen and drop, to display a new visualization.

The dashboard which contains the charts (4)(5)(6) fills the rest of the screen.

The first visualization is a bar chart (4). The chart's y-axis indicates the amount of publications the searched author has published. The x-axis indicates the year in which he published the specific amount of publications. If there are publications with no year given, the x value is set to zero. The ticks on the y-axis mark the amount of publications in one year, starting from zero.

The list (5) displays the different co-authors of the searched author. The list is scrollable and displays the data in the format: first name, last name, amount. The amount depends on the number of publications they published together.

The bubble chart visualization (6) displays the main topics that the searched author explored in his publications. It has a slider to filter the main topics by their occurrence.



**Fig. 4.** Wireframe interactive behavior of the dashboard. Positioning and Resizing

Figure 4 shows the communication between the right navigation bar and the dashboard. As mentioned on the arrow the user may drag any visualization from the side navigation into the dashboard and drop it there. The dashboard shows the new visualization depending on which of the items the user dragged into it.

(1) The user can change the position of every visualization on the screen with drag and drop. The initial positioning of the visualizations is just a suggestion. The visualizations affected by the drag, because the dragged item is situated above them, are dodging to the next free position automatically. The dragged item has an on-hover effect. Marked in the dashboard by a blue shimmering underneath, as shown in figure 4. This is displaying where the visualization will be dropped on the screen.

(2) shows how to resize any visualization on the dashboard. Every visualization is inside a container which has a resize handler in the lower right corner. The user can resize the container and the visualizations resize too.<sup>1</sup>

### Analysis Phase (Data Zooming and Filtering)

After gaining a first impression on an author's current research activity and the main research of his studies a more detailed view on the data can help investigate the work of an author. For tasks like those we provide a more detailed view on the data as shown in figure 5.

<sup>1</sup> There is a video of the first prototype of the system available on: <http://s.vis.h-da.de/video-vis-dashboard> (last accessed: 31/08/2020)



**Fig. 5.** Detail view on the dashboard application

(1) The first chart is not a typical chart. It shows the ranking of a specific facet like topics or authors displayed as slices of different width. The width refers to the amount in comparison to its neighbors. The vertical position refers to the ranking that facet has, as they are sorted descendant. The colors of the slices change when a user hovers and highlights the same facet over different years, making it easy to analyze the temporal development of a facet over the years.

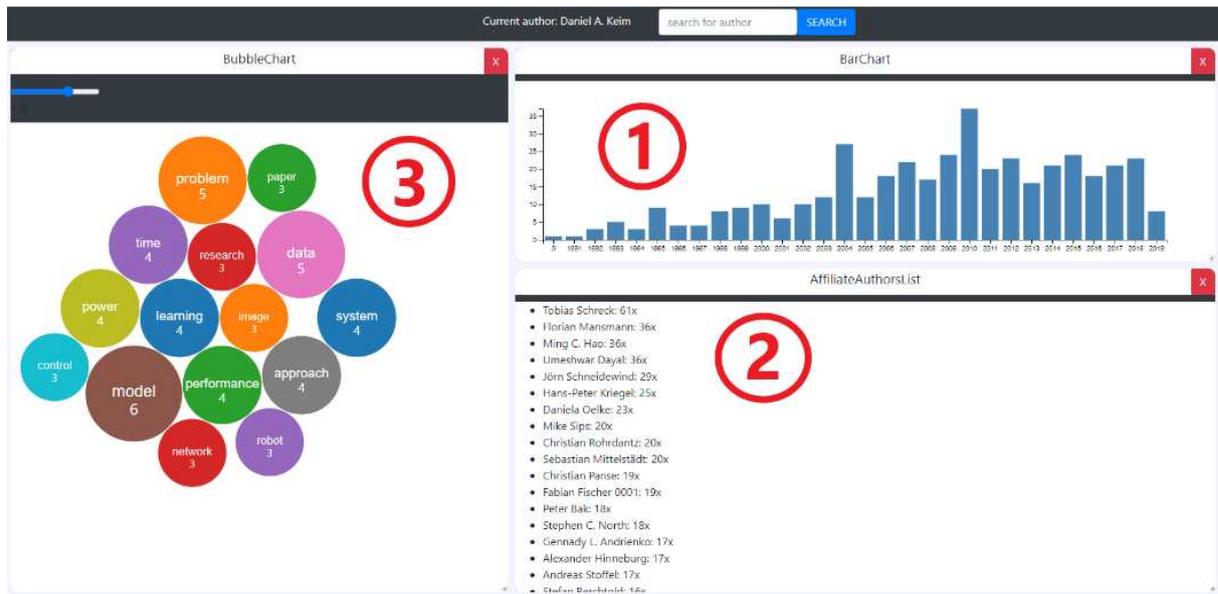
(2) The second chart is a list view for specific publications by the author. It delivers information about the title and the involved authors. List items expand on click and display several other relevant pieces of information like the abstract of the publication, the main topics, and hyperlinks to the publications.

### Details Phase

A competitor analysis can require very specific knowledge of the work of an author. It is therefore necessary to provide as much details as possible. Figure 6 shows on the right side the view on one specific publication. It provides links to the full texts of the publications.

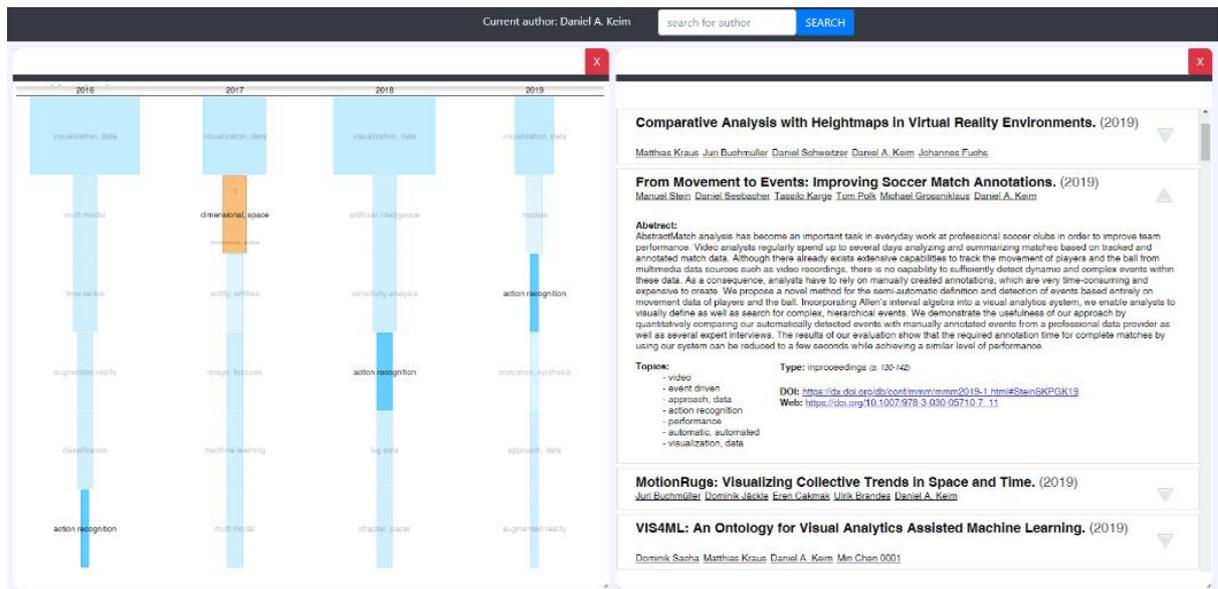
## 3.4 Prototypical Implementation

Figure 6 shows the prototypical implementation of the overview dashboard which is currently created for the author “Daniel A. Keim”. The implementation strictly sticks to the planning in the wireframes.



**Fig. 6.** An overview dashboard which was created for the author Daniel A. Keim

The prototypical implementation of the analysis and details phases can be viewed inside figure 7.



**Fig. 7.** A detail view dashboard for the author Daniel A. Keim

## 4 Use Case Competitor and Leading Domain Experts Observations

The application developed is an expert system which may deliver specific insights into the research work of competitors. For that reason, the general audience of the system is a small group of decision-makers.

To outline the benefit of our developed approach better, we will now write our description further based on the exemplary scenario. To be aware of all activities relating to your market, almost any leading actor is interested in a solution to observe research and developments of its field and particularly those of scientific influencers and competitors. So, they need a software system that allows them to identify the rest of the newly formed research team and they also need to clarify the main topics of their former and current re-search. Moreover, they must keep track of that research groups activities every month to investigate their current state of research.

### **Use Case Objective**

Let us imagine your role in a company is to be head of department toward visual analytics solution. In this position you want to be aware of any research activities in the field e.g. in form of news or research papers, especially those that are relevant for your market. Of course, you also want to know about any activities of you competitors. From the research field you will follow researchers in the R&D domain such as Daniel Keim. Similarly, you would do it with competitors.

### **Use Case Description**

In the first step an overview dashboard for the author Daniel Keim is created.

As one can see in figure 6, the overview dashboard provides different visualizations. This helps receiving different information about the searched author Daniel A. Keim.

The first information (1) delivers a temporal overview of the amount of publications over the last years. As one can see, Keim has been publishing significantly over the last years.

(2) provides an overview of his co-authors. The top co-author is Tobias Schreck. They have 61 publications together.

(3) provides a basic overview of the main topics of his publications.

As the overview dashboard - which is initially created - only delivers a basic overview, we need to dive deeper into the data to find out more about that author's work.

Figure 7 shows details about that author's publications. Chart (1) shows a ranking of the main topics for the last four years.

Chart (2) shows a list of the publications that author has published.

## **5 Conclusion**

All in all, one can say that metadata gathered by text mining algorithms (visualized, and juxtaposed inside visual layouts like dashboards), provides specific insights and knowledge into technological trends and people. This comes in handy in innovation management and competitor research. This is a specific need a lot of enterprises have. They need to be aware of changes in their market early and react to those changes fast in order not to fail the future customers' requirements.

The prototype developed in this paper tried to show a comprehensible use-case on how such a dashboard can be used to analyze a specific competitor. Even though it is

only a prototype it still presents a variety of benefits to observer market trends of specific persons. It is important to remind ourselves that for almost any market and business just a few influencing persons exist beside well-known competitors. This makes this approach so effective. It allows us to be aware of changes and new directions in a field by just observing a handful of people.

The future goal is to implement the dashboard functions for observations next to the already existing analysis solution for visual trend analytics to achieve an encompassing solution for trend analysis.

## Acknowledgements

This work was partially funded by the Hessen State Ministry for Higher Education, Research and the Arts within the program “Forschung für die Praxis” and was conducted within the research group on Human-Computer Interaction and Visual Analytics (<http://www.vis.h-da.de>).

## References

1. Liheng, Z., Aggarwal, C., Qi\*, G.: Stock Price Prediction via Discovering Multi-Frequency Trading Patterns. In: Matwin, S., Yu, S., Farooq, F.: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17), pp.2141-2149, New York, NY, USA (2017). doi: 10.1145/3097983.3098117
2. Burkhardt, D., Nazemi, K., Kuijper A., Ginters E.: A mobile visual analytics approach for instant Trend Analysis in mobile Contexts. In Proceedings of the 5th International Conference of the Virtual and Augmented Reality in Education (VARE2019), pp. 11–19, I3M (2019). doi: 10.5281/zenodo.3473040
3. Hwang, M., In, M., Ha, S., Lee, K.: TASIS: Trend analysis system for international standards. ITU Kaleidoscope: Challenges for a DataDriven Society (ITU K), pp. 1-8, IEEE (2017). doi: 10.23919/ITU-WT.2017.8247002.
4. Kontostathis, A., Galitsky, L. M., Pottenger, W. M., Roy, S., Phelps, D. J.: A survey of emerging trend detection in textual data mining. In Berry, M. W.: Survey of Text Mining, pp. 185-224. Springer, New York, USA (2004). doi: 10.1007/978-1-4757-4305-0\_9
5. Sharma, S., Swayne, S. A., Obimbo, C.: Trend analysis and change point techniques: a survey. Energy, Ecology and Environment, pp. 123-130 (2016). doi: 10.1007/s40974-016-0011-1
6. Chandler, R. E., Scott, M.: Statistical methods for trend detection and analysis in the environmental sciences. John Wiley & Sons. Chichester, UK (2011). doi:10.1002/9781119991571
7. Gray, K. L.: Comparison of Trend Detection Methods. PhD thesis, Department of Mathematics, University of Montana, USA (2007).
8. Gaubinger, K., Rabl, M., Swan, S., Werani, T.: Innovation and Product Management. Springer, Heidelberg (2015). doi: 10.1007/978-3-642-54376-0
9. Keim D., Bak P., Bertini E, Oelke D., Spretke D., Ziegler H.: Advanced Visual Analytics Interfaces, In Santucci, G.: Proceedings of the International Conference on Advanced Visual Interfaces (AVI'10) the International Conference. Roma, Italy, pp. 3-10, ACM Press, New York, USA (2010). doi: 10.1145/1842993.1842995

10. Nazemi, K.; Retz, R.; Burkhardt, D.; Kuijper, A.; Kohlhammer, J.; Fellner, D. W.: Visual Trend Analysis with Digital Libraries. Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business. pp. 14:1–14:8, ACM, Graz, Austria, (2015). doi: 10.1145/2809563.2809569
11. Nazemi K., Burkhardt D. (2018): „Juxtaposing Visual Layouts – An Approach for Solving Analytical and Exploratory Tasks Through Arranging Visual Interfaces“. In proceedings of 4th International Conference of the Virtual and Augmented Reality in Education, I3M (2018). doi: 10.5281/zenodo.2542952
12. Nazemi, K., Burkhardt, D.: Visual Analytics for Analyzing Technological Trends from Text. In Proceedings of the 23<sup>rd</sup> International Conference Information Visualization (IV2019), Paris, France, pp. 191-200, IEEE (2019), doi: 10.1109/IV.2019.00041.
13. Shneiderman B.: The eyes have it: A task by data type taxonomy for information visualizations (1996). doi: 10.1109/VL.1996.545307
14. Blei, D. M., Ng A. Y., Jordan M. I.: Latent Dirichlet Allocation. Journal of Machine Learning Research. Journal of Machine Learning Research 3 (2003). doi: 10.1162/jmlr.2003.3.4-5.993
15. Nazemi, K., Breyer, M., Burkhardt, D., Fellner, D. W.: Visualization Cockpit: Orchestration of Multiple Visualizations for Knowledge-Exploration. In International Journal of Advanced Corporate Learning, 3 (4), pp. 26-34, ISSN: 1867-5565, International Association of Online Engineering, Vienna, Austria (2010). doi: 10.3991/ijac.v3i4.1473
16. Nazemi, K., Burkhardt, D., Breyer, M., Stab, C., Fellner, D. W.: Semantic Visualization Cockpit: Adaptable Composition of Semantics-Visualization Techniques for Knowledge-Exploration Conference. In International Conference on Interactive Collaborative Learning (ICL), Kassel University Press, Germany (2010).
17. Nazemi, K., Burkhardt, D., Kaupp, L., Dannewald, T., Kowald, M., Ginters, E.: Visual Analytics in Mobility, Transportation and Logistics. In Workshop on ICTE in Transportation and Logistics, pp. 82-89, Springer (2020). doi: 10.1007/978-3-030-39688-6\_12

# Emerging Knowledge Extraction and Visualization in Medical Document Corpora

Christian Nawroth<sup>1</sup>, Marc Herrmann<sup>1</sup>, Felix Engel<sup>2</sup>, Paul Mc Kevitt<sup>2</sup>, and Matthias Hemmje<sup>2</sup>

<sup>1</sup> Chair of Multimedia and Internet Applications, Faculty of Mathematics and Computer Science, University of Hagen  
christian.nawroth@fernuni-hagen.de marc.herrmann@studium.fernuni-hagen.de

<sup>2</sup> Research Institute for Communication and Cooperation, Dortmund  
{fengel, pmckevitt, mhemmje}@ftk.de

**Abstract.** In this paper, we demonstrate our concept of emerging Named Entities (eNEs) for the tasks of Emerging Knowledge Extraction and Visualization in medical document corpora. We derive four use cases that utilize eNEs in medical document corpora to support medical expert users accessing emerging knowledge. We design the visual Emerging Named Entity Recognition and Information Retrieval System (visual eNER-IRS), supporting three of these use cases. We demonstrate proof-of-concept emerging knowledge visualizations for the different use cases. Finally, we present a detailed user evaluation of our visualization approach. The evaluation concludes that our approach helps users utilize eNEs on a corpus and a single document level. Overall, this paper demonstrates the benefits of our approach for the related project RecomRatio by providing recent and emerging knowledge for evidence-based medical use cases. Hence, the main contribution is a visualization of new medical concepts and emerging knowledge in literature for medical experts for supporting medical information, decision-making, and reporting in medical research and treatment.

**Keywords:** Emerging Named Entity Recognition (eNER) · Emerging Knowledge Visualiazation, emerging Named Entities (eNEs), Emerging Named Entity Recognition and Information Retrieval System (visual eNER-IRS)

## 1 Introduction and Motivation

This article is based on the results of a master thesis [6] and a dissertation project at the Chair of Multimedia and Internet Applications at the University of Hagen. Both works are related to the project *Recommendation Rationalisation* (RecomRatio [2]). RecomRatio is a DFG funded research project that aims to support expert health professionals during informed decision-making processes (e.g., for or against a certain diagnosis/therapy) by providing evidence through textual arguments found in the medical literature and documents. Within RecomRatio, we intend to make emerging Named Entities (eNEs) [12] and emerging

Argument Entities available for Information Retrieval (IR) in medical document corpora supporting medical argumentation engineering. Following our previous work on this topic [9][12], medical eNEs are names for medical entities (e.g., for diseases, drugs) that are in use in a medical document corpus (e.g., PubMed / MEDLINE). Yet, they are not formally acknowledged through the expert community, i.e., by adding them to a medical vocabulary. In addition, to support the underlying medical argumentation use cases within ReconRatio we defined an emerging Argument Entity (eAE) as an argument that contains an eNE in its premise or the conclusion element [14]. We argue that eNEs usually represent the most recent knowledge in a domain. Hence, emerging Named Entity Recognition (eNER) aims to recognize them in the document corpus and make them available for medical IR use cases. To recognize eNEs, we propose a hybrid approach combining textual Natural Language Processing (NLP) with Machine Learning (ML) techniques on temporal features [12]. Whilst our previous publications focused on the recognition of eNEs, here we explain why and how we provide eNEs to the user through visual interfaces that support four use cases for medical expert users. For our tasks presented here, we use two corpora: PubMed MEDLINE Baseline 2020<sup>3</sup> (MEDLINE) and PubMed Open Access (PMC OA) Subset<sup>4</sup>. Whilst the former generally only consists of the title and abstract, the latter also contains the full texts, so we decided to use both for our Document Engineering project. Between 1970 and 2019, the number of citations added to MEDLINE grew from 219.337 entries per year to 1.406.789, based on our corpus index statistic derived from our experimental corpora. So the yearly growth rate increased by a factor 6.4 within 50 years. Furthermore, we outline how our work links knowledge from these two corpora to the ClinicalTrials<sup>5</sup> document corpus.

## 2 State-of-the-Art and Related Work

Our work is related to the task of realtime Emerging Topic Detection in Microblogs as presented by Chen et al. [4], which also utilizes ML techniques on non-textual features to detect emerging topics within microblogs. Our approach differs as it does not focus on realtime detection, but long-term eNEs in a scientific text corpus and therefore, it uses different non-textual temporal features compared to Chen et al. [4]. Furthermore, our approach for eNER aims to recognize eNE in scientific corpora and hence combine Document Engineering techniques from traditional NER and ML. In recent work, Wang et al. [17] apply hot topic detection to the field of academic big data, which they call Academic Hot Topic Detection. Like our approach, they combine a textual NER approach in the first stage with a feature learning approach. Their main features are a co-occurrence graph and word embeddings amongst additional document related features. In contrast, we focus on eNEs in a solely temporal way, not yet analyzing whether these topics are “hot”, i.e. setting a trend of popular information

<sup>3</sup> [https://www.nlm.nih.gov/databases/download/pubmed\\_medline.html](https://www.nlm.nih.gov/databases/download/pubmed_medline.html)

<sup>4</sup> <https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/>

<sup>5</sup> <https://clinicaltrials.gov/>

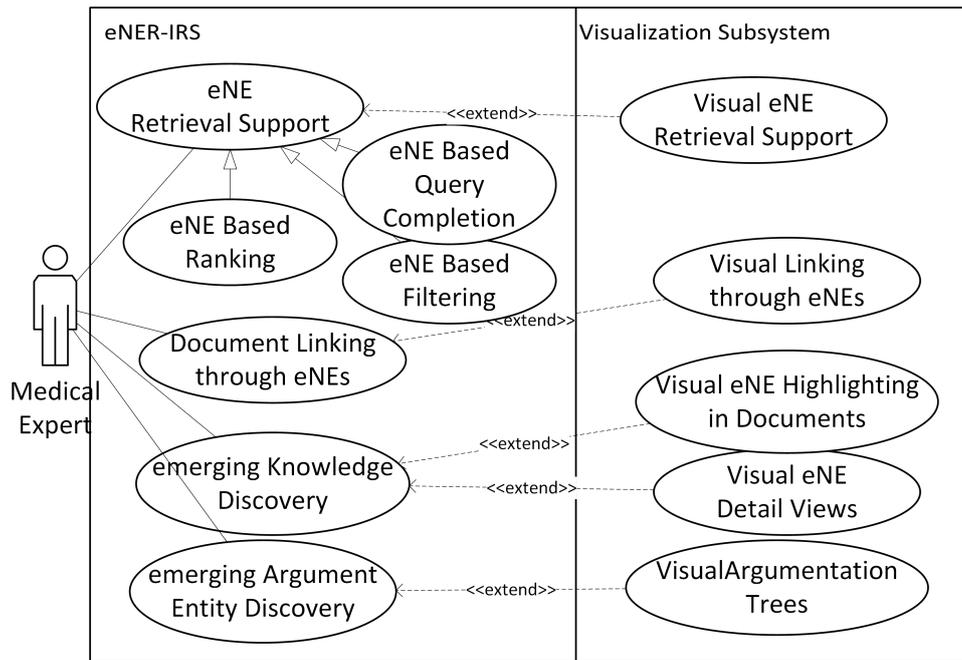
need/interest. The design of the visualization subsystem generally follows the IVIS4BigData Framework introduced by Bornschlegl et al. [3]. The framework describes a method to transform raw data from big data sources into “advanced visual user interfaces for Big Data Analysis” to allow “efficient and effective” Human-Computer Interaction (HCI). A major component of the IVIS4BigData is a pipeline that consists of different steps to provide data insight and effectuation based on raw data. These steps are Raw Data Collection, Data Structures, Visual Structures, and Views. Between Data Collection and Data Structures, the framework includes an analytics layer that comprises an analytical component of the underlying big data use cases.

### 3 Visual eNER-IRS System Design

To design the visual Emerging Named Entity Recognition and Information Retrieval System (visual eNER-IRS), we apply a user-centered design approach [13]. Therefore, we introduce the four use cases of the visual eNER-IRS, give a brief overview of the general system architecture, and derive the architecture of the visualization subsystem that will become the basis of the later Argument Visualization System. Here, we focus on the visualization subsystem and explain the underlying architecture briefly to ensure general understanding. A detailed description of the eNER pipeline is given in [12, 11].

#### 3.1 Use Cases of the Visual eNER-IRS

The visual eNER-IRS is intended to support four different information retrieval use cases [11]. These are eNE retrieval support, document linking through NEs, emerging Knowledge Discovery, and (later) emerging Argument Entity discovery, as shown in Fig. 1. These four use cases are supported by one or two generic visual use cases provided through the visualization subsystem. In the following, we briefly introduce the four general use cases summarizing [11]. The first, eNE Retrieval Support (see Fig. 1) aims at providing functionality that utilizes eNEs to enhance and support several standard retrieval methods, like query completion, filtering, faceted search, and boosting of ranking results depending on eNEs. The associated visual use case is visual eNE Retrieval Support. This visualization use case is intended to highlight eNEs during several steps of user interaction with the retrieval system. The second general use case supported by one visualization use case is document linking through eNEs. In this use case, eNEs are utilized to provide a link between documents from different corpora. For example, a user finds a clinical trial in the ClinicalTrials (CT) Corpus that contains eNEs that represent new medical knowledge in the respective clinical trial. Then these eNEs can be used to search for documents in another text corpus, e.g., MEDLINE, to retrieve new and emerging knowledge from that text corpus too. The associated visual use case *Visual Linking through eNEs* is intended to provide an interactive graphical representation of that use case, i.e., a network graph showing links between documents from different corpora based



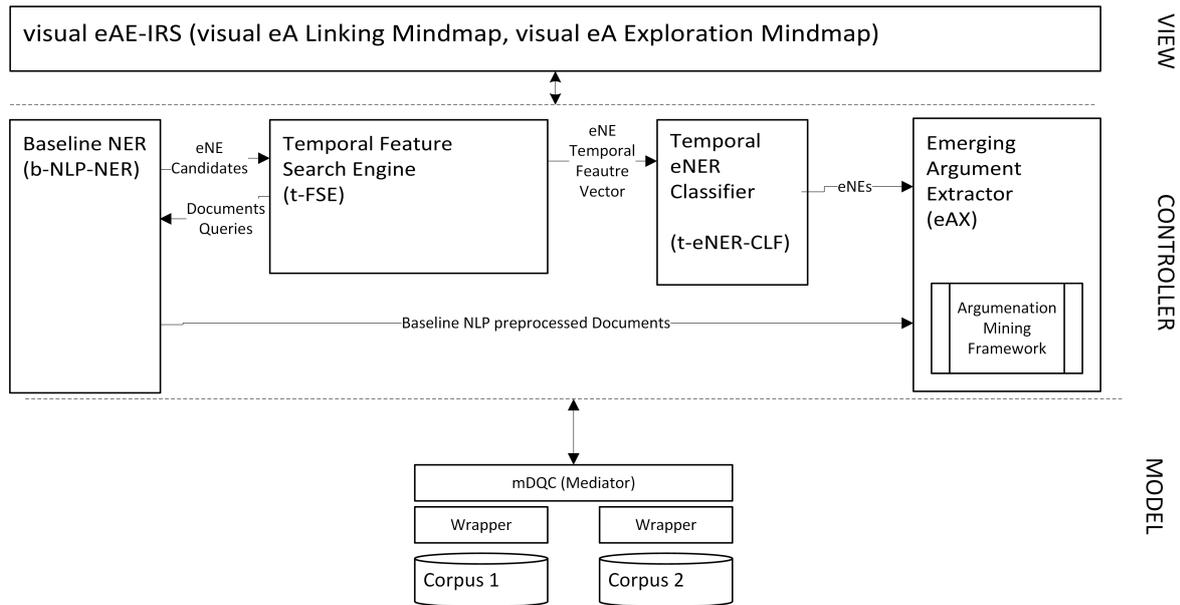
**Fig. 1.** UML eNER-IRS Use Case Model [11]

on eNEs. The third general use case is visual emerging Knowledge Discovery. This use case has an exploratory characteristic and enables users to explore new knowledge on document level and the single emerging entity level. The associated visual use cases (visual eNE highlighting in documents, visual eNE detailed views) provide views for both exploratory aspects, which means visual highlighting of eNEs in selected papers and providing detailed information on selected eNEs based on the textual and temporal analysis of the eNER-IRS. The fourth use case is emerging Argument Entity Discovery. Based on emerging Argument Entities (e.g., from a survey article) in arguments’ premises or conclusions, the expert medical users can retrieve, link, and visualize arguments that cover the most recent medical knowledge. This use case is not covered by this paper but published in [10].

### 3.2 General Conceptual Architecture

Following the motivation and the three initial use cases, our architectural modelling approach (see Fig. 2) for recognizing eNEs in a medical document and query corpus combines methods from NLP, NER, IR, and ML [12, 11]. Our approach follows the Model View Controller (MVC) paradigm [7].

Here, we focus on the conceptual design of the view layer that contains the visualization components of the eNER-IRS. As the View layer interacts with the controller layer, we introduce the controller layer for a general understanding. A more detailed description and evaluation of the underlying eNER pipeline in the controller layer is published in [12, 11]. The core components of the controller layer that are referenced in the View layer are the medical document corpus, the baseline NLP and NER, the temporal features search engine, and



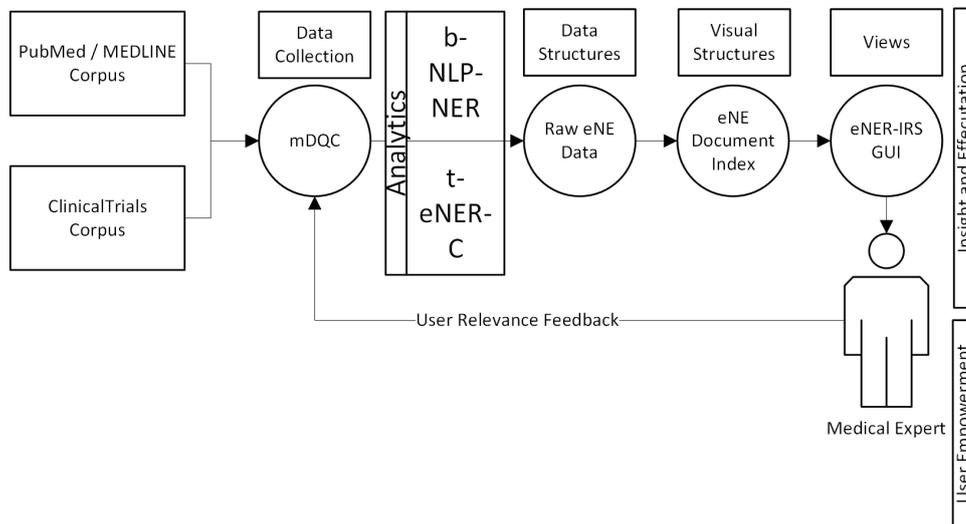
**Fig. 2.** Conceptual System Architecture adapted from [11]

the temporal eNER Classifier. The medical document corpus is the indexed document corpus containing all medical documents within the system. The baseline NLP and NER provides the extraction of eNER candidates based on textual features. The temporal features search engine extracts temporal features of the eNE-candidates from the medical document corpus (e.g., year of first use). The temporal eNER Classifier is the ML component that finally classifies eNE candidates based on these temporal features. Based on the extracted eNEs and the baseline preprocessed documents, the emerging Argumentation Extractor aims at identifying eAEs in arguments. It will rely on a state-of-the-art argumentation mining framework, e.g., ArgumenText [16]. Following this overview of the core eNER components in the controller layer in the next subsection, we introduce the emerging knowledge visualization subsystem’s conceptual design.

### 3.3 Conceptual Design of the Visual Subsystem

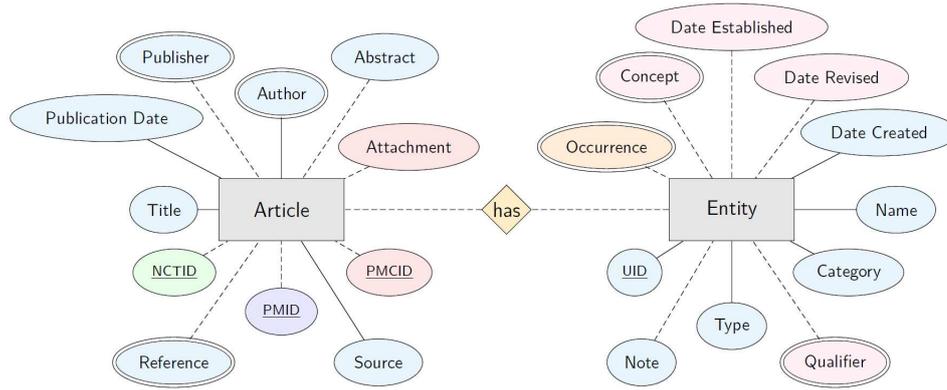
The conceptual design of the visualization subsystem has to consider challenges posed by the big data characteristics of the underlying document corpora and vocabularies. Hence, to address these challenges for the visualization subsystem’s conceptual design, we applied the IVIS4BigData Framework. In general, IVIS4BigData aims to make big data resources available and beneficial for users through visualization. Fig. 3 shows how we use the IVIS4BigData Framework to transform raw textual data from medical corpora into views that allow medical expert users’ visual data insight and emerging knowledge effectuation. Compared to the full IVIS4BigData Framework in our work, the pipeline part and a feedback channel (user empowerment) are implemented. Furthermore, we focus on the end-user’s view, but we do not implement the views on the first three components of the full IVIS4BigData pipeline. Textual Big Data Sources for our

system are the two corpora PubMed and MEDLINE, as introduced above. The raw textual data is collected in the medical document corpus from the system architecture design (see Fig. 2). In the IVIS4BigData, the medical document corpus refers to the “Data Collection” component. The analytics layer of IVIS4BigData in our system design is represented by the two components that perform analytic tasks (eNER) on the raw data and turn the raw data into data structures: The baseline NLP and NER and the temporal eNER classifier in the controller layer as described above. The following IVIS4BigData component “Data Structures” in our work is represented by a JSON structure that encodes the mapping between a corpus document and the automatically recognized eNEs. These mappings are then transformed into a visual structure that is stored in a search engine index whose Entity-Relationship Model (ERM) is shown in Fig. 4. The *Article* entity has several attributes derived directly from the underlying corpus metadata, such as Author, Abstract, or its corpus ID (e.g., for PUBMED / MEDLINE: PMID). In contrast, *Entity* itself and its attributes are not taken from metadata but extracted through the eNER-IRS. For example, those attributes contain the Date Created (time of first use in the corpus), the name, and a possible category of it. Hence, more generally speaking, the mapping introduced above maps already existing corpus knowledge to new knowledge extracted by the eNER-IRS. The final component of the IVIS4BigData Framework is the eNER-IRS-GUI



**Fig. 3.** Visual eNER based on IVIS4BigData [3].

that we discuss in the remainder of this paper. We present a prototypical proof-of-concept implementation following the three initial use cases (eNE retrieval support, document linking through NEs, emerging Knowledge Discovery) in the following section.



**Fig. 4.** Entity Relationship Model of Visual Structure [6]

## 4 Proof-of-Concept Implementation

This section describes the prototypical proof-of-concept implementation of the visual eNER-IRS use cases based on the IVIS4BigData framework using several methods and packages. First, we give a brief overview of the technical implementation followed by descriptions of each of the visual eNER-IRS use cases followed by a description of the graphical user interface (GUI) aspects.

### 4.1 Technical Implementation

In the following, we describe the prototypical technical implementation of the previously introduced visual eNER-IRS use cases based on eNEs. The underlying concept of the Visualization Subsystem (see Fig. 1) can be considered as two independent software systems based on the IVIS4BigData Framework as follows: To transform data structures into visual structures (see Fig 4) as a first step, we developed a batch application based on the Spring Batch Java-Framework. It converts all the eNER-IRS output JSON files, including data about eNEs such as its name and occurrences in medical documents, into processed and consolidated XML files. These files are then indexed by Apache Solr to create the eNEs visual data structures (IVIS4BigData: Visual Mappings). These files are also used in another batch processing step that reads the raw data of the different medical corpus (PMC, ClinicalTrials), extracts relevant data attributes, and enriches the data by adding information about entities such as MeSH concepts and eNEs. Also, these XML output files are indexed by Apache Solr, and as a result, two visual data structures for medical documents and eNEs are created. The second system is a client-server architecture software based on the Spring Web Model-View-Controller (MVC) Java-Framework. It generally reads the visual structure data from Apache Solr and displays it on different web pages to provide the described visual eNER-IRS uses cases (IVIS4BigData: View Transformations). The web page design and functionality is based on frameworks such as Boot-

strap<sup>6</sup> and jQuery<sup>7</sup>. Additionally, for the specific use case of visual linking and highlighting of eNEs, we rely on the JS-Library D3.js<sup>8</sup> for diagram and PDF.js<sup>9</sup> for document visualization. To transport data from server to client, multiple REST API endpoints are available and tailored for the specific visual eNER-IRS use case. In the following, the prototypical implementations of the use cases are explained.

## 4.2 Visual eNE Retrieval Support

The document search enables users to browse through the different document collections (PubMed, ClinicalTrials) easily with two filter categories (*eNEs* and *Medical Subject Headings*): The main search field and the result of the document search are displayed in the right view area. Above, next to the search field, there is an additional button to control on which document collection the search should be performed. Below, the matching documents, including title, publication date, unique identifiers, entity categories, and source document collection, are listed separately. The detailed view of a paper shows the authors, the assigned entities. If available, the abstract in which the assigned entities are highlighted, is shown. The left sidebar contains all necessary controls to conveniently browse the dataset of *Emerging Named Entities* and *Medical Subject Headings*. As a result, the documents are filtered based on the selected entities. Furthermore, the sorting of the search result can be influenced by the green Learning To Rank (LTR) button next to the main search field. Generally speaking, the documents are sorted by an individual score in descending order. This score is a measurement for how relevant each document is for the given user query. The gray shaded numerical value reflects the default sorting, whereas the green shaded value also considers the information about eNEs. In detail, this score increases based on the number of assigned eNEs and on whether the user query also matches these entities. To improve user search experience, the main search field is extended by an auto-completion functionality (see Fig. 5) that lists query suggestions from multiple datasets based on the current user input. Depending on the selected document collections, the first suggestions are made by matching document titles and eNEs. Also, further suggestions derive from the entity categories *eNEs* and *Medical Subject Headings* (see Fig. 6). Selecting an entity suggestion results in automatically adding the entity as a filter criterion in the left sidebar.

## 4.3 Visual Linking through eNEs

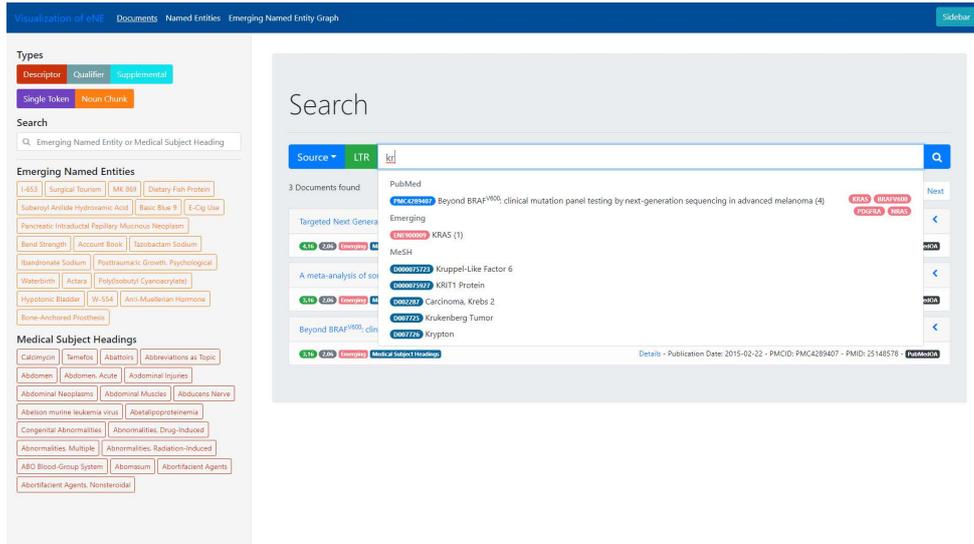
The relationships between entities are created once multiple entities appear in a single document. The greater the number of documents in which two related entities appear, the closer their relationship is. Such relationships can be researched interactively with the help of the network graph accessible under the

<sup>6</sup> <https://getbootstrap.com/>

<sup>7</sup> <https://jquery.com/>

<sup>8</sup> <https://d3js.org/>

<sup>9</sup> <https://mozilla.github.io/pdf.js/getting-started/>

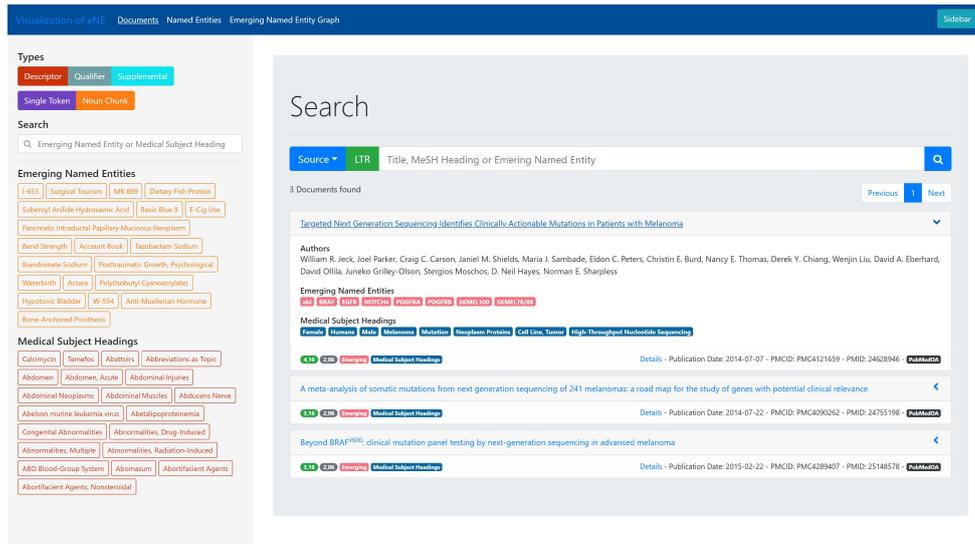


**Fig. 5.** Visual eNER-IRS: Autocompletion for User Query [6].

navigation item *Emerging Named Entity Graph* (see Fig. 7). In the left sidebar, all entities with at least one relationship are displayed. The list can be filtered regarding type, reference, and name. Additionally, the list automatically updates once an entity is selected by only showing entities in a direct relationship. The network graph itself is shown in the right view area and refreshes automatically once the entities' selection is updated. A node represents each chosen eNE, and its size depends on the total number of documents from the different collections it is assigned to. The links between nodes display the connecting documents, and their amount is represented by edge width. By clicking on the link details of the relationship are revealed.

#### 4.4 Visual eNE Highlighting in Documents

The detailed view of a document contains all mandatory attributes such as title, unique identifiers, document collection and also optional attributes such as authors, assigned entities (eNEs and non-eNEs (MeSH)) and the abstract or original PDF-document. The availability of the optional attributes depends on the collection source of the document. For example, only for documents from the PubMed Central Open Access (PMCOA) collection, the original document in PDF-format can be displayed. In the left sidebar, the assigned eNEs and non-eNEs are listed as interactive buttons. Also, these entities are highlighted in the continuous text of the abstract or the PDF-document, if available (see Fig. 8). In case of viewing a PDF-document, additional buttons to page backwards and forwards and download the document are displayed. For eNEs, the interactive button can be expanded to a drop-down list showing all the document pages on which the term appears.



The screenshot shows the 'Visualization of eNE' interface. On the left, there is a sidebar with several sections: 'Types' (Descriptor, Qualifier, Supplemental), 'Search' (with a search bar and filter 'Emerging Named Entity or Medical Subject Heading'), 'Emerging Named Entities' (a grid of entity boxes like 'I-653', 'Surgical Tourism', etc.), and 'Medical Subject Headings' (a grid of heading boxes like 'Calcimycin', 'Tetanus', etc.). The main area is titled 'Search' and shows a search bar with 'Source' and 'LTR' dropdowns, and a search query 'Title, MeSH Heading or Emerging Named Entity'. Below the search bar, it indicates '3 Documents found' and shows a list of results. The first result is 'Targeted Next Generation Sequencing Identifies Clinically Actionable Mutations in Patients with Melanoma' with a dropdown arrow. Below this, there are sections for 'Authors' (listing names like William R. Jeck, Joel Parker, etc.), 'Emerging Named Entities' (with buttons for I653, EGFR, etc.), and 'Medical Subject Headings' (with buttons for Melanoma, Mutation, etc.). Each result entry includes a 'Details' link, a publication date, and PMCID/PMID numbers.

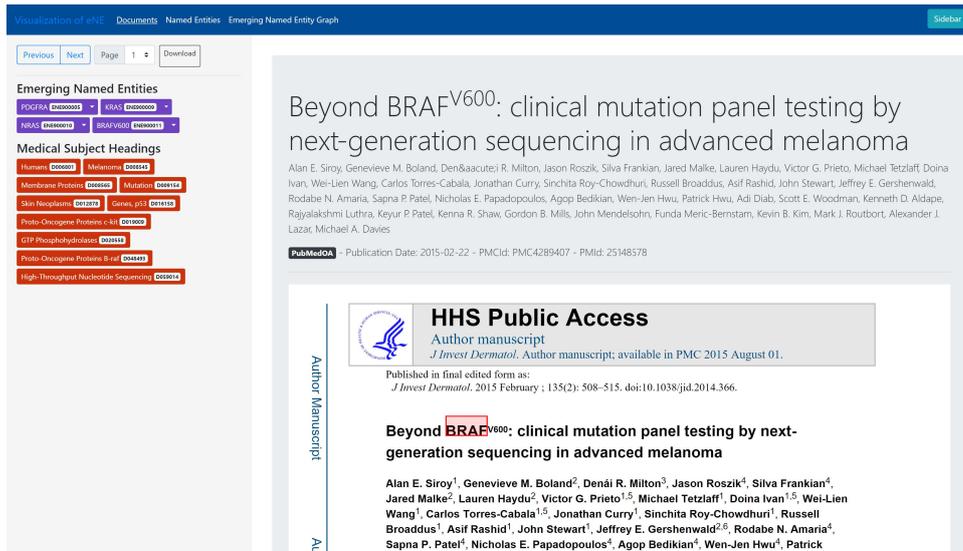
**Fig. 6.** Visual eNER-IRS: Search for Medical Documents [6].

#### 4.5 Visual eNE Detailed Views

This visual eNER-IRS use case provides the highest interaction between expert medical users and the eNER-IRS processes. Hence, the user can acknowledge or reject an eNE candidate suggested by the eNER-IRS. The requested feedback (and the respective interface) is intentionally binary (ACKNOWLEDGE / REJECT), keeping in mind that expert medical users may lack data science knowledge to give a more differentiated assessment. However, for those expert medical users with data science / ML skills, the visualization provides two metadata parameters from the eNER-IRS for their decision process. The data types of the result set of the ML process for this visual eNER-IRS use case are terms/tokens representing eNEs and temporal (statistical) metadata derived from the big data analysis of the temporal feature search engine. Besides the temporal metadata, the classification threshold from the underlying eNER-IRS component is displayed (see Fig. 9).

The detailed view of an entity contains all mandatory attributes. These are unique identifiers, category, type, name, and also optional attributes such as references to documents from different collections, overall frequency, and multiple dates related to the creation, revision, and establishment (See Fig. 10). The buttons in the left sidebar grouped by document collection reflect the related entities. An additional bar chart showing the frequency of occurrence on an annual basis is displayed depending on data availability. Two drop-down lists can change the plotted range of years on the right side above the diagram. Additionally, related entities' occurrence data can be added interactively by the buttons shown on the right-hand side.





The screenshot displays a web interface for document visualization. On the left, a sidebar titled 'Emerging Named Entities' lists various categories with counts, such as 'PSGIRA' (2000000), 'KRAS' (2000000), 'NRAS' (2000000), and 'BRAFV600' (2000000). Below this, 'Medical Subject Headings' are listed, including 'Human' (200000), 'Molecular' (200000), 'Membrane Protein' (200000), 'Mutation' (200000), 'Skin Neoplasms' (200000), 'Cancer, p53' (200000), 'Proto-Oncogene Proteins c-kit' (200000), 'GTP Phosphohydrolases' (200000), 'Proto-Oncogene Proteins B-raf' (200000), and 'High-Throughput Nucleotide Sequencing' (200000). The main content area shows a document titled 'Beyond BRAF<sup>V600</sup>: clinical mutation panel testing by next-generation sequencing in advanced melanoma'. The document text includes the title, authors (Alan E. Siroy, Genevieve M. Boland, Deni R. Milton, Jason Roszik, Silva Frankian, Jared Malke, Lauren Haydu, Victor G. Prieto, Michael Tetzlaff, Doina Ivan, Wei-Lien Wang, Carlos Torres-Cabala, Jonathan Curry, Sinchita Roy-Chowdhuri, Russell Broadus, Asif Rashid, John Stewart, Jeffrey E. Gershenwald, Rodabe N. Amaria, Sapna P. Patel, Nicholas E. Papadopoulos, Agop Bedikian, Wen-Jen Hwu, Patrick Hwu, Adi Diab, Scott E. Woodman, Kenneth D. Aldape, Rajyalakshmi Luthra, Keyur P. Patel, Kenna R. Shaw, Gordon B. Mills, John Mendelsohn, Funda Meric-Bernstam, Kevin B. Kim, Mark J. Routbort, Alexander J. Lazar, Michael A. Davies), and a PubMed ID. A 'HHS Public Access' banner is visible, indicating the author manuscript is available in PMC 2015 August 01. The document title is highlighted in red in the original image.

**Fig. 8.** Visual eNE Highlighting In Documents [6] (Article: [15]).

For each of the scenarios, we defined 2-5 multiple choice test questions (TQs). For each TQ, one or more answer options are correct. Users' answers with all correct answer options are classified as *correct*, with some correct answer options as *partly\_correct* and with no correct answer options as *wrong*. Fig. 11 shows the evaluation for the test questions (overall results and results per question). Turns out that there is a share of  $\geq \frac{2}{3}$  of correct or partially correct answers for all three scenarios. This finding concludes that, in general, users were able to fulfill the three test scenarios defined above. However, within the questions of the particular scenarios, a variance regarding the outcome can be observed. In scenario (1), the TQ04, and scenario (2), the TQ06 has an outcome of correct answers of  $< 0.5$ . TQ04 deals with the highlighting of single eNEs in a document, TQ06 is about ranking statistics. This finding concludes that only the visualization details must be improved while the overall system is usable and performs well. Users were able

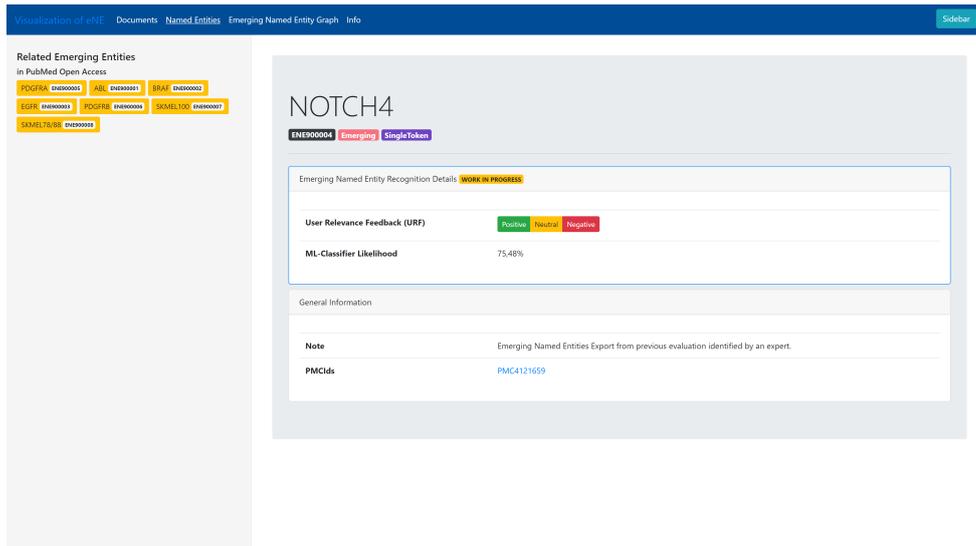
### 5.3 Usability Metrics

The goal of this step is to assess the perceived usability of the user based on ISO 9241-11. Therefore, the Usability Metric for User Experience (UMUX) with its four-item Likert scale is used [5]. with the following questions:

- Q01 The system's capabilities meet my requirements.
- Q02 Using [this system] is a frustrating experience.
- Q03 [This system] is easy to use.
- Q04 I have to spend too much time correcting things with [this system].

First, Fig. 12 shows the results of Q01 and Q03.

It turns out that a more than 0.5 of the participants gave a positive or neutral assessment for both questions. However, for Q03, turns out that a significant



**Fig. 9.** Acknowledgement of eNEs [6].

portion of participants do not think that the system is easy to use. We argue that this is not surprising but emphasize that the system is an expert system that may require more extensive training to be used beneficially. Secondly, Fig. 13 shows the results of Q02 and Q04, now with a reversed colour scale compared to Fig. 12 as explained before.

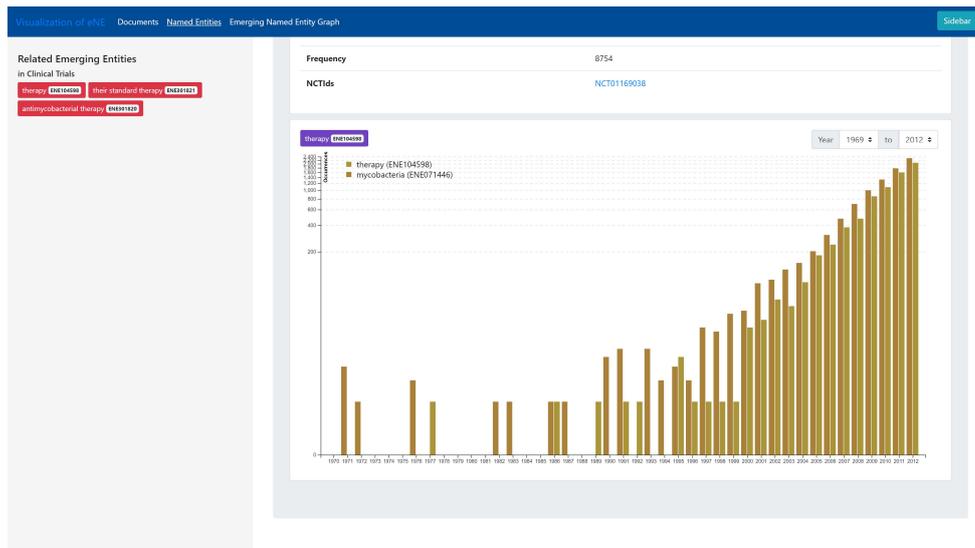
Again, turns out that more than half of the participants have a positive or neutral assessment regarding Q02 and Q04.

Thirdly, we plot the mean results and the standard deviations for all questions (See Fig. 14). The plots of the mean and standard deviation show reflect the results introduced earlier. For Questions Q01 and Q03, the mean is greater or equal to the *neutral* assessment, while for Q02 and Q04, it is below. We argue that the relatively strong standard deviations result from the heterogeneous participant group evaluating our system, with different experiences in the medical domain, and using expert retrieval systems.

Overall the usability evaluation showed a positive outcome, leading to the conclusion that the system, in general, has reliable usability whilst there are again improvements in visualization details. Furthermore, it shows the need for sufficient training on the system for users inexperienced in using expert systems or who are new to the medical domain.

## 5.4 Added Value in Professional Terms

The following questions are intended to assess the added value in professional terms related to certain areas in the prototypical application. In contrast to UMUX, here the 5-point Likert scale is used to express how much the tester agrees (5) or disagrees (1) with a particular statement [1], (Neutral: (3), N/A: (0)):



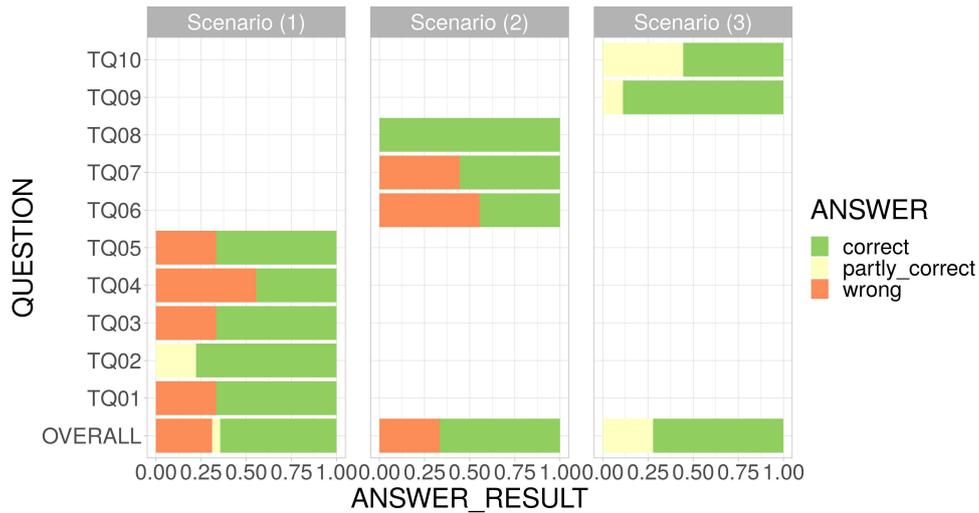
**Fig. 10.** Visual eNE detailed views [6]

- Q05 The possibility to filter medical documents with the help of eNEs can generate a true added value during the research process.
- Q06 The visually highlighted eNEs in continuous text and original medical documents support the assessment in terms of their quality and professional relevance.
- Q07 The possibility to download the original medical document with visually highlighted eNEs allows distributing the information with expert colleagues simply.
- Q08 The interactive bar chart is a valuable visualization to display the occurrences of eNEs.
- Q09 The interactive network graph is a valuable visualization to display the relationships between eNEs.
- Q10 The visualized relationships between eNEs support the assessment in terms of their quality and professional relevance.

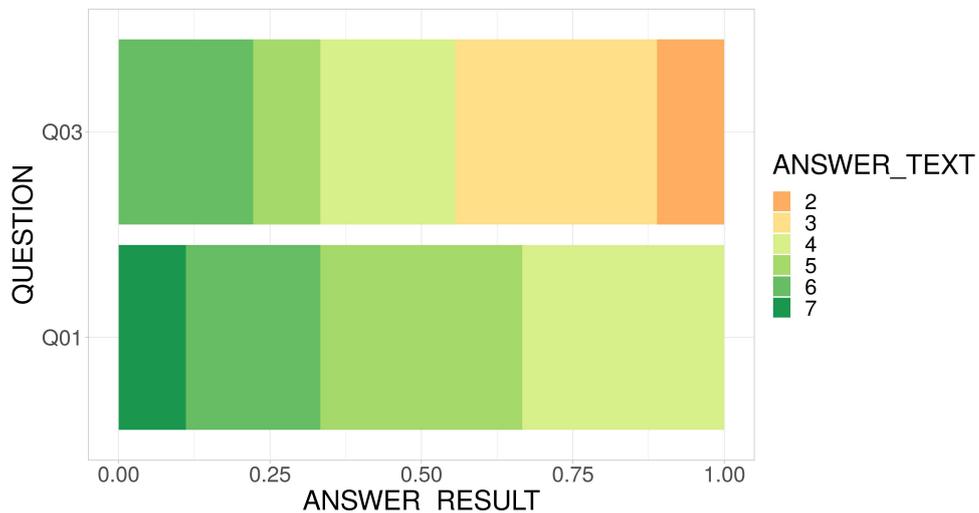
Fig. 15 envisions the results of the questions above.

Here, turns out that again all questions have a *neutral* or better outcome in more than half of the answers given. Except for Q08, more than 50% have a *better than neutral* outcome. The high ratings for Q05, Q06, and Q09 are promising. These questions reflect the core of our work and our use case. They show that our concept of eNE and its utilization within information retrieval use cases and their visualization are seen as beneficial by most participants. In contrast, the questions Q07 and Q08 focus on detail visual implementations. They emphasize that there is room for improvement regarding the aspects of the visualization. Again, we plotted the mean results per question, including the standard deviation (see Fig. 16).

For all questions, it shows mean values significantly above the neutral value of (3). Question seven has the strongest standard deviation, while the other standard deviations are more moderate. That again reflects a strong variance among



**Fig. 11.** Test Questions' Results

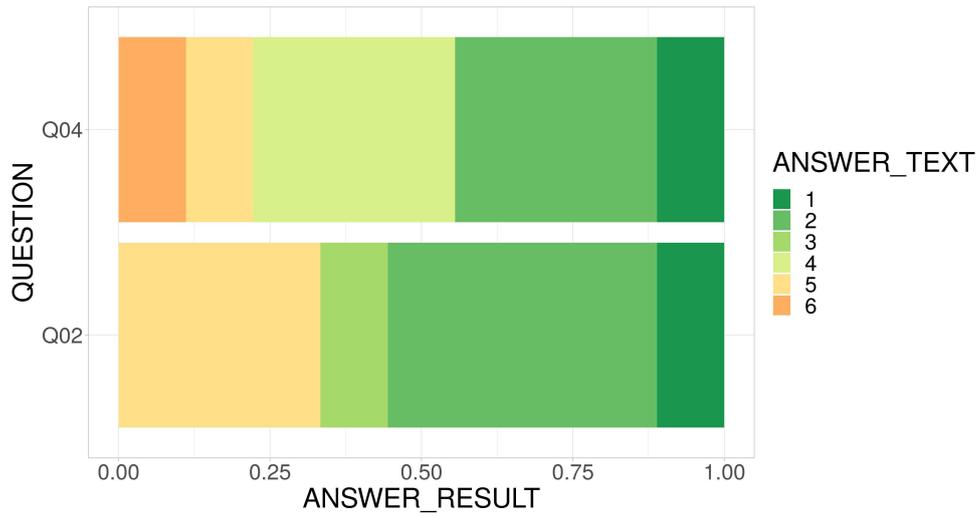


**Fig. 12.** Usability Assessment I

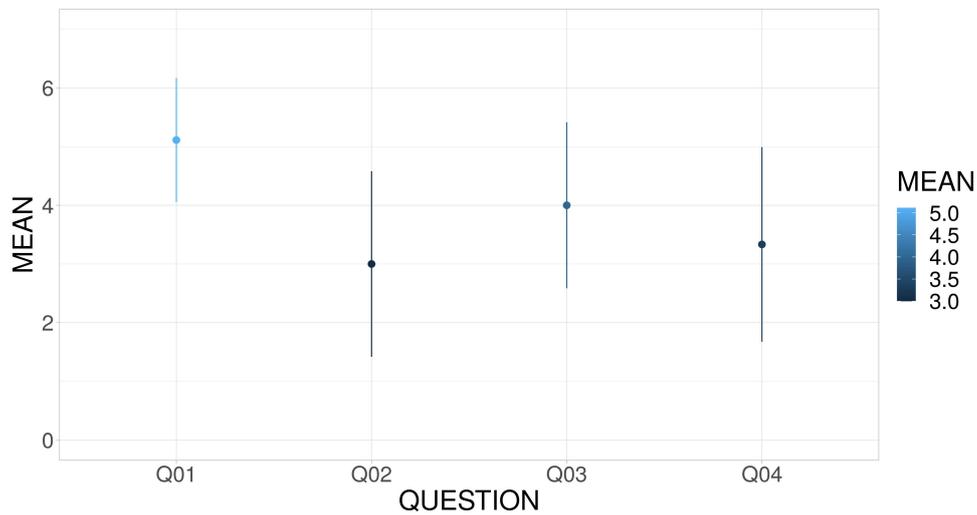
the participants when it comes to using a detailed implementation feature. The high means for questions Q05, Q06, and Q09 demonstrate that our concept of eNEs and their visualization is useful and beneficial amongst the participants.

## 6 Conclusion and Discussion

In this paper, we outlined a complete workflow utilizing and visualizing our concept of emerging knowledge represented by eNEs. We introduced four visual use cases to utilize eNEs by medical experts in document corpora. We derived a system design for recognizing and visualizing them to support medical retrieval and argumentation use cases. We designed a visualization subsystem on the IVIS4BigData Framework, visualizing eNEs in documents and document corpora and demonstrated a proof-of-concept implementation. Our evaluation

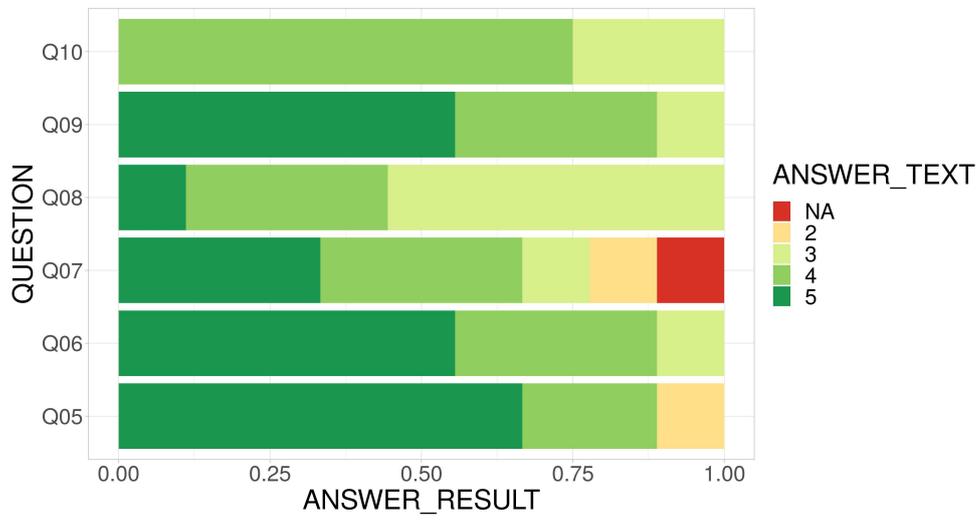


**Fig. 13.** Usability Assessment II

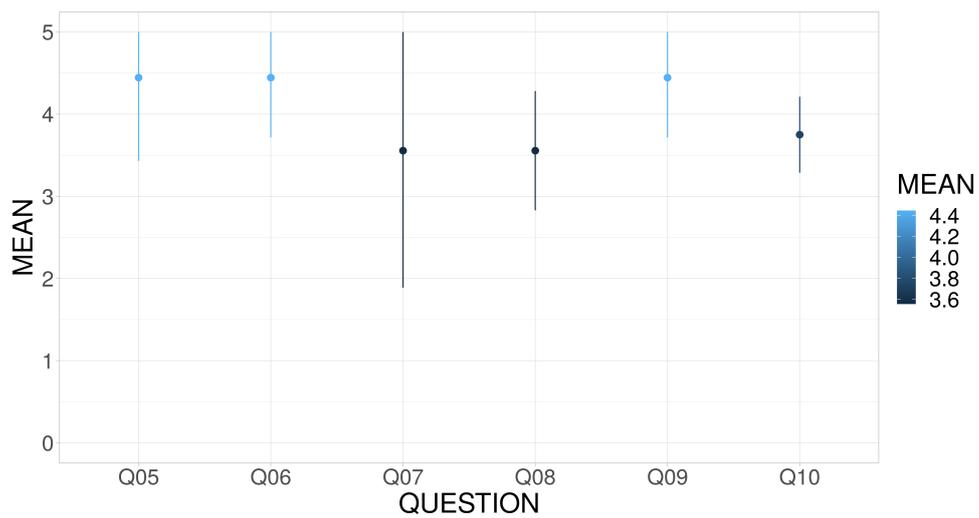


**Fig. 14.** Overall Usability Assessment (Means)

indicated that our three use cases and their four visualization could help expert medical users searching for document-based medical evidence. Furthermore, we identified items for improving our concept in visual details. We showed the necessity for sufficient user training with the complex task of eNER in a medical document corpus. Concerning future work, we outlined the emerging Argumentation Extraction system.



**Fig. 15.** Added Value in Professional Terms



**Fig. 16.** Added Value in Professional Terms (Means)

## References

1. 5-Point Likert Scale. In: Preedy, V.R., Watson, R.R. (eds.) *Handbook of Disease Burdens and Quality of Life Measures*, pp. 4288–4288. Springer, New York, NY (2010). [https://doi.org/10.1007/978-0-387-78665-0\\_6363](https://doi.org/10.1007/978-0-387-78665-0_6363), [https://doi.org/10.1007/978-0-387-78665-0\\_6363](https://doi.org/10.1007/978-0-387-78665-0_6363)
2. of Bielefeld, U.: *Rationalizing Recommendations (RecomRatio): Project-Homepage*. Bielefeld (2017), <http://www.spp-ratio.de/de/projekte/ratiorec/>
3. Bornschlegl, M.X., Berwind, K., Kaufmann, M., Engel, F.C., Walsh, P., Hemmje, M.L., Riestra, R.: *IVIS4BigData: A Reference Model for Advanced Visual Interfaces Supporting Big Data Analysis in Virtual Research Environments*. In: Bornschlegl, M.X., Engel, F.C., Bond, R., Hemmje, M.L. (eds.) *Advanced Visual Interfaces. Supporting Big Data Applications*. pp. 1–18. *Lecture Notes in Computer Science*, Springer International Publishing, Cham (2016). [https://doi.org/10.1007/978-3-319-50070-6\\_1](https://doi.org/10.1007/978-3-319-50070-6_1)
4. Chen, Y., Amiri, H., Li, Z., Chua, T.S.: *Emerging topic detection for organizations from microblogs*. In: *SIGIR 2013 - Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 43–52 (2013). <https://doi.org/10.1145/2484028.2484057>
5. Finstad, K.: *The usability metric for user experience. Interacting with Computers* **22**(5), 323–327 (2010), publisher: Oxford University Press Oxford, UK
6. Herrmann, M.: *Visualisierung von Emerging Named Entities im Rahmen von Information Retrieval in medizinischen virtuellen Forschungsumgebungen*. Masterthesis, FernUniversität in Hagen, Hagen (2020)
7. Krasner, G.E., Pope, S.T., others: *A description of the model-view-controller user interface paradigm in the smalltalk-80 system*. *Journal of object oriented programming* **1**(3), 26–49 (1988)
8. Laubheimer, P.: *Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests* (2018), <https://www.nngroup.com/articles/measuring-perceived-usability/>
9. Nawroth, C., Engel, F., Eljasik-Swoboda, T., Hemmje, M.: *Towards enabling emerging named entity recognition as a clinical information and argumentation support*. In: *DATA 2018 - Proceedings of the 7th International Conference on Data Science, Technology and Applications* (2018)
10. Nawroth, C., Duttenhöfer, A., Hemmje, M.: *Argumentationsunterstützung durch emergentes Wissen in der Medizin*. In: to appear in: Wilhelm Bauer, Joachim Warschat, *Innovation durch Natural Language Processing - Mit Künstlicher Intelligenz die Wettbewerbsfähigkeit verbessern* (2020)
11. Nawroth, C., Engel, F., Hemmje, M.: *Emerging Named Entity Recognition in a Medical Knowledge Management Ecosystem*. In: *Accepted for: Proceedings of 12th International Conference on Knowledge Engineering and Ontology Development*. Budapest (2020)
12. Nawroth, C., Engel, F., Mc Kevitt, P., Hemmje, M.L.: *Emerging Named Entity Recognition on Retrieval Features in an Affective Computing Corpus*. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. pp. 2860–2868 (Nov 2019). <https://doi.org/10.1109/BIBM47256.2019.8983247>
13. Norman, D.A., Draper, S.W.: *User centered system design: New perspectives on human-computer interaction*. CRC Press (1986)
14. Palau, R.M., Moens, M.F.: *Argumentation mining: the detection, classification and structure of arguments in text*. In: *Proceedings of the 12th international conference on artificial intelligence and law*. pp. 98–107 (2009)

15. Siroy, A.E., Boland, G.M., Milton, D.R., Roszik, J., Frankian, S., Malke, J., Haydu, L., Prieto, V.G., Tetzlaff, M., Ivan, D., Wang, W.L., Torres-Cabala, C., Curry, J., Roy-Chowdhuri, S., Broaddus, R., Rashid, A., Stewart, J., Gershenwald, J.E., Amaria, R.N., Patel, S.P., Papadopoulos, N.E., Bedikian, A., Hwu, W.J., Hwu, P., Diab, A., Woodman, S.E., Aldape, K.D., Luthra, R., Patel, K.P., Shaw, K.R., Mills, G.B., Mendelsohn, J., Meric-Bernstam, F., Kim, K.B., Routbort, M.J., Lazar, A.J., Davies, M.A.: Beyond BRAFV600: Clinical Mutation Panel Testing by Next-Generation Sequencing in Advanced Melanoma. *Journal of Investigative Dermatology* **135**(2), 508–515 (Feb 2015). <https://doi.org/10.1038/jid.2014.366>, <http://www.sciencedirect.com/science/article/pii/S0022202X15371189>
16. Stab, C., Daxenberger, J., Stahlhut, C., Miller, T., Schiller, B., Tauchmann, C., Eger, S., Gurevych, I.: ArgumenText: Searching for Arguments in Heterogeneous Sources. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. pp. 21–25. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). <https://doi.org/10.18653/v1/N18-5005>, <https://www.aclweb.org/anthology/N18-5005>
17. Wang, B., Yang, B., Shan, S., Chen, H.: Detecting Hot Topics From Academic Big Data. *IEEE Access* **7**, 185916–185927 (2019), publisher: IEEE

# A Holistic UX Methodological Framework for Measuring the Dynamic, Adaptive and Intelligent Aspects of a Software Solution

Vivien Johnston<sup>1</sup>[0000-0002-9388-7637], Michaela Black<sup>2</sup>[0000-0002-5140-2566],  
and Jonathan Wallace<sup>1</sup>[0000-0002-8415-4001]

<sup>1</sup> School of Computing, Ulster University, Jordanstown, Co. Antrim, Northern Ireland  
{[vjohnston-v7](mailto:vjohnston-v7@ulster.ac.uk), [jg.wallace](mailto:jg.wallace@ulster.ac.uk)}@ulster.ac.uk

<sup>2</sup> School of Computing, Engineering & Intelligent Systems, Ulster University, Londonderry,  
Co. Londonderry, Northern Ireland  
[mm.black@ulster.ac.uk](mailto:mm.black@ulster.ac.uk)

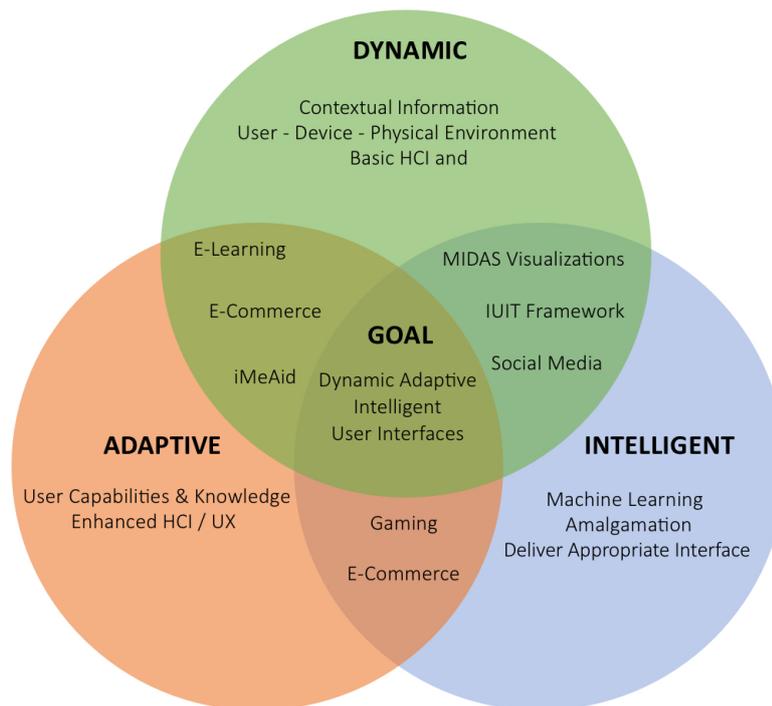
**Abstract.** The aim of this paper is to present research that proposes a methodological assessment framework based upon criteria/variables which have been grouped into three core aspects namely how Dynamic, Adaptive and Intelligent the user experience is. The framework aims to enhance the user experience of an application by analyzing these aspects and providing recommendations to a developer to produce a more enhanced Dynamic Adaptive Intelligent User Interface. Research into this field has identified that not all current applications are aware of, or capable of measuring all the aspects. The framework is based upon a three phased approach: phase one will measure the variables within each aspect and produce a score that indicates to a developer the degree to which their current application fits within each aspect; phase two highlights the areas of growth within each aspect and provide recommendations that would enhance the application's user experience; and phase three validates the framework by highlighting the application's user experience progress from previous measurements. All three phases are combined to produce a robustly proven tool that aids the developer with validated advice in order to enhance their products user experience.

**Keywords:** User Experience, Methodological Framework, Measure, System Scoring.

## 1 Introduction

Previous research includes the proposal of a framework that would be able to enhance the user experience (UX) of an application [1]. UX relates to the design, usability and functionality of an application's interface [2]. This work identified three aspects that are key to the creation of the framework, namely: Dynamic, Adaptive; and Intelligent (Figure 1) [1]. Each aspect contains parameters, a parameter is defined as being a measurable piece of information that is linked to one of the aspects of the framework [3].

The Dynamic aspect of the framework contains parameters relating to the contextual information of an end-user, their device; and their physical environment. The parameters for the Dynamic aspect include: the type of device the end-user is using; and the time of day they are accessing the application [1]. The Adaptive aspect measures existing parameters about each end-user, such as: their knowledge set, capabilities; and their goal for using the application [1]. The Intelligent aspect uses data analysis to help identify patterns and trends within data. This aspect will help enhance the UX for each cohort that an end-user would belong to by working in conjunction with the previous two aspects [1].



**Fig. 1.** The Intersections of Dynamic, Adaptive and Intelligent Aspects with Examples [1].

This paper extends research that was carried out in the work of [1]. It has been identified through previous research that not all forms of measurement have been capable of measuring all three recognized aspects however, the Hawthorne Effect has been identified during UX analysis.

The Hawthorne Effect has been identified during many forms of observational analysis by each end-user when providing feedback on the UX of an application's interface. It is the feeling of pressure whilst being observed during a task, this then leads to unusual interactions by each end-user [4]. This is a factor that the framework in question will take into consideration, it will bypass any end-user feeling pressurized. To avoid an end-user feeling pressurized when supplying feedback, it would be helpful if there was a framework that could: measure the degree to which an application fits within each aspect and illustrate this with a score, highlight the areas of growth and provide recommendations that will enhance the UX of an application; and provide validation to highlight the UX progress from previous measurements. The framework would benefit developers within multiple domains and assist with their software development process, and overall enhance the UX of their products/service solutions.

In order for this framework to be created, identification of additional parameters within each of the three aspects mentioned above is needed. A form of measurement will be highlighted as to the degree of each parameter. Once this framework is fully integrated into the developer's software development process, it will assist in producing a Dynamic Adaptive Intelligent User Interface (DAIUI).

The remainder of this paper is structured as the following: section 2 is related work, section 3 is the methodology; and section 4 contains the conclusion and future work.

## 2 Related Work

UX design can be described as being a narrative, by providing a story to the end-user via a network of events. A narrative could be portrayed as one of two approaches: task or experience. Task is in relation to a goal that an end-user may have or is carrying out when using an application. Experience, however, is in relation to the types of emotions and meaning behind the interactions from each end-user [5]. The main narrative that is important is the UX of an application, and feedback regarding this is normally provided by end-users' through a variety of evaluation methods.

Evaluation methods are categorized into three segments: self-reported, observational; and physiological measurements [6]. Self-reported measurements relate to an end-user documenting their thoughts and feelings via a survey or questionnaire, observational refers to observing an end-user whilst they interact with an application; and physiological relates to sensors attached to an end-user that monitor their physical movement in the form of quantifiable data [6]. Measurements including surveys can be time consuming and there could also be the issue of subjective bias, such as: are there a mix of quantitative and qualitative questions; and whether it is completed alone or alongside an observer face-to-face. This approach could sway the end result in favor of the observer or, the end-user could tell the observer what they want to hear, as opposed to what they really think themselves. The structure of how each question is presented could ultimately impact an end-user's decision. Observational methods consist of Cognitive Walkthroughs and Think-Aloud sessions, these methods help understand the thoughts and decisions an end-user makes whilst navigating an application [7]. All of these issues link back to the Hawthorne Effect, as to whether an observer is influencing their decision [4]. This is where the framework would be of benefit to the developer.

The UX industry has been using evaluation methods throughout their design process however, evaluation metrics are a developing area. Evaluation metrics measure the UX of an application to calculate a score. UserZoom created a single UX metric called qxScore [8]. qxScore benchmarks the experience of an application by evaluating two areas: behavior, relating to task success rate; and attitude, including trust, ease of use and appearance [8]. A qxScorecard generates the results by indicating the quality of experience from 0 to 100, >45 being very poor and 91-100 being great. This UX metric covers the fundamentals of UX evaluation however, it is then up to the developers and stakeholders within the company to decide on what improvements to make that will enhance the UX of their application, and there lies a gap. Alternative methods of evaluation have been used within other domains, such as education.

Within [9], they used a UX/UI evaluation framework within the cyberlearning environment to evaluate two main areas: usability; and utility. Other areas of interest included: technology, users; and context. Usability attributes consisted of problem-based learning and ease of use evaluations via Cognitive Walkthroughs and Heuristic surveys [9]. Utility attributes used pre/post-test and final scores for learning achievements, and a UX/UI survey to document the evaluation of user satisfaction [9]. These are appropriate evaluation methods and are possibly more manual and traditional compared to qxScore. Although it may be a time-consuming process, there are other factors that could assist and make the evaluation process more engaging, such as gamification.

Gamification is the incorporation of gaming elements, such as: point scoring, leader boards; and levels [10], all of which were incorporated into [9]. This kept the students engaged for longer, return often to complete assignments; and allow them to be in competition against their fellow classmates. In return, this led to honest feedback. This is due to the lack of pressure as nobody was watching, and gamification assisted in providing a sense of enjoyment whilst also keeping them engaged. In addition, gamification also allows for gaps to appear that highlight topic areas that a student might be struggling with, this has been demonstrated within M-Elo.

M-Elo incorporated gamification elements to identify the knowledge gap of each student, whilst also considering their parameters [11]. These parameters were independent and helped model each end-user's knowledge state which assisted in the recommendation process. A visualization widget allowed each end-user to track their current knowledge against their peers. In return, the application provided questions based upon their largest knowledge gap [11]. A Likert-scale survey was then used to capture end-user feedback, covering areas such as: motivation, rationality; and trust [11]. Students detailed that the incorporation of peer comparison provided a sense of trust that encouraged motivation in order to progress their education to the next level, based upon the appropriate recommendations provided [11]. This in return can reduce their cognitive load, this is the amount of cognitive effort required to understand the topic, presentation and sequence of events whilst using an application [12]. All of these factors detailed will be considered for the framework in question to assist the developer within their software development process. The framework will objectively measure and indicate where an application fits within the three aspects (this is currently not provided within other research), highlight the areas of growth and make appropriate recommendations, without an observer influencing its decisions. The framework will not only improve the user journey for each end-user, but it will provide a clear direction for the developer.

### 3 Methodology

The framework is based upon a three phased approach and therefore three studies will fulfil each phase, phase one is currently underway.

Phase one of this framework will input data from specific domains (which is still ongoing research) in relation to the UX of the application, for example education. A hierarchy flow weight measurement will assist in producing a score. This is the main score that the developer receives about how Dynamic, Adaptive and Intelligent the UX

of their application is. In order for a score to be established, the following measurement process must take place.

The measurement takes into account that the UX of an application as a whole is marked out of 100%. This total percentage is divided between each aspect of the framework: 33.33% Dynamic, 33.33% Adaptive; and 33.33% Intelligent. Each aspect holds parameters that are specific to each aspect, these have been identified and assigned during initial research. Each parameter within its assigned aspect is assigned a weight out of 100% and is based upon how much influence and value it would contribute to the UX of an application - the heavier the weighting, the greater influence on the scoring. In addition, each parameter contains sub-parameters. These sub-parameters detail what is required within each parameter in order to achieve the full weighting listed. Each sub-parameter has its own weighting in accordance to the influence and value that is required in order to enhance the UX. Table 1 below illustrates the parameters and sub-parameters in each aspect and their weights which have been weighed out of 100%.

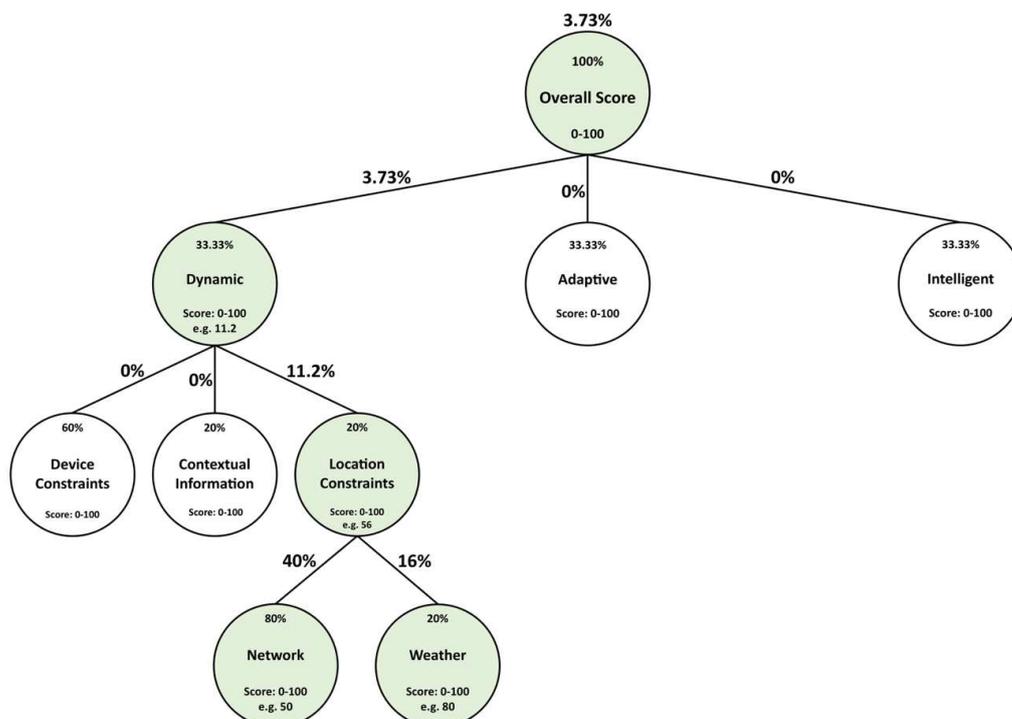
**Table 1.** Sub-Parameters and Weights within Assigned Parameters.

	Aspect	Parameter	Weighting	Sub-Parameter	Weighting
<b>Overall Score 0-100</b>	<b>Dynamic</b> 33.33%	Device Constraints	60%	Device Type	25%
				Screen Size	25%
				Battery Life	20%
				Processor Speeds	20%
				Memory Storage	5%
				Internet Speed	5%
		Contextual Information	20%	Home	30%
				Work	30%
				Walking	10%
				Time of Day	30%
		Location Constraints	20%	Network	80%
				Weather	20%
	<b>Adaptive</b> 33.33%	Capabilities	25%	Visual	40%
				Hearing	10%
				Cognitive	15%
				Physical	35%
		Knowledge Set	15%	Cohort	75%
				Technical Expertise	25%
		User Goal	60%	Searching	25%
Browsing				25%	
Submitting Forms				25%	
Checking off a Task	25%				
<b>Intelligent</b> 33.33%	Knowledge Based Rules	100%	Personalized to User	100%	

Figure 2 below is a hierarchy diagram providing a hypothetical example as to how the measurement and scoring would work. The circles that are highlighted in green (darker circles) will be used for demonstration purposes to showcase one set of parameters. Nodes with no colour detail the other weighting percentages that have been distributed.

As mentioned, an application as a whole is marked out of 100% (root node), and each of the three aspects (parent nodes) have been distributed a percentage of it: 33.33% Dynamic, 33.33% Adaptive; and 33.33% Intelligent. Every node from the root, to a leaf node within Figure 2 will be scored between 0 and 100, this score works in conjunction with the weighting that each parameter and sub-parameters have been allocated. For example, to understand the scoring detailed, we would start from the Network and Weather leaf nodes at the bottom of Figure 2 and work upwards.

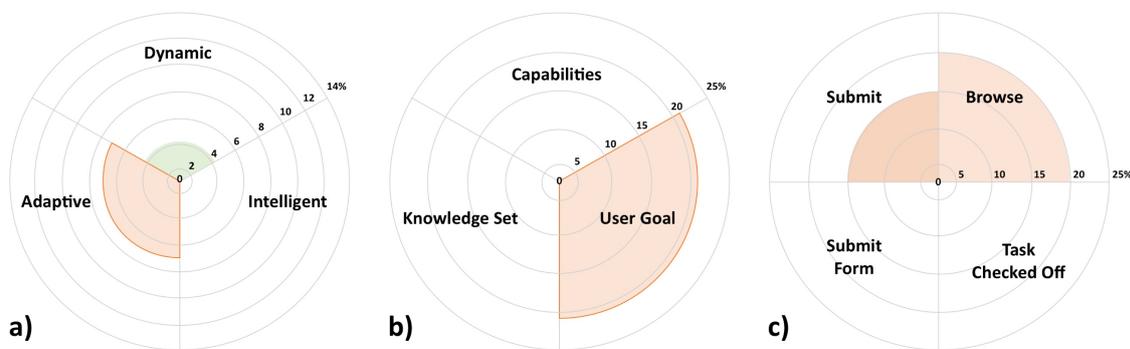
The Network (sub-parameter) leaf node is in relation to the Location Constraints parameter. As illustrated within Figure 2, the Network leaf node has been allocated a score of 50/100, this makes converting to a percentage easier. For example, the Network leaf node has been given a weight of 80%, the score of 50 equates to 50% of the 80% weight, this means that Network has a score of 40%. The same principle is applied to the Weather leaf node, it has been given a score of 80 which equates to 80% of the 100 possible marks. 80% of the 20% weighting is equal to 16%. The percentages from each leaf node (40% and 16%) are added together to form the score for Location Constraints which is 56 out of the 100 possible marks. 56% of 20% Location Constraint allocation is equal to 11.2% and this then equates to the total score for the Dynamic parent node. As the Dynamic aspect is worth 33.33% of the 100 possible marks, the same calculation applied to the 11.2%, 11.2% of the 33.33% allocation to the Dynamic parent node is 3.73%. This means that the overall score within this example works out at 3.73% as only one aspect is being demonstrated (as an example due to the limited of space available), which is a very poor score. In order for a developer to understand if this score is poor or not, a form of gamification would be applied.



**Fig. 2.** An Example of Measurement and Scoring Allocation for the Dynamic Aspect via a Hierarchy Diagram.

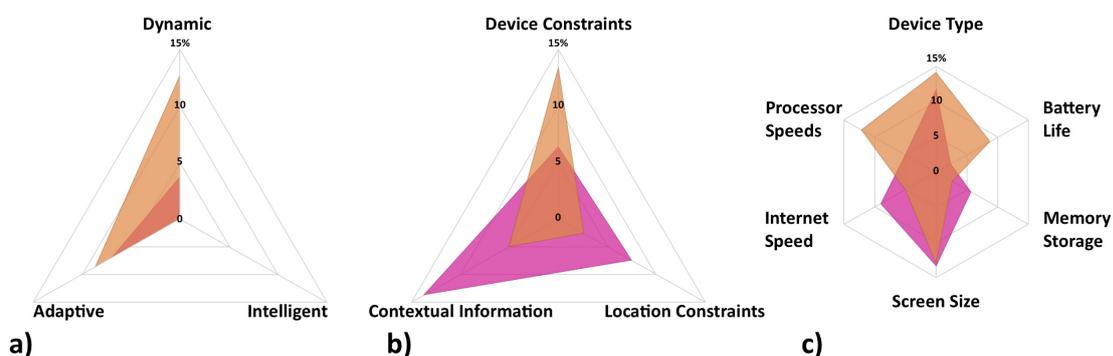
In relation to gamification, most car racing games award bronze, silver or gold medals to those players who finish 3rd, 2nd or 1st. The same principle can be applied to the framework in question, these could be known as scoring boundaries. For example: scores between 0-39 would be bronze, 40-89 would be silver; and 90-100 would be gold. This would be the main score that a developer would see, as it is the aggregation of scores from the three core aspects.

Within education, students use a virtual learning environment known as Blackboard Learn (BBL). To produce a score for BBL, multiple forms of media have been taken into consideration: data, the application and screenshots supplied, the hierarchy diagram from Figure 2 was used to produce a final score. Figure 3 below illustrates the results: chart A indicates the percentage of each aspect that is currently being fulfilled, the highlighted border around the Adaptive aspect indicates what will be shown in chart B, chart B is drilled down from the Adaptive aspect in chart A detailing the parameters; and chart C is drilled down from chart B indicating the percentage of sub-parameters.



**Fig. 3.** BBL Results via Three Radar Visualizations.

The radar charts above show the scoring of selected parameters and sub-parameters. Based upon the scoring boundaries within this framework previously mentioned, BBL is 11.1%. As the numbers are between 0 and 39, it is categorized as being bronze. Phase two of this framework would then indicate the areas of growth and provide recommendations. The radar charts being used in Figure 3 are good to use for individual software solutions, while those provided in Figure 4 below illustrate what two software solutions would look like when compared and scored against the framework. Figure 4 is illustrating hypothetical results, and this would allow for a similarity score.



**Fig. 4.** Hypothetical Results Overlaying Two Software Solutions that are being Measured Against the Framework.

Phase two of the framework continues on from phase one and this will be the second study of the PhD. The main focus of phase two is to take the gaps (areas of growth) that are clearly visible from Figure 3 and provide recommendations to help the application improve. The recommendations would assist the developer by advising them on what their application needs in order to improve not only its UX, but its scoring that the framework has supplied.

As an example, based upon the results from BBL, it would be helpful if recommendations could be supplied in relation to the Intelligent aspect to boost its scoring. By boosting the Intelligent aspect, it would allow the application to provide material that is relevant to that particular student, whilst working with other parameters, such as type of device. Recommendations could be as simple as a tooltip, to draw the developer's attention to the parameter and sub-parameter from the charts illustrated above.

Phase three of the framework is the final phase. The main purpose of this phase is to validate the recommendations, and to accept that they do in fact enhance the UX of an application. In order to provide a form of validation, progress history would be an important factor for the developer. Progress history is in relation to the previous measurements and scores that the framework has produced from the same application. This allows the developer to see the progress their application is making in order to produce a more DAIUI. By viewing a progress history, it reassures the developer that the framework is having an impact on the UX of their application.

## 4 Conclusion and Future Work

The work presented here has detailed that not all applications are aware, or capable of measuring the three recognized aspects of: Dynamic, Adaptive; and Intelligent. In order for these aspects to be recognized, a holistic UX methodological based assessment framework has been outlined. It will assist and benefit developers within a variety of domains with their software development process. Overall, this will enhance the UX of their products/service solutions within their preferred domains.

Besides phases two and three of the framework, future work will consist of the translation of measurements that each sub-parameter contains. Translating them into a format that the framework will understand. Further work would entail how to automate the identification and measurement of each aspect in order to produce a score.

### Acknowledgments

The authors wish to acknowledge the support of the European Commission on the H2020 project MIDAS (G.A. nr. 727721) and DfE NI (Department of the Economy Northern Ireland).

## References

1. V. Johnston, M. Black, J. Wallace, M. Mulvenna, and R. Bond, “A Framework for the Development of a Dynamic Adaptive Intelligent User Interface to Enhance the User Experience,” in *Proceedings of the 31st European Conference on Cognitive Ergonomics*, New York, NY, USA: Association for Computing Machinery, 2019, pp. 32–35.
2. “What is User Experience (UX) Design?,” *The Interaction Design Foundation*. <https://www.interaction-design.org/literature/topics/ux-design> (accessed Dec. 20, 2018).
3. “What is parameter? - Definition from WhatIs.com,” *WhatIs.com*. <https://whatis.techtarget.com/definition/parameter> (accessed Nov. 07, 2018).
4. R. Macefield, “Usability studies and the Hawthorne effect.” Usability Professionals’ Association, May 01, 2007.
5. A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, “User experience evaluation methods: current state and development needs,” in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, Reykjavik, Iceland, Oct. 2010, pp. 521–530, doi: 10.1145/1868914.1868973.
6. J. Hussain *et al.*, “A Multimodal Deep Log-Based User Experience (UX) Platform for UX Evaluation,” *Sensors*, vol. 18, no. 5, Art. no. 5, May 2018, doi: 10.3390/s18051622.
7. S. Martin, “Measuring cognitive load and cognition: metrics for technology-enhanced learning,” *Educational Research and Evaluation*, vol. 20, no. 7–8, pp. 592–621, Nov. 2014, doi: 10.1080/13803611.2014.997140.
8. “Introducing UserZoom’s single UX metric for experience benchmarking: the qxScore,” *UserZoom*, Apr. 15, 2019. <https://www.userzoom.com/using-userzoom/introducing-userzooms-single-ux-metric-for-experience-benchmarking-the-qxscore/> (accessed Jun. 28, 2020).
9. H. W. Alomari, V. Ramasamy, J. D. Kiper, and G. Potvin, “A User Interface (UI) and User eXperience (UX) evaluation framework for cyberlearning environments in computer science and software engineering education,” *Heliyon*, vol. 6, no. 5, p. e03917, May 2020, doi: 10.1016/j.heliyon.2020.e03917.
10. “gamification | Definition of gamification in English by Oxford Dictionaries,” *Oxford Dictionaries | English*. <https://en.oxforddictionaries.com/definition/gamification> (accessed Dec. 20, 2018).
11. S. Abdi, H. Khosravi, S. Sadiq, and D. Gasevic, “A Multivariate Elo-based Learner Model for Adaptive Educational Systems,” *arXiv:1910.12581 [cs]*, Oct. 2019, Accessed: Mar. 13, 2020. [Online]. Available: <http://arxiv.org/abs/1910.12581>.
12. W. L. in R.-B. U. Experience, “Minimize Cognitive Load to Maximize Usability,” *Nielsen Norman Group*. <https://www.nngroup.com/articles/minimize-cognitive-load/> (accessed Mar. 13, 2020).

# A Conceptual Architecture for AI-based Big Data Analysis and Visualization Supporting Metagenomics Research

Thoralf Reis, Thomas Krause,  
Marco X. Bornschlegl, and Matthias L. Hemmje

University of Hagen, Faculty of Mathematics and Computer Science, 58097 Hagen,  
Germany {thoralf.reis, thomas.krause, marco-xaver.bornschlegl,  
matthias.hemmje}@fernuni-hagen.de

**Abstract.** This paper targets to introduce an architecture for Artificial Intelligence (AI) based Big Data Analysis and Visualization supported metagenomics research based on the AI2VIS4BigData Reference Model. Metagenomics research covers the examination of huge amounts of data to improve the understanding of microbial communities. Technological and methodical improvements in Big Data Analysis drive progress in metagenomics research and thereby support practical applications like, e.g., the analysis of cattle rumen with the research goal of reducing the negative impact of cattle breeding on global warming. AI2VIS4BigData is a reference model for the combined application areas of Big Data Analysis, AI, and Visualization. Its purpose is to support scientific and industrial activities with guidelines and a common terminology to enable efficient exchange of knowledge and information and thereby prevent "reinventing the wheel". The general applicability of the AI2VIS4BigData model for metagenomics has been validated in a previous publication. As a next step, this paper derives a conceptual architecture that specifies a possible adaption of AI2VIS4BigData for metagenomics. For this, three new metagenomic publications utilizing AI and Visualizations are assessed.

**Keywords:** Metagenomics · Big Data · AI · Visualization · AI2VIS4BigData.

## 1 Introduction and Motivation

Metagenomics research analyzes relationships within whole microbial communities while genomics research focuses on the analysis of genes or the genome of a single organism [1]. A practical example for metagenomics research is the investigation of the rumen microbiota regarding its influence in cattle greenhouse gas emissions and food conversion efficiency [2] as cattle are a major contributor to climate change and relevant for food security, two significant challenges society is facing [2]. The demand for data in metagenomics research is significantly bigger than for regular genomics research: the investigation of relationships and coherence between organisms or genes in and between metagenomic samples

[1] requires biological researchers to process, store, and exchange big amounts of data e.g. via specialized bioinformatics databases [3]. Hence, metagenomics research benefits on a large scale from progress and development in Big Data Analysis such as decreasing costs for storage and processing of huge amounts of data. With the EU-funded MetaPlat<sup>1</sup> project, scientists from different research institutions with either Big Data Analysis or bioinformatics background worked together to develop the MetaPlat platform. This cloud based Big Data Analysis platform is specialized to analyze metagenomics data like, e.g., rumen microbiota [2]. For an effective analysis of Big Data, the platform empowers the researchers to utilize cutting-edge technology such as Artificial Intelligence (AI) [2] and various forms of Information Visualization (IVIS) [4] to provide the researchers with visual feedback of their activities and enable them to identify new insights.

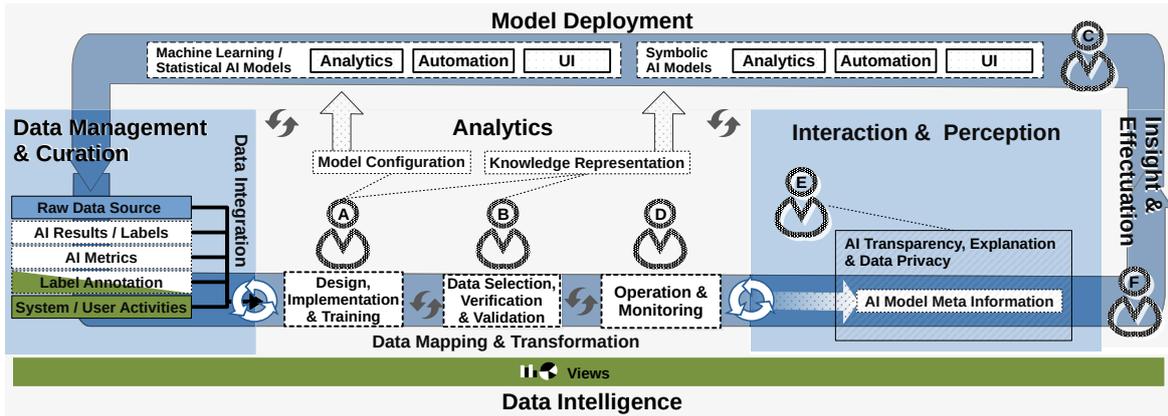
To define the vague term Big Data, a popular approach is to follow the data management challenges outlined by Doug Laney [5]. These challenges comprise three dimensions (the three v's): variety (ambiguous data manifestations regarding e.g. data format, data structure or data semantics), volume (big amount of data), and velocity (high frequent data inflow) [5]. By this definition, the sheer volume of data in metagenomics research allows labeling it as Big Data. The collective term AI summarizes techniques and methods such as symbolic AI or Machine Learning (ML) to implement intelligence for machines (in contrary to human or animal natural intelligence) [6]. Example application scenarios of AI in metagenomics research of rumen are the analysis of data through clustering [3] or the training of classifiers to categorize data samples [2]. Big Data and AI are closely connected to each other [6]: Big Data is very useful to derive, validate, apply, and enhance AI models while AI-driven algorithms enable the exploration of Big Data and its potential. Visualization of data, processing steps as well as AI model development is an important link between both application areas. It enhances comprehension and decreases entry barriers for new users. In addition, visualization offers the chance to meet the growing demand for explainability and transparency of AI

With [7], the AI2VIS4BigData Reference Model for research and practical applications in the application areas Big Data Analysis, AI, and Visualization was introduced. Its objective is to provide a common specification as well as a common basis for discussion and thereby reduce the risk of inefficiency through reinventing the wheel and solving problems that have already been solved elsewhere. The reference model's theoretical applicability was evaluated in an expert round table workshop featuring presentations from three practical application domains: health care, economics, and metagenomics [8]. Until now, the reference model was validated only for one metagenomics research application [9]. This paper targets to validate three further metagenomics research applications from the MetaPlat project to assess if an architecture can be derived.

Within the remainder of this paper, the AI2VIS4BigData Reference Model and the three assessed metagenomics research publications from the MetaPlat project are introduced, the pursued architecture modeling approach is presented

---

<sup>1</sup> <https://metaplat.eu>



**Fig. 1.** AI2VIS4BigData - A Reference Model for AI Supporting Big Data Analysis

(Section 1.3), a multi-layered conceptual architecture is introduced in Section 2 together with an initial validation in Section 3 before this paper concludes with outlining its contributions and providing an outlook (Section 4).

### 1.1 AI2VIS4BigData Reference Model

The AI2VIS4BigData Reference Model (Figure 1) was derived through projecting the AI lifecycle phases of AIGO’s AI System Lifecycle [10] onto the IVIS4BigData Reference Model [11] considering the different AI models (ML or statistical AI models as well as symbolic AI models) for supporting Big Data Analysis, AI data, and AI user stereotypes [7]. It contains the three processing steps *Data Management & Curation*, *Analytics*, *Interaction and Perception* accompanied with a data intelligence layer for user interaction and User Interfaces (UI) of IVIS4BigData [11], a reference model that target to “close the gap in research with regard to information visualization challenges of Big Data Analysis as well as context awareness” [11]. AI2VIS4BigData introduces a model deployment layer that spreads over the three processing steps [7]. AI models are executed directly within the data and information loop which links the deployed models to the input data they need for execution and compute output data that is fed back into the Big Data Analysis system [7]. The remaining activities of AI system life cycle phases are displayed within the analytics layer as *Design, Implementation & Training*, *Data Selection, Verification & Validation* as well as *Operation & Monitoring* and interconnected through bidirectional arrows emphasizing the iterative nature of AI model design [7]. The different reference model elements are linked to five clearly distinguished AI user stereotypes (model designer, domain expert, model deployment engineer, model operator, model end user, and model governance officer) and four clearly distinguished Big Data Analysis user stereotypes (system owner, data scientists, management consultants as well as directors including C-levels) [7].

### 1.2 Assessed Metagenomic Use Cases

In [9] a conceptual workflow for metagenomic studies was presented and demonstrated using two previously published metagenomic use cases. The first of these

use cases [12] was the visualization of gene dependencies using a whole-genome approach and a new framework for improved correlation measurement between genes. The second publication [13] analyzed the relationship between microbiomes in feces and rumen using a taxonomic analysis of partial genome sequences (barcode sequences). Together they cover the two main branches of metagenomic analyses (taxonomic and functional). It was shown in [9] that the metagenomics analysis workflow extracted from these publications can be mapped directly onto the AI2VIS4BigData Reference Model therefore validated its relevance for the field of metagenomics.

This section will introduce three additional publications in metagenomics research that serve as a base to further validate and transform this conceptual workflow into a generic architecture. The publications were selected as they are practical examples for metagenomic analysis (which can represent Big Data Analysis applications depending on the sample size), carried out by different researchers and most importantly, they describe the usage of statistical methods or ML as well as Visualization. Although all selected publications originate from the MetaPlat project, they represent different research approaches like, e.g., the analysis of genes or the analysis of OTUs. In addition, the homogeneous MetaPlat terminology eases the architecture derivation.

The publication *A Metagenomics Analysis of Rumen Microbiome* [2] by P. Walsh et al. demonstrates a metagenomic analysis of the "Bos taurus" rumen microbiome using ML models in a cloud based environment. For optimal performance and scalability, a queueing system is used between individual components, thus enabling asynchronous and parallel execution. After importing raw sequence data into the system, it is written to one of these processing queues which feed into a similar metagenomic analysis workflow that uses the QIIME toolset to perform data cleanup and clustering of sequences into Operational Taxonomic Units (OTUs). The workflow assigns taxonomic labels to these OTUs. In an analytics step, various ML models are used to classify the samples into phenotypes using the taxonomic data of the previous steps as an input. Finally, the publication showcases various visualizations ranging from a taxonomic composition chart to plots of algorithmic accuracy and other AI metrics.

In *Analysis of Rumen Microbial Community in Cattle through the Integration of Metagenomic and Network-based Approaches* [3], H. Wang et al. functionally analyze the rumen microbial community in cattle through application of a network-based approach: the authors construct a co-abundance network utilizing the "relative abundance of 1570 microbial genes" [3] that enables them to identify functional modules. In doing so, they present a method to automatically determine a cutoff threshold value to generate the co-abundance network in the first place [3]. While the first publication [2] uses partial sequences sufficient to identify and analyse the taxonomic composition, this publication is based on whole genome data which enables the analysis of genes. Together they cover the two main branches of metagenomic studies. To construct the co-abundance network used in the publication, the short reads generated by next-generation sequencing platforms are assembled into longer sequences. These sequences are

then matched to the KEGG<sup>2</sup> database to identify genes (and associated meta-data) present in the samples. Using the relative abundances of these genes, correlations can then be computed by analyzing how the abundance of one gene affects the abundance of other genes across the various samples. Since the presence or absence of a correlation is not always distinguishable from statistical noise, a suitable cutoff value is then determined using an automated computational method. Using the cutoff values, a network graph can be constructed that represents genes as nodes and the correlation strength as the length of edges connecting these nodes.

As third and last assessed publication, M. Wang et al., the authors of *Understanding the relationships between rumen microbiome genes and metabolites to be used for prediction of cattle phenotypes* [4] combined metabolomics with metagenomics in order to identify differences in diets and methane emissions from rumen metabolites and microbial genes. They analyzed 36 rumen samples and identified the difference in the response of rumen microbes to different basal diets which down the road affect cattle methane emissions [4]. The study starts from gene abundance data of cattle rumen obtained from previous studies on the experiment designed by Roehe et al. [14]. The abundance data was cleansed and transformed before conducting multiple activities to determine correlations between genes and metabolites related to the differences in diets in the experiment design. The correlation data was then used to build correlation networks as well as various other plots and result tables.

### 1.3 Discussion, Conclusion, and Identification of remaining Architectural Challenges

In order to arrive at a generic architecture that enables the management, analysis, and visualization of metagenomic data as well as the fusion with other health related data and knowledge, the first step is mapping the introduced metagenomics publications to the generic stages of the AI2VIS4BigData Reference Model ("Data Management & Curation", "Analytics" and "Interaction & Curation"). This is easy to validate as all three publications include steps to ingest, manage or cleanup metagenomic sequences, all of them include statistical or ML methods for analytics and also all of them produce one or more visualizations. The same was previously already demonstrated in [9]. Therefore, it is proposed that an architectural model should explicitly model these stages.

Looking at the papers in detail, further requirements for a comprehensive architectural model can be derived: The first publication describes the importance of using individual components that communicate through asynchronous mechanisms like, e.g., queuing systems to achieve high performance and scalability. The impact of Big Data and ML in Metagenomics is also mentioned as a challenge in [9]. A suitable architecture should therefore aim to separate individual parts and components of the system where possible so that they can operate and scale individually. The second publication [3] shows the need of additional

---

<sup>2</sup> Kyoto Encyclopedia of Genes and Genomes, <https://kegg.jp>

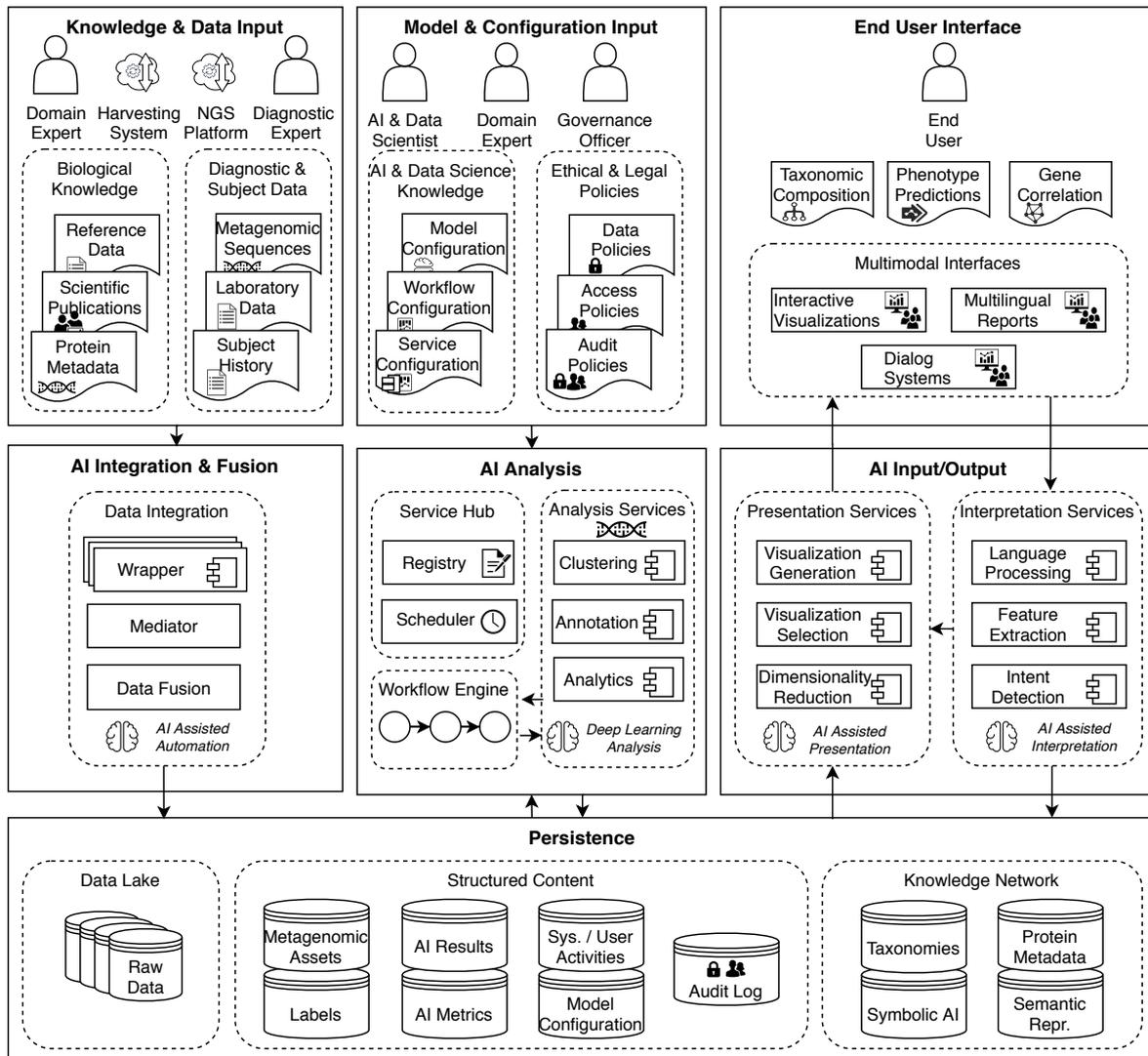
knowledge sources like, e.g., gene databases for the analysis of metagenomic sequences. Our proposed architecture should therefore support the ingestion and persistence of these additional data sources into a knowledge network that can be used by metagenomic workflows. The third publication [4] is important as it does not start from raw sequence data but from intermediate results obtained from other studies. Our architecture should be able to reuse the same intermediate results for several distinct analyses thus requiring the persistence of these intermediate results. This requirement also partially addresses the challenge of "Reproducibility" mentioned in [9] and the area of "AI Transparency, Explanation & Data Privacy" of the AI2VIS4BigData model. All three publications differ significantly in the exact steps executed in the analysis phase and the visualizations produced. It is therefore important that the analysis is done in a modular fashion where the order and type of steps is dynamic and that a wide range of visualizations is supported.

## 2 AI2VIS4BigData Conceptual Architecture supporting Metagenomics Research

This paper introduces the AI2VIS4BigData architecture (Figure 2) for processing and analysis of metagenomic data in an AI and Big Data environment. It was designed by extending the Big Data Analysis and Visualization architecture of IVIS4BigData [11] with AI and metagenomic aspects in order to fulfill the metagenomics requirements outlined in Section 1.3. The architecture is vertically split into three pillars separating the components for metagenomics data integration and processing (domain-specific input), AI and data science modeling and configuration (AI analysis input) from the components responsible for result visualization and data generation (output). This is based on the design principle of Separation of Concerns (SoC) [15] and makes it easier to develop, scale or exchange the components separately. Each of these three pillars is structured into three layers following the Model View Controller (MVC) pattern [16] with a shared persistence layer interconnecting all three pillars. The bottom layer represents the model, the top layers represent the view while the middle layers contain the controllers. Metagenomics-specific architecture elements are a dedicated user, knowledge and data artifacts within the input layer, assets and knowledge networks in persistence layer as well as domain-specific end user interfaces. The following rough description of the individual layers and components follows the flow of data, starting from the top left at data input and ending with result visualization at the top right corner:

**Knowledge & Data Input.** Within this layer, expert users or systems ingest metagenomics-related knowledge and data into the system. These information comprise biological and genetical knowledge (e.g. protein metadata or knowledge automatically extracted from scientific publications) as well as diagnostic and subject data (e.g. metagenomic sequences).

**AI Integration & Fusion.** This layer contains all services and methods to integrate the various domain-specific inputs into the system, to perform a data fusion and persist it as structured content or knowledge network. The se-



**Fig. 2.** AI2VIS4BigData Conceptual Architecture Supporting Metagenomics Research  
 mantic integration is realized through implementation of the mediator wrapper approach.

**Model & Configuration Input.** The necessary knowledge and information for configuring the AI applications within the system is provided by AI and data science expert users within this layer. The input contains the required knowledge to register and schedule all AI services and to select appropriate analysis methods and algorithms. The additional AI2VIS4BigData role of the *Governance Officer* ensures legal compliance and maintaining ethical standards through providing relevant constraints.

**AI Analysis.** The middle layer is responsible for performing analysis tasks on behalf of the user. A workflow system together with a service registry allow for flexible configuration of the required analysis steps while the scheduler manages the execution of these steps on distributed or local computing nodes. Intermediate and final results are stored persistently.

**Persistence.** The persistence layer targets to store various types of data and enable data exchange between overlying layers. Raw data is stored in a data lake with little to none processing performed to improve reproducibility and trans-

parency of the system. Structured data includes parsed genetic sequences, intermediate results from analysis processes and other kind of schema-bound data. Lastly a knowledge network tries to represent biological and medical knowledge as well as semantic rules required for Symbolic AI in a machine readable way.

**AI Input/Output.** The purpose of this layer is to intelligently interpret the intentions of the system's end user (e.g. through applying natural language processing) and present the information that is relevant for them in a suiting form (e.g. after performing a dimensionality reduction or selecting appropriate visualization techniques).

**End User Interface.** The end user interface layer contains the multimodal interfaces through which the system's end users access its data and information. These interfaces comprise visualizations, reports and dialogue systems that present the domain-specific artifacts (e.g. taxonomic compositions).

### 3 Initial Validation and Remaining Challenges

The proposed architecture specifies all areas of the AI2VIS4BigData Reference Model. The area "Data Management & Curation" of the reference model is addressed by the left pillar. The "Analysis" area is covered by the second pillar and especially the "AI Analysis" layer. Finally, the "Interaction & Perception" area is implemented through the right pillar. The architecture also implements all requirements that were outlined previously. A detailed mapping of the requirements to the architecture elements would be beyond the scope of this paper, yet is planned for future work. Individual components for input, analysis and visualization are strictly split by the three pillars and communicate asynchronously through the persistence layer allowing for flexible scaling and Big Data processing. Additional knowledge sources are supported by providing a data agnostic input layer together with a mediator wrapper architecture for data integration. The persistence layer ensures that reproducibility and transparency is possible by storing intermediate and final results. Finally, a flexible workflow system and a service registry support the heterogeneity of metagenomic studies and allow easy integration of new analysis methods. The remaining challenges for the architecture comprise a harmonization with the IVIS4BigData architecture, a generalization for application domains beyond metagenomics research, a technical specification as well as a proof of concept technical implementation. Since the selected publications were limited to the MetaPlat project, the assessment of practical applicability for the introduced architecture in metagenomics research beyond MetaPlat is a further remaining challenge.

### 4 Conclusion and Outlook

In the course of this paper, three MetaPlat publications were assessed that analyze rumen microbiota through metagenomics research utilizing Big Data Analysis, AI as well as visualization. Objective of this assessment was the derivation of a AI2VIS4BigData-based conceptual architecture for real-life application in this three-fold research area. The resulting AI2VIS4BigData conceptual architecture supporting metagenomics research was introduced in Section 2. It consists of

seven layers arranged alongside the three levels of the MVC pattern. As outlook, future work is planned to overcome the challenges introduced in Section 3.

## References

1. F. Engel, M. Fuchs, P. M. Kevitt, M. Hemmje, and P. Walsh, “A Metagenomic Content and Knowledge Management Ecosystem Platform,” 2019.
2. P. Walsh, C. Palu, B. Kelly, B. Lawor, J. T. Wassan, H. Zheng, and H. Wang, “A Metagenomics Analysis of Rumen Microbiome,” *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 2077–2082, 2017.
3. H. Wang, H. Zheng, F. Browne, R. Roehe, R. J. Dewhurst, F. Engel, M. Hemmje, and P. Walsh, “Analysis of Rumen Microbial Community in Cattle through the Integration of Metagenomic and Network-based Approaches,” *2016 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 198–203, 2017.
4. M. Wang, H. Zheng, H. Wang, R. J. Dewhurst, and R. Roehe, “Understanding the relationships between rumen microbiome genes and metabolites to be used for prediction of cattle phenotypes,” in *BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering*. VDE, 2019, pp. 1–5.
5. D. Laney, “3D Data Management: Controlling Data Volume, Velocity, and Variety,” META Group, Tech. Rep., 2001.
6. ISO, “ISO/IEC JTC 1/SC 42 Artificial Intelligence,” 2018. [Online]. Available: <https://isotc.iso.org/livelink/livelink/open/jtc1sc42>
7. T. Reis, M. X. Bornschlegl, and M. L. Hemmje, “Towards a Reference Model for Artificial Intelligence Supporting Big Data Analysis,” *To appear in: Proceedings of the 2020 International Conference on Data Science (ICDATA’20)*, 2020.
8. —, “AI2VIS4BigData: Qualitative Evaluation of a Big Data Analysis, AI, and Visualization Reference Model,” *To appear in: Lecture Notes in Computer Science*, vol. LNCS 10084, 2020.
9. T. Krause, B. Andrade, H. Afli, H. Wang, H. Zheng, and M. Hemmje, “Understanding the Role of (Advanced) MachineLearning in Metagenomic Workflows,” *To appear in: Lecture Notes in Computer Science*, vol. LNCS 10084, 2020.
10. OECD, *Artificial Intelligence in Society*, 2019.
11. M. X. Bornschlegl, “Advanced Visual Interfaces Supporting Distributed Cloud-Based Big Data Analysis,” Dissertation, University of Hagen, 2019.
12. H. Zheng, H. Wang, R. Dewhurst, and R. Roehe, “Improving the Inference of Co-occurrence Networks in the Bovine Rumen Microbiome,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2018.
13. B. G. N. Andrade, F. A. Bressani, R. R. C. Cuadrat, P. C. Tizioto, P. S. N. de Oliveira, G. B. Mourão, L. L. Coutinho, J. M. Reecy, J. E. Koltjes, P. Walsh, A. Berndt, L. C. A., J. C. P. Palhares, and L. C. A. Regitano, “The structure of microbial populations in Nelore GIT reveals inter-dependency of methanogens in feces and rumen,” *Journal of animal science and biotechnology*, vol. 11, p. 6, 2020.
14. R. Roehe, R. J. Dewhurst, C.-A. Duthie, J. A. Rooke, N. McKain, D. W. Ross, J. J. Hyslop, A. Waterhouse, T. C. Freeman, M. Watson, and R. J. Wallace, “Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance,” *PLOS Genetics*, pp. 1–20, 2016.
15. E. W. Dijkstra, “On the role of scientific thought,” in *Selected writings on computing: a personal perspective*. Springer, 1982, pp. 60–66.
16. M. Fowler, *Patterns of enterprise application architecture*. Addison-Wesley Longman Publishing Co., Inc., 2002.

## Chapter 4

---

# COVID-19 Research and Smart Healthcare



# Understanding the Impact of COVID-19 on Behavior Changes and Decision Making of Chinese Students and Researchers in the UK

Hui Yang<sup>1</sup>, Huiru Zheng<sup>2\*</sup>, Xi Liu<sup>1</sup>

1 School of Social Work, China University of Labor Relations, Beijing, China, 100048;  
redyanghui@163.com;

2 School of computing, Ulster University, Northern Ireland, UK, BT37 0QB;  
h.zheng@ulster.ac.uk

\*: corresponding author

**Abstract** With the outbreak of COVID-19 in Wuhan and the subsequent worldwide pandemic, overseas Chinese students and researchers (CSRs) have been hit twice, i.e. worrying about their families during the first phase and then experiencing the pandemic outbreak in a foreign country (UK) as a second hit. This study surveyed 179 CSRs in the UK via an online questionnaire to understand their behavioral patterns, risk assessment, and decision-making intentions during the two phases of the pandemic. The survey showed that (1) CSRs have experienced greater changes in behaviors such as leaving their rooms and going outside, engaging social interaction, hygiene habits, and hoarding; and fewer changes in physical health, daily routines, and sleep patterns; (2) During the second phase, the concern on risk of contracting COVID-19 was significantly increased from 7.8% to 20%; (3) About 60% of CSRs were not confident that they would get effective help in the UK, which was mainly related to the UK's pandemic prevention policy, public awareness, NHS ambulance service regulations, and the medical treatment of COVID-19 in the UK; and (4) A quarter of CSRs had returned or were planning to return to China, and nearly 3/4 of CSRs had decided to remain in the UK because of factors such as academic work, financial pressure (flight tickets have become very expensive), potential risk of infection on travel, taking a wait-and-see attitude, or being optimistic that the pandemic would be over soon.

**Keywords:** COVID-19 and overseas Chinese; COVID-19 and behavior change; COVID-19 and decision making

## 1 Introduction

COVID-19, the infection caused by a novel coronavirus, broke out in Wuhan, Hubei province, China, around early January, 2020, and then swept through most parts of China. When the epidemic had been well-controlled in China, unfortunately the epidemic has evolved into a global pandemic, as announced by the World Health Organization (WHO)<sup>[1]</sup>. Overseas Chinese students and researchers (CSRs) are unique

in the sense that they have been going through both these two phases of COVID-19 pandemic; During first phase they worried about their families and friends back in China, and then in the second phase they have been worrying about themselves in a foreign country. Moreover, during the coronavirus crisis anti-Asian crimes were up 21% in the UK <sup>[2]</sup>, therefore CSRs have to bear the threat and pain caused by stigmatization during the pandemic. In addition, according to *the International Students in the UK Report 2020*<sup>[3]</sup> published by the Association of British Universities, the number of international students from mainland China in the UK has reached 12.6 million (undergraduate and above), accounting for 35% of all non-EU students, as of 2018--2019 academic year statistics, and this is a big number. The above three factors make CSRs as a group, being worth investigating their behaviors during COVID-19 pandemic. What is the impact of this crisis on daily life of CSRs in the UK? Do they think they are safe? And what do they think about their chances of getting effective medical support in the UK? Answers to these questions will help UK understand the needs of CSRs, so that the country might become more attractive to CSRs, and even to international students from other countries as well.

Since the outbreak of the pandemic, research has been carried out on the status of various groups in China affected by the pandemic, such as the anxiety levels of public, medical staff, university students (Zhao et al<sup>[4]</sup>; Hu et al<sup>[5]</sup>) and the impact of the pandemic on the status of international students studying abroad (ABCP<sup>[6]</sup>), but limit research has been conducted for addressing the above questions on the impact of the pandemic on individual lifestyle of international students.

## 2 Methods

### 2.1 Participants and Study Design

Participants were recruited through the Chinese Students and Scholars Associations in the UK and the CSRs social media Wechat Groups. The survey was designed in Chinese, and recruitment was carried out via Chinese social media platform WeChat, which is the most popular social media platform among the Chinese community, and most UK universities and Chinese Students and Scholars Associations have a public account. The survey was launched at a Chinese survey platform Wenjuanxin (<https://www.wenjuan.com/s/mq2Mvqs/>). The study received ethical approval from the Research Ethical Committee of Faculty of Computing, Engineering and Building Environments, Ulster University, UK.

### 2.2 Research Aim

The aim of this research is to carry out a survey on behavior changes among CSRs (including undergraduate students, post-graduate students, PhD researchers, post-docs, and visiting researchers from China) in the UK when the COVID-19 outbreak in China and the UK.

### 2.3 Survey Design

The study was conducted using a self-administered questionnaire. The questionnaire was designed to collect information on CSRs in the UK on their concerns and daily behaviors during the pandemic. It included four sections: (1) Participants' demographics information, such as educational status and study/research area; (2) Behavioral changes after the outbreak in China and the following pandemic in the UK. Questions focus on the impact of COVID-19 on the participant's social communication, shopping, sleep, hygiene, health, and information source related to the COVID-19; (3) Self-assessment of an individual's risk to the COVID-19 and the reasons; (4) Reasons behind behavioral decision-making.

### 2.4 Statistical Analysis

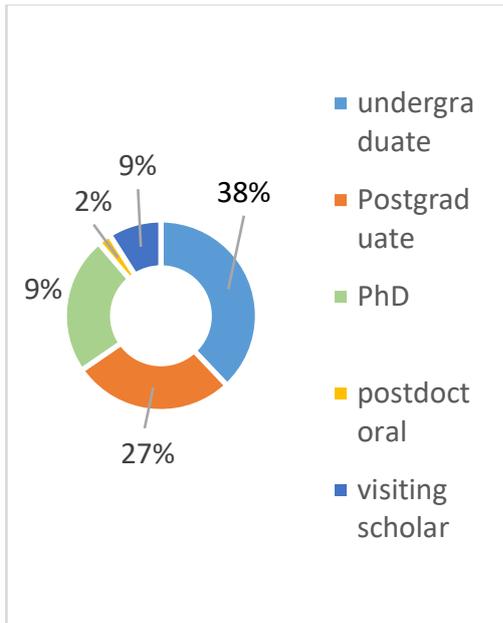
Data were analyzed using SPSS (v 25). All variables were analysed using frequencies and descriptive statistics to determine the number and percentages for each variable.

## 3 Results

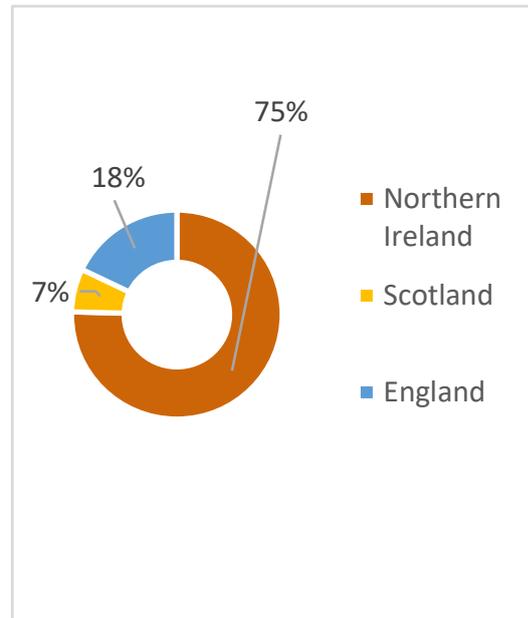
The results consist of four parts: (1) Demographics information of participants; (2) Behavioral changes during the Pandemic; (3) Perceived risk and support; and (4) Decision making and reasons behind.

### 3.1 Demographics of participants

This section includes five pieces of information about the participant's age group, gender, the region of hometown, education level, and study area in the UK. We received 179 questionnaires which are 100% valid. Of the 179 participants, 97 (54%) were female and 82 (46 %) were male. The age range of participants was between 18 and 34 years, among which 145 (81%) participants were aged 18 to 28 years. As shown in Figure 1 and Figure 2, a total of 89% of the participants' educational backgrounds were undergraduate, postgraduate, and doctoral. The majority (75%) of all participants studied in Northern Ireland.



**Fig. 1.** Education Background



**Fig. 2.** Geographical Area of Study

As shown in Figure 3, participants were originally coming from 26 regions of China. 12.8% were from the Guangdong Province, followed by 11.2% from Beijing and 6.1% were from the Shaanxi Province.

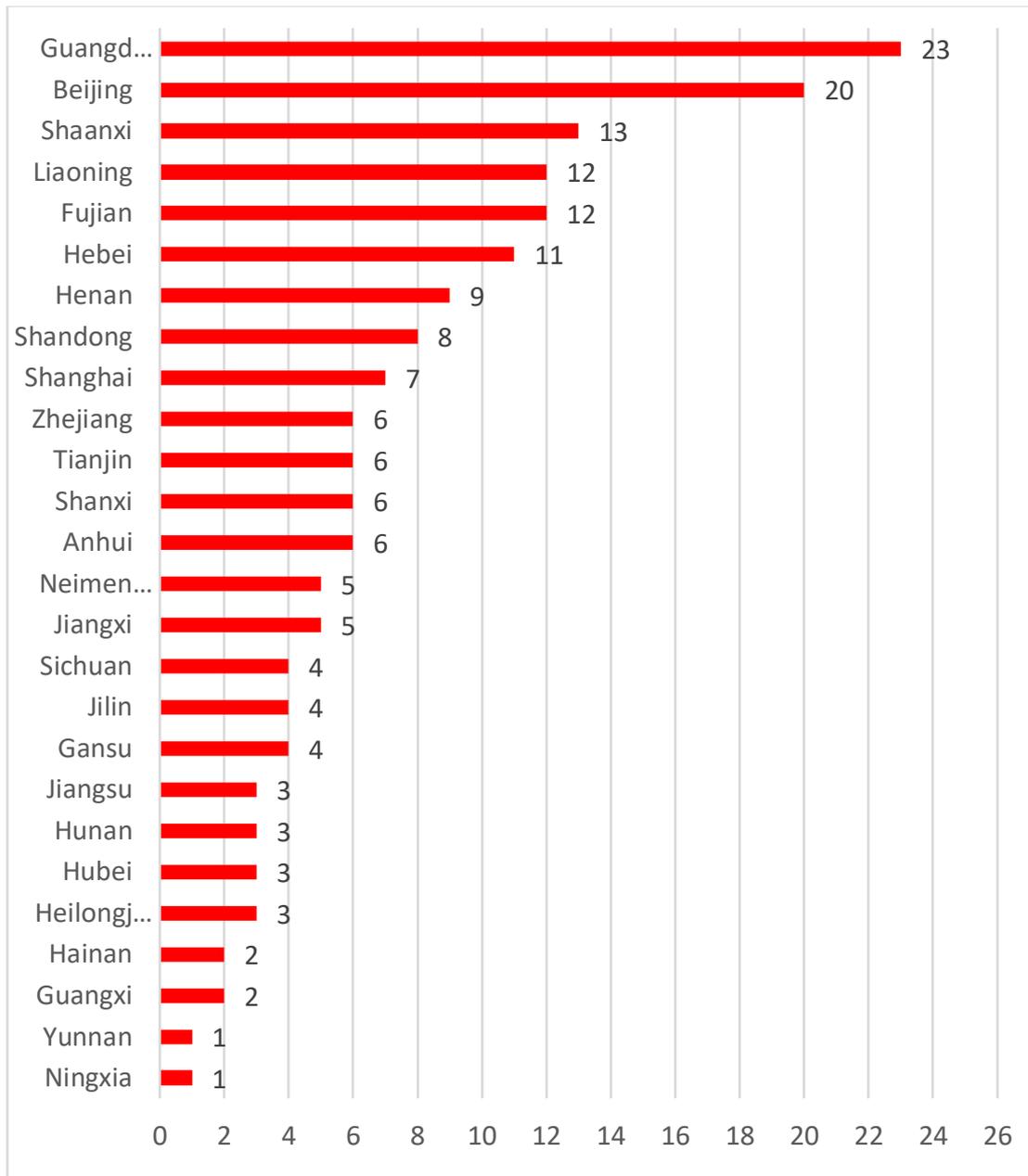


Fig. 3. Geographic Distribution of GSRs

### 3.2 Behavioral Changes of Students in the Pandemic

The pandemic has affected people's behavior and daily life. The CSRs, including undergraduate students, post-graduate students, PhD researchers, post-docs, and visiting researchers from China, in the UK may have been hit twice by this pandemic in two main stages: 1. the COVID-19 broke out in China; 2. the COVID-19 broke out in the UK. It is important to understand how the pandemic has an impact on their daily life, communication, attitude to future career and how they search for information and they may seek support when they need.

This section investigates changes in the behavior of CSRs in the aftermath of the outbreak, including how they first learned about the news, attitudes toward wearing

masks as a foreigner in the UK, changes in the time when they began searching for information, and changes in their lifestyles.

*Question3.2.1: From which source you heard the outbreak?*

*Family / neighbors, schoolmates, friends / Internet (news medias, social media) / others*

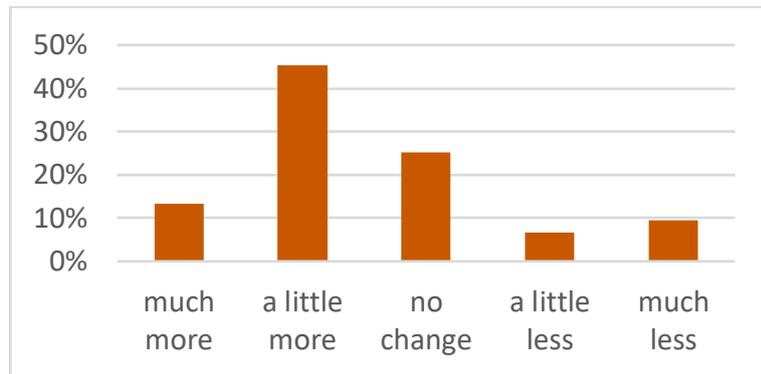
Most (152, 85%) of participants heard this news about pandemic from the internet (news, social media, etc.), 7% of participants got this news from neighbors, schoolmates, friends, and 8% of participants learned from family members.

*Question3.2.2: Did you put on face masks after you heard the outbreak of COVID-19? YES / NO / Maybe later/ Wanted but didn't*

Almost 32% of participants wore the mask after the COVID-19; while there were 31% of participants chose to go unmasked; 19% said they might wear masks in the future, and 18% wanted to wear a mask, but for some reason, they didn't.

*Question3.2.3: Time spent searching for information about COVID-19 compared to time spent searching for information in the past.*

*Much less / a little less / same / a little more / much more (1-5)*



**Fig. 4.** The Time Change for Searching COVID-19 Related Information

As shown in Figure 4, most participants (45.3%) spend a little more time to search for information about COVID-19; and some participants (13.3%) spend much more time to search for information about COVID-19. In total, 58.8% of participants have spent more time on information searching than before and only 16% of participants reduced the searching time.

*Question3.2.4: When COVID-19 outbreak occurred in China, how did it impact on your daily life changed? (stage 1)*

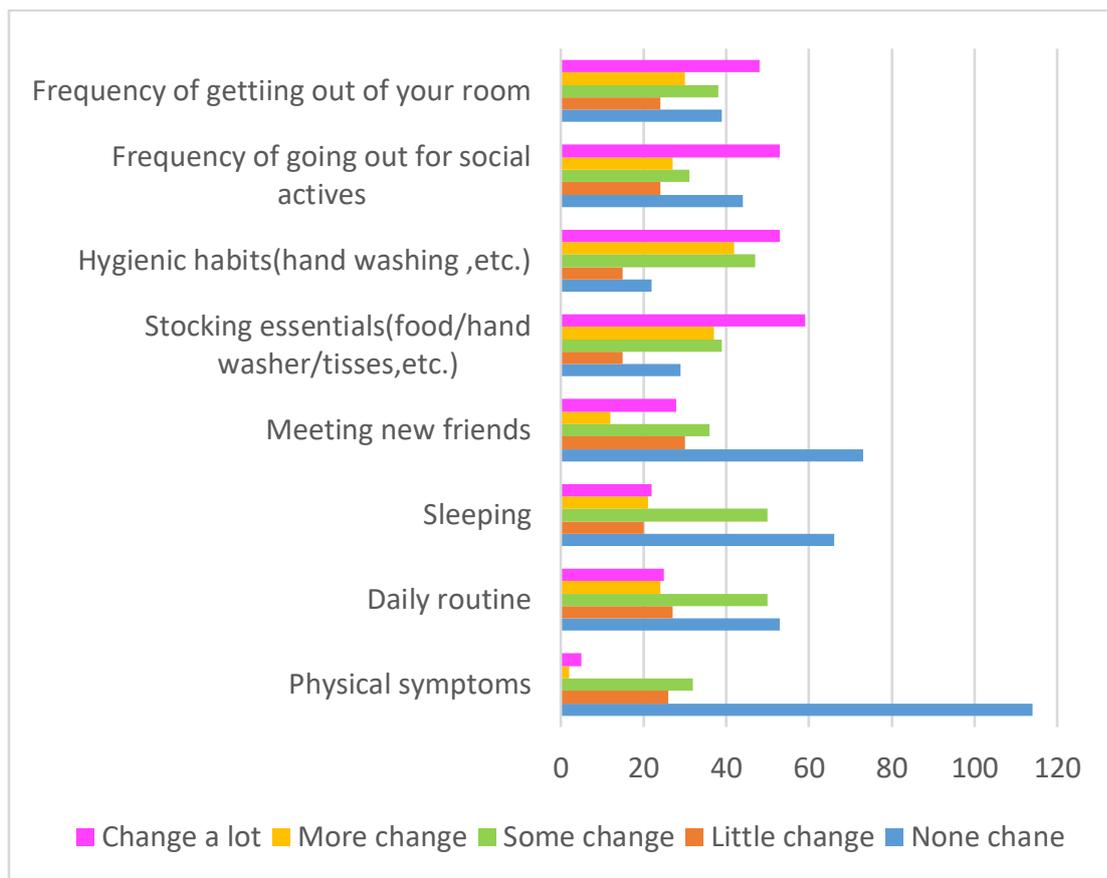
*1None change-2--3-4 -5 Change a lot*

The changes in lifestyle frequency were compiled based on actual observations of international students, but there was no direction in the item, which could lead to misunderstanding by readers. According to the frequency of leaving the room,

participating in social activities and meeting new friends all decreased, while hygienic habits such as washing hands increased, and stockpiling of goods increased based on the rush of supermarkets in the UK. Sleep, daily routines and physical symptoms are difficult to show direction based on the current survey.

As shown in Figure 5, when COVID-19 outbreak started in China, the following four daily behaviors have changed the most from:- "Frequency of leaving your room"(27%, 48), "Frequency of going out for social activities"(30%, 53), "Hygienic habits (handwashing, etc.) "(30%, 53), "Stocking essentials (food/hand washer/tissues, etc.) "(33%, 59). Basically, about 1/3 of the participants chose "change a lot". The four daily behaviors remained the same were" physical Health symptoms (headache/sore throat/stomachache, etc.), " "Meeting new friends", "Sleeping", "Daily routine".

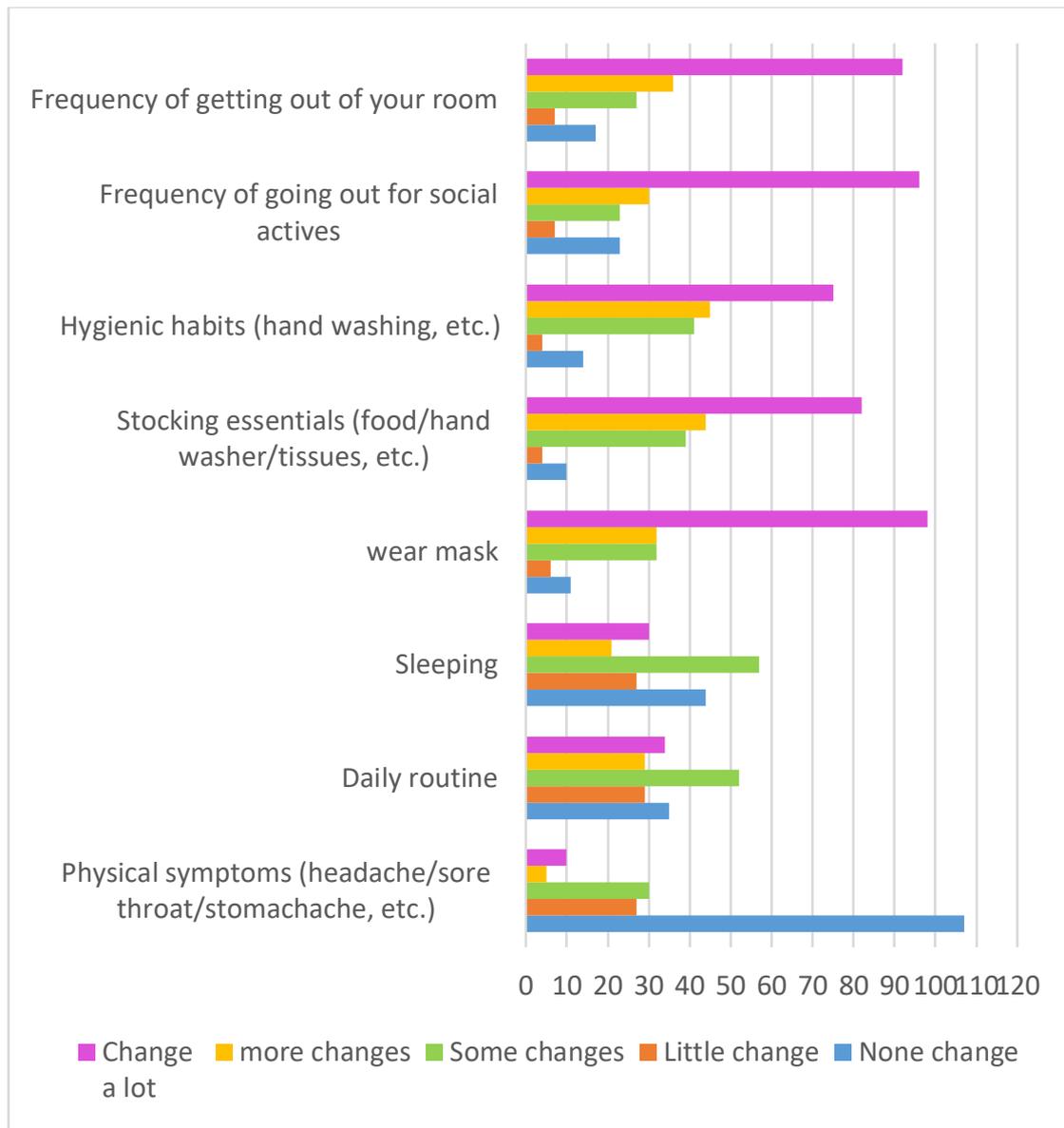
It can be seen that there is a big change in the existing behaviors and habits of interacting with others, while non- interpersonal behaviors such as "physical health symptoms", "sleep", and "daily routines" have not changed much. "Making new friends", on the other hand, maybe because the original life has not changed much, so there is not much change.



**Fig. 5.** Change in Behavior (stage 1)

*Question3.2.5: When COVID-19 outbreak occurred in the UK, how did it change on your daily life? (stage2 ) ( 1None change-2-3-4-5 Change a lot )*

With the outbreak in the UK, this stage added the question asking about wearing a mask, and it was clear that over 55% of the participants changed very much on this option. As shown in Figure 6, when the COVID-19 outbreak happened in the UK, the five behaviors that participants changed most in their daily life were " Wearing mask", " Frequency of going out for social activities", " Frequency of leaving your room", " Stocking essentials (food/hand washer/tissues, etc.) ", "Hygienic habits"; The three behaviors that remained none change in the participants' daily life were " Physical health symptoms (headache/sore throat/stomachache, etc.) ", " Sleeping", and " Daily routine".



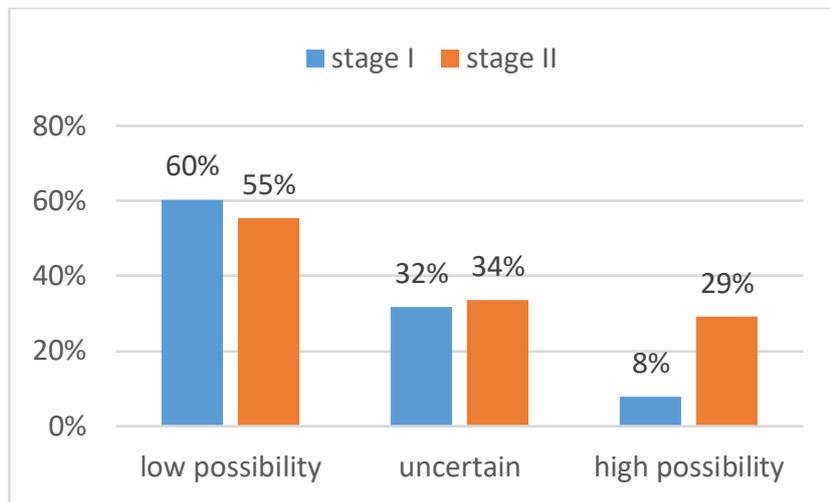
**Fig. 6.** Change in Behavior (stage 2)

### 3.3 Perceived Risk and Support

It is also important how CSRs perceive their own risk/safety at both stages. This section includes 5 questions: the assessment of the likelihood of being diagnosed;

confidence in receiving effective medical assistance when needed; the reasons for that confidence assessment; the support expected; and the support received.

*Question3.3.1: How likely you thought you would be diagnosed as COVID-19? Please indicate the extent to which you agree with each of the following statements using one of the three options.(high possibility; uncertain; high possibility)*

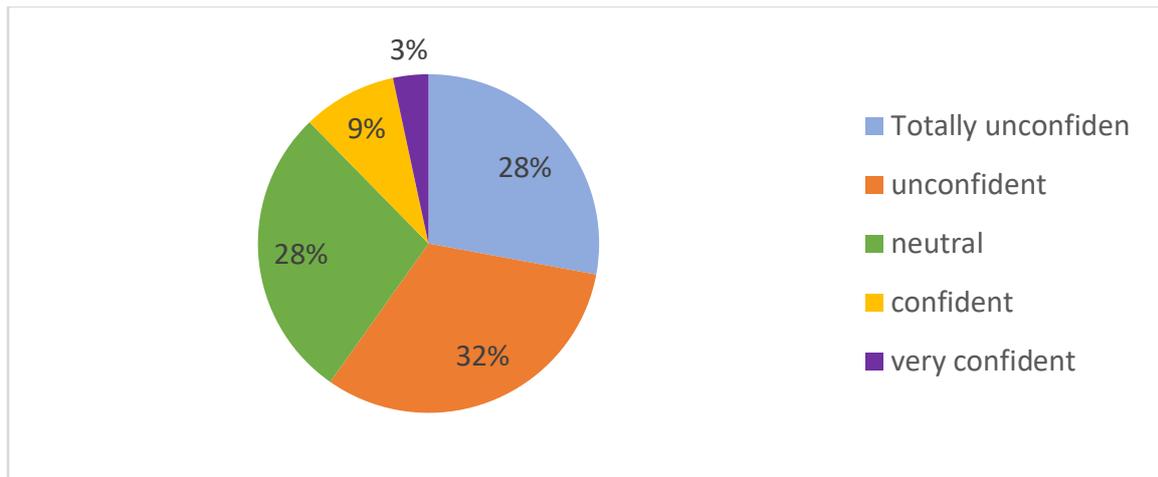


**Fig. 7.** Possibility of Diagnosed as COVID-19

It can be seen that in stage I (Figure 7), the number of participants who think they could have a high risk of infection is relatively small, only 7.82%; mostly (60%) think that they could have a very low risk of infection, etc., and about 32% of participants think that there is uncertainty.

In stage II, the number of participants who thought there was uncertainty about infection increased by 2%, while the proportion of those who thought there was a high risk of infection increased significantly, with the proportion rising by more than 20% from 8% to nearly 1/3(29%) of the participants.

*Question3.3.2: How much confidence you have for receiving effective medical treatment in case you were diagnosed with COVID-19? Please indicate the extent to which you agree with each of the following statements using one of the three options (low confidence; uncertain; high confidence).*

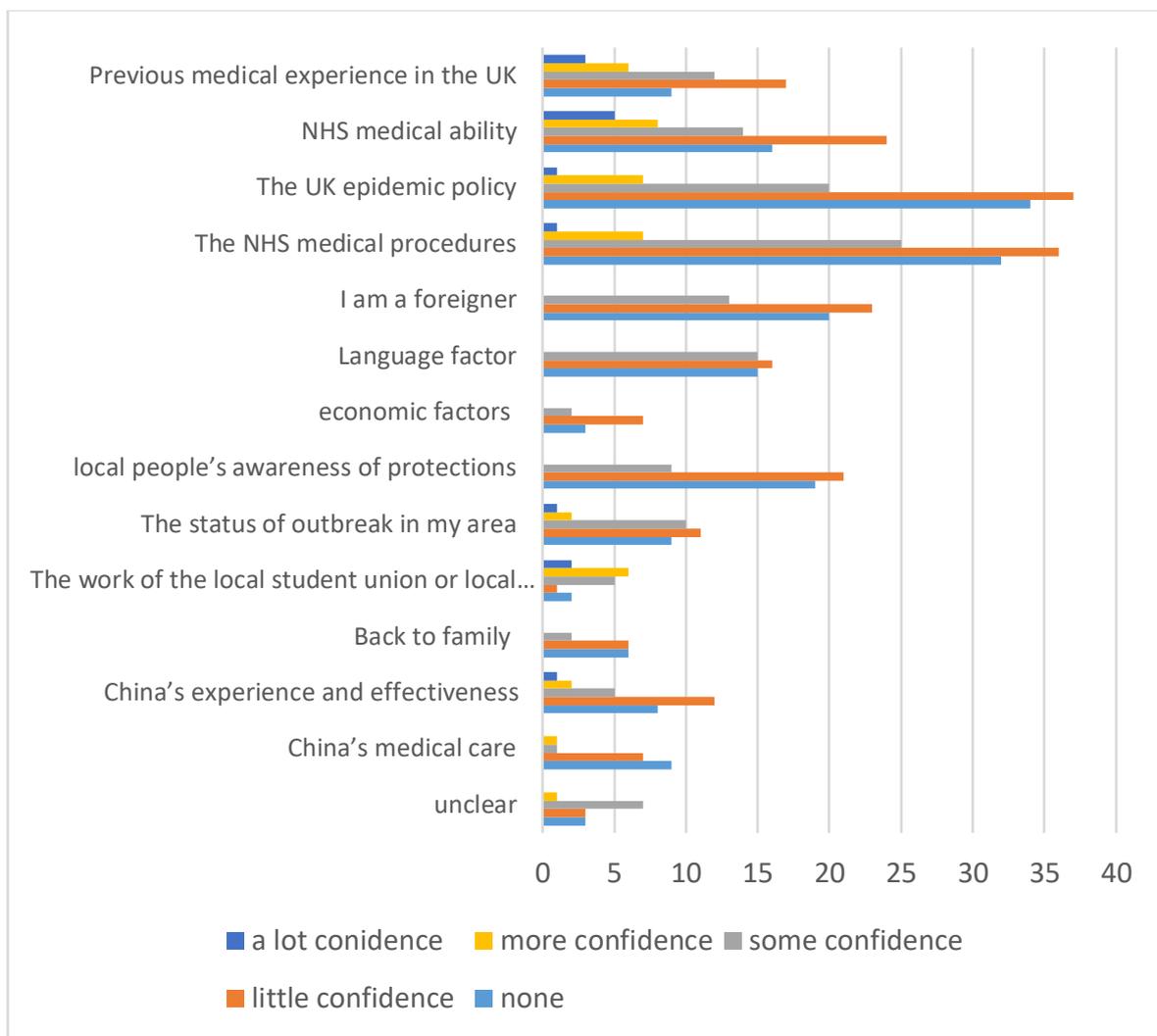


**Fig. 8.** Confidence in Receiving Effective Treatment

Whether or not you have access to effective medical assistance is the most important indicator of international students' sense of security. More than half (60%) of participants have low confidence in this, of which nearly half (28%) participants have no confidence at all, and the rest also have very low confidence. Nearly one-third of participants (28%) are neutral about this, and only about 12% (9%+3%) participants are confident about this.

*Question 3.3.3: Why do you think so? (up to three selections)*

Among the reasons for getting effective support (as shown in Figure 10), participants showed high confidence in the option of "the support from the local (Chinese) student union or consulate"; it reflects the active support they have provided. On the "language factor", although some participants had some confidence, overall no one showed high confidence in this factor; on all other options low confidence was more prominent, reflecting the overall low confidence of participants in the various environmental factors involved.



**Fig. 9.** Summary of Reasons for Confidence in Receiving Effective Treatments

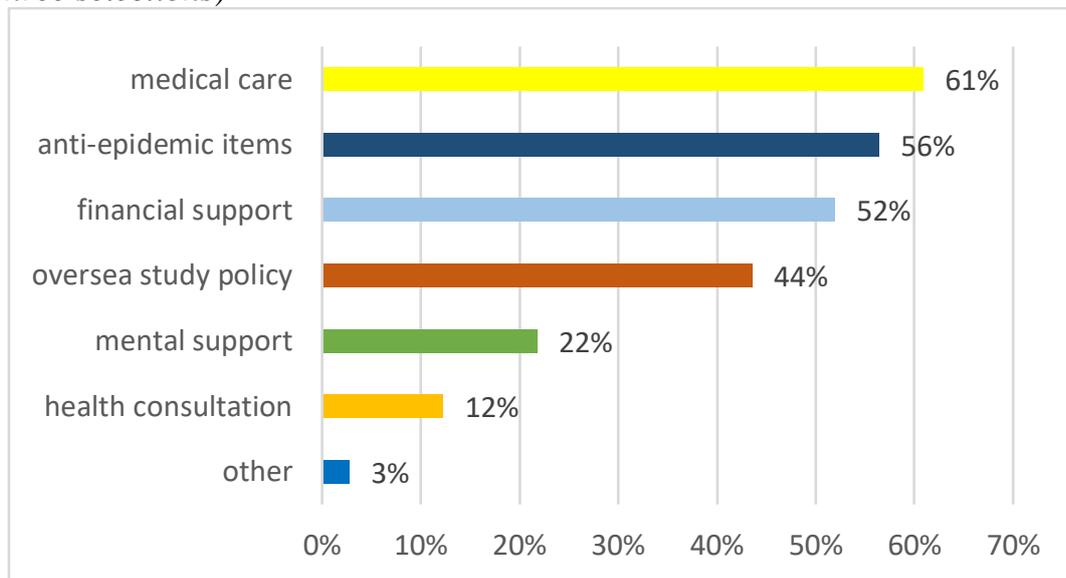
**Table1** chi-square test: the reason for effective treatment

Q45	none	little confidence	some confidence	more confidence	a lot of confidence	chi-square sig
Unclear	3	3	7	1		0.412
<i>China's medical care</i>	9	7	1	1		0.081
<i>China's experience and effectiveness</i>	8	12	5	2	1	0.627
<i>Back to family</i>	6	6	2			0.323
<i>The work of the local student union or local consulate</i>	2	1	5	6	2	0.000*
<i>The status of the outbreak in my area</i>	9	11	10	2	1	0.973
<i>local people's awareness of protections</i>	19	21	9			0.003*

<i>economic factors</i>	3	7	2			0.283
<i>Language factor</i>	15	16	15			0.068
<i>I am a foreigner</i>	20	23	13			0.006*
<i>The NHS medical procedures</i>	32	36	25	7	1	0.093
<i>The UK epidemic policy</i>	34	37	20	7	1	0.006*
<i>NHS medical ability</i>	16	24	14	8	5	0.049*
<i>Previous medical experience in the UK</i>	9	17	12	6	3	0.281

According to the results of the chi-square test (Table1), the participants' opinions were significantly different on these options, such as "the local population's awareness of protection", "I am a foreigner", "the UK's pandemic prevention policy", "NHS pandemic prevention and medical care level" and "policy of the local federation or consulate".

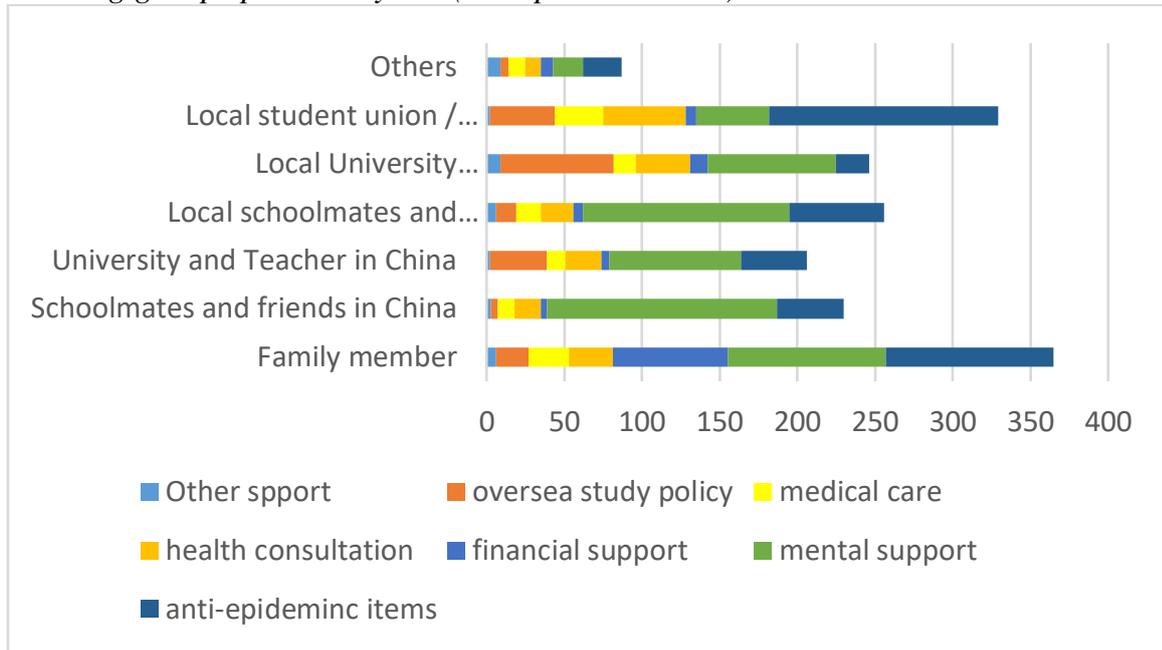
*Question3.3.4: At your area, which is your most desired anti-epidemic help? (up to three selections)*



**Fig. 10.** Summary of Desired Support

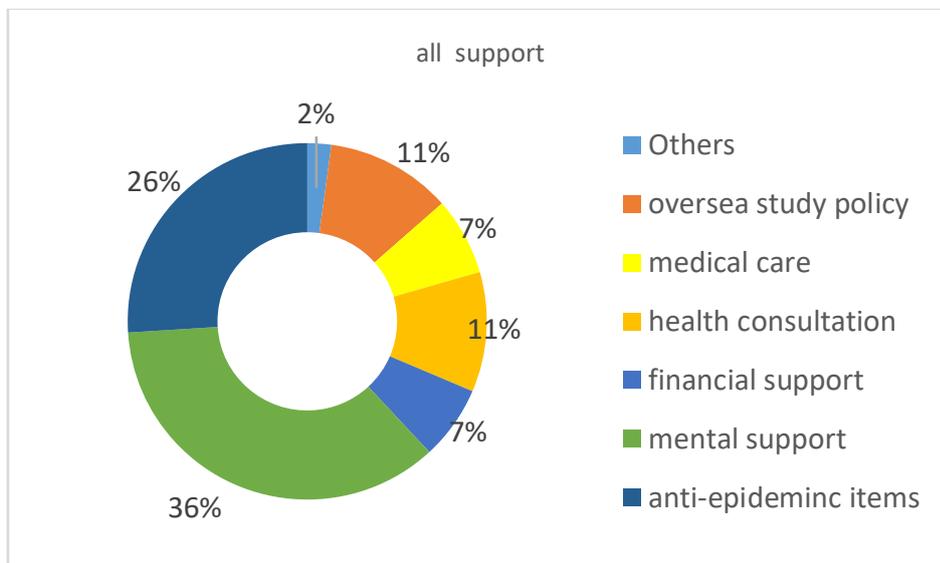
Over 50% of the demanding options are "medical care", "anti-pandemic items" and "financial support". It means that at this stage, material needs related to pandemic protection are most important to the participants. Subsequently, the options for international students were "overseas study policy", "mental support", and "health counseling for medical care and pandemic prevention". About 3% of the participants chose "Other" and all of them indicate "Airfare", "Airline tickets", and such on.

*Question3.3.5: When COVID-19 outbreak in the UK, what's kind of supports did the following groups provided you? (multiple selections)*



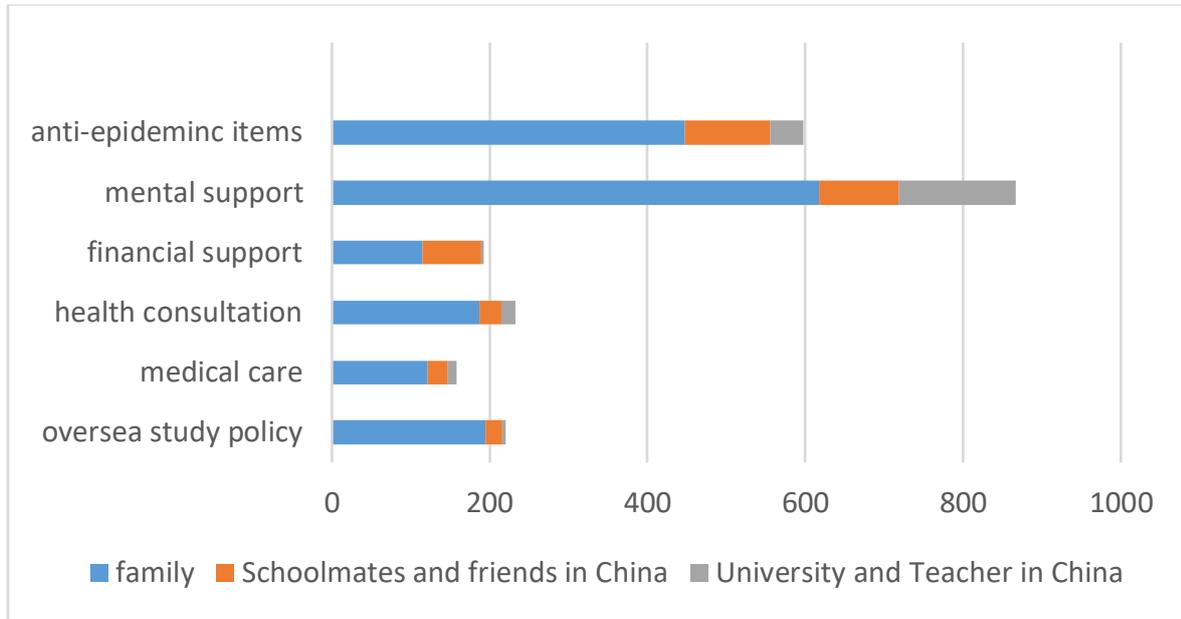
**Fig. 11.** Overview of Support Received

The "medical care," which ranks first in Figure 10, was provided relatively little by seven support groups. The "anti-pandemic items" and "mental support" were both provided a lot by the seven support groups.



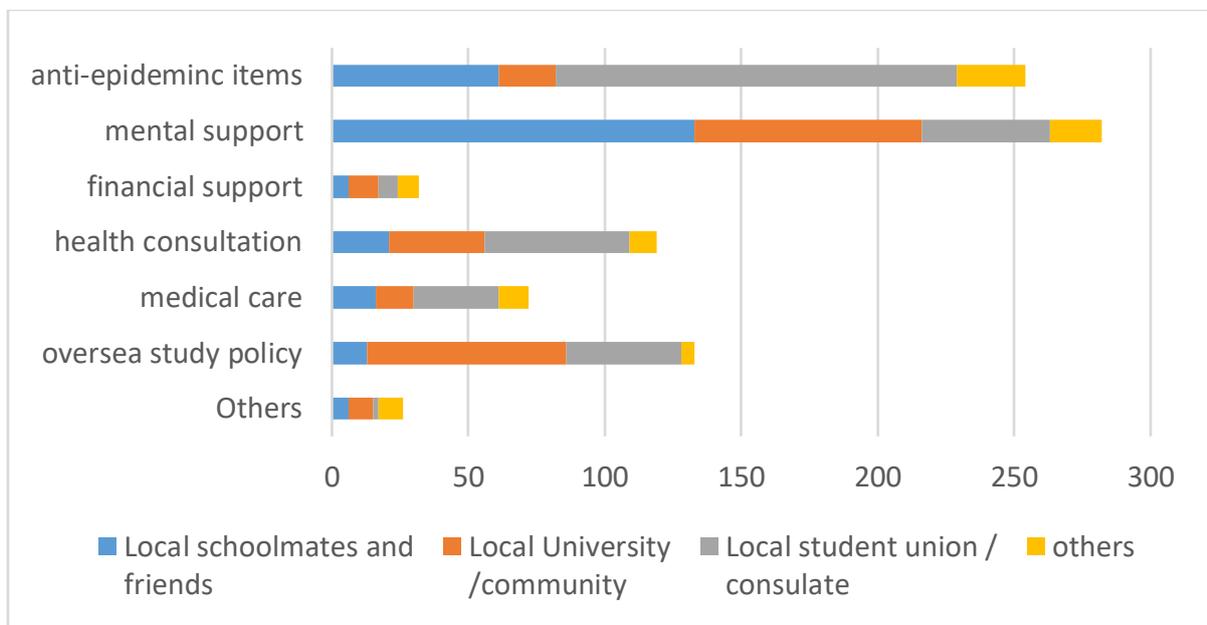
**Fig. 12.** Percentage of Support Received

Participants have received the most support in the form of "mental support", followed by "anti-pandemic items", "health counseling for medical care and pandemic prevention" and "overseas study policy", "financial support" and "medical care", and finally, "other elements of uncertainty".



**Fig. 13.** Received Support from China

The support received from families is mainly "anti-pandemic material", "mental support" and "financial support" for vaccination. The support from Chinese classmates and friends is mainly "mental support", "anti-pandemic material" and "health consultation" for anti-pandemic information. Chinese universities and teachers mainly provided support in forms of "mental support", "anti-pandemic material" and "overseas study policy".



**Fig. 14.** Received Support from the UK

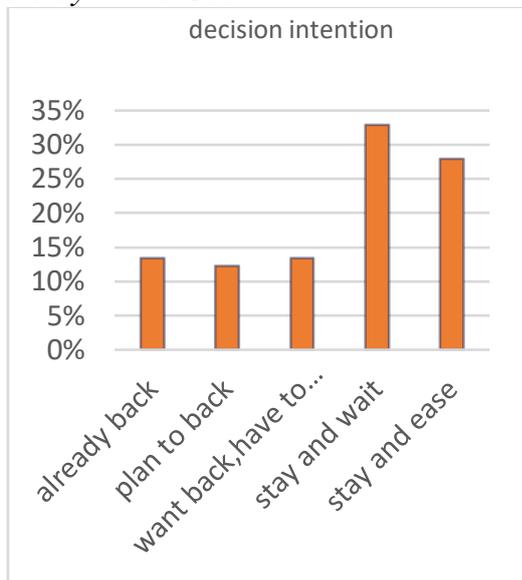
Local students/friends mainly provided "mental support" and "anti-pandemic material" support to CSRs. Local universities/communities provided "mental support" as well as support on "overseas study policy". Local Chinese student

unions/consulates mainly provide "anti-pandemic material ", "mental support", "health consultation" for pandemic-prevention information and "oversea study policy".

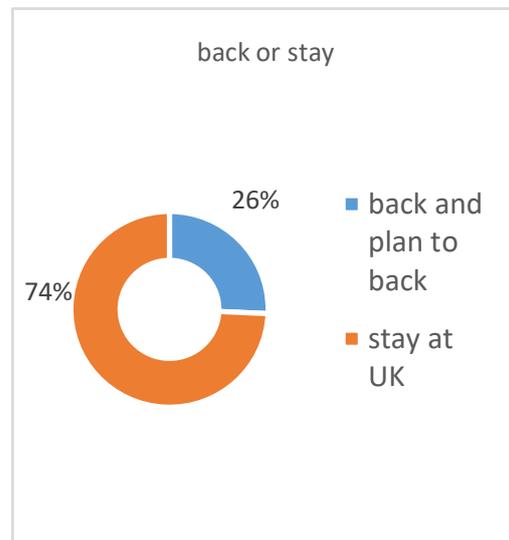
### 3.4 CSRs’ Decision Making and Reasons Behind

This part includes 2 questions: the CSRs' decision and the reasons to support decision making.

*Question3.4.1: After the COVID-19outbreak occurred in the UK, your decision is? Plan to go back; Already go back; Stay at ease, believing the epidemic will pass; Temporally stay in the UK, acting depends on the situation; Want to go back but have to stay in the UK.*



**Fig. 15.** Distribution of Five Decisions

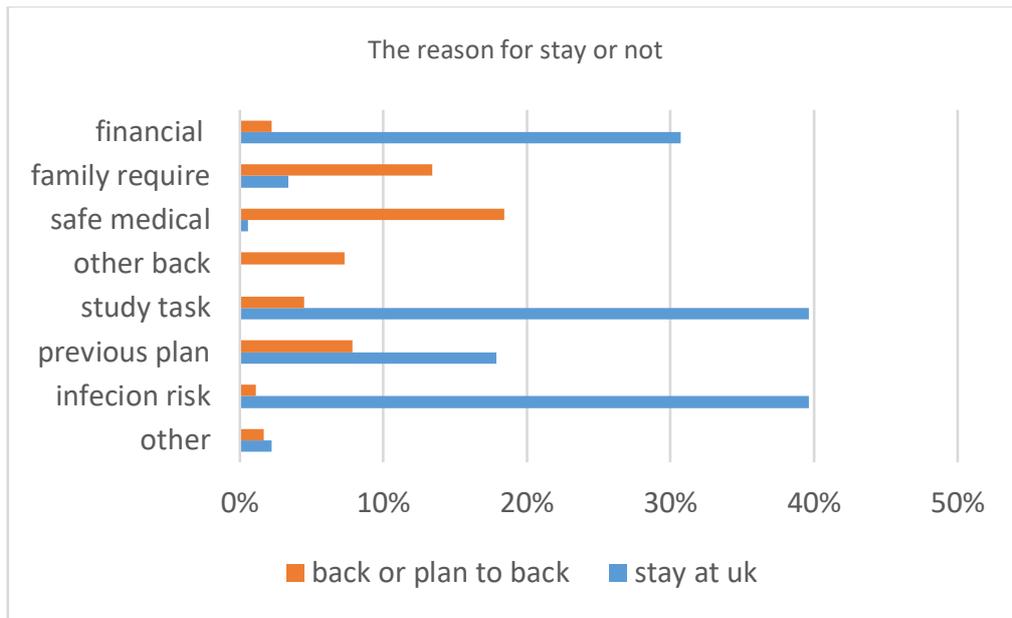


**Fig. 16.** Final Decision

In terms of specific choices, “wait-and-see” is the most popular choice, at 33%. This was followed by those who thought the pandemic would pass and waited with confidence (28%). The rest participants are those who have gone back, intend to go back, and want to go back but have to stay(Figure 15).

If we distinguish between suburbs based on the current intention to go back or not, the current situation is as shown above (Figure 16). Only about 1/4 of participants have returned or are planning to return, and nearly 3/4 of participants are still staying in the UK whether they want to or not.

*Question3.4.2: The reasons you made the decision? (up to 3 selections)*



**Fig. 17.** Reasons Behind the Decision Making

Table2 Chi-square Test: reasons behind the decision

<i>Choice reason</i>	<i>stay</i>	<i>back</i>	<i>chi-square sig</i>
<i>other</i>	4	3	0.289
<i>other back</i>	0	13	0.000*
<i>*family require</i>	6	24	0.000*
<i>same as previous</i>	32	14	0.394
<i>financial issue</i>	55	4	0.000*
<i>infection risk on the way</i>	71	2	0.000*
<i>study task</i>	71	8	0.000*
<i>safe medical consideration</i>	81	33	0.188
<b>sum</b>	<b>320</b>	<b>101</b>	

We can see from the results of Chi-square test (Table2) that there are significant differences between the two types of choices for participants to go back or not. The participants who choose "stay" are mainly based on the economic factor, the risk of cross-infection on the road, and the task of studying afterward; these are mostly objective factors. The participants who chose "back" were mainly based on interpersonal factors such as "the request of family members" and "all other international students around have gone back".

## 4 Discussion

The outbreak of the pandemic in China has made CSRs make significant changes in their lifestyles and behaviors, such as taking more protective measures than local people. When the outbreak had not yet broken out and the UK government was not

enforcing home quarantine, CSRs had already begun to make significant changes in "wearing masks", "leaving their rooms less often", "having less contact outside", and "changing their hygiene habits".

Behavioral changes were in the area of multiple interpersonal interactions and little change in self-related aspects such as sleep, routines, and physical and mental symptoms.

CSRs have multiple factors influencing their assessment of their own well-being in the event of a foreign pandemic. In general, with the outbreak of pandemics in the UK, CSRs think that they are more likely to be infected with pandemics, and they are less confident that they will be able to get effective medical treatment, based on their awareness of the UK government's "Buddhist anti-pandemic" measures and their status as foreigners. More than half (60%) of participants have low confidence in this, and only about 12% (9%+3%) participants are more confident about it.

CSRs in crisis need medical assistance support most, followed by support such as anti-pandemic items and financial support, and then "soft support" such as overseas study policy, and mental support. CSRs receive the most support in the form of anti-pandemic items and mental support, and almost all of these resources are provided at both home and abroad. Financial support is mainly provided by families. The work of local student federations/consulates (distributing anti-pandemic materials and providing information on pandemic prevention) has had a significant effect. It is almost impossible for CSRs to secure the content of the overall national regulation, such as the UK's pandemic prevention policy, medical and ambulance protection, and China's aviation policy for pandemic prevention requirements.

As of July 5, there were significant differences between the two final behavioral decision choices of CSRs to "go back to China" and "stay in the UK". The three-quarters of participants chose to stay in the UK, which included the proactive choices of "want back, have to stay", "wait and see how the pandemic develops" and those who believe the pandemic will pass (stay and ease). The participants who chose to "stay in the UK" were influenced by a variety of factors, such as academic needs, financial pressure (flight restrictions), risk of cross-contamination on the road, and their original plans, etc. The participants who chose to "go back" were influenced by interpersonal factors, such as "family" and "people around them have left".

In conclusion, the majority (3/4) of participants stayed in the UK although they experienced two pandemic shocks in the UK and did not think they have enough safety in the pandemic. On the one hand, it is based on the security of pandemic prevention materials, psychological care, and study abroad related policies at home and abroad, and on the other hand, it is also based on the consideration of future development. In terms of the "fight or flight" mode of crisis response, 1/4 chose to run away and 3/4 chose to fight.

## 5 Limitation of this study

This study had some limitations. Firstly, our sample was small. Furthermore, limitations of the current study include the use of a self-designed questionnaire and

the reliance upon self-reporting in the midst of the constantly changing prevalence of COVID-19. In addition, this study focused on changes in CSRs' behavioral change, risk of COVID-19 infection, confidence in receiving effective assistance, and final choice to return home or not, and did not measure their psychological statuses, such as anxiety and depression levels. Future studies could add this component and be comparable to most current studies in China (health care workers, university students, and other groups).

## 6 Conclusion

The conclusion of this paper is: Despite the fact that CSRs has a significantly higher assessment of infection risk in the second stage (34% uncertain, 29% high risk), and perceive themselves as less likely and less confident to be rescued (60% low confidence), 3/4 of the participants still chose to stay in the UK based on a number of realistic and subjective factors (subsequent studies, flight tickets, risk of infection on the road, etc.). As a result, our findings are based on cross-sectional data from local convenient samples. However, in reality, from May to July, there have been some changes in the UK's epidemic prevention and control policies and public responses. The future study could explore the changes in the mindset of CSRs and the influence of relevant factors in this changing environment of epidemic prevention and control in the UK.

## 7 Reference:

1. WORLD HEALTH ORGANIZATION, WHO director-general's opening remarks at the media briefing on COVID-19 11 March 2020.  
<https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
2. Anti-Asian crimes up 21% in UK during coronavirus crisis. The Guardian, May 13, 2020.
3. Association of British Universities, International Student Statistics in UK 2020.  
<https://www.studying-in-uk.org/international-student-statistics-in-uk/>
4. Hongyu Zhao, Xiaoyi He, Guanhua Fan, Liping Li, Qingjun Huang, Qinming Qiu, Zhewei Kang, Taifeng Du, Ling Han, Lei Ding, Haiyun Xu: COVID-19 infection outbreak increases anxiety level of general public in China: involved mechanisms and influencing factors. *Journal of Affective Disorders*, volume 276:446-452 (2020).  
<https://doi.org/10.1016/j.jad.2020.07.085>
5. Wei Hu, Li Su, Juan Qiao, Jing Zhu, Yi Zhou: Countrywide quarantine only mildly increased anxiety level during COVID-19 outbreak in China. Preprint, (2020).  
<https://doi.org/10.1101/2020.04.01.20041186>
6. ABCP (Association of British Chinese Professors): The Impact of COVID-19 on Chinese Applicants' Decisions to Pursue a University Degree in the UK in 2020-21 Academic Year, Results from an Online Survey, (2020). <http://abcp.org.uk/covid-19-2020-survey-report/>

# A Close Look on the *Corona* Impact on Surveillance Radar Channel Loads

Roman Raekow<sup>1,2</sup>, Michael Kuhn<sup>2</sup>, and Bernd-Ludwig Wenning<sup>3</sup>

<sup>1</sup> Deutsche Flugsicherung GmbH, Langen, Germany [roman.raekow@dfs.de](mailto:roman.raekow@dfs.de)

<sup>2</sup> h\_da Hochschule Darmstadt, Germany [michael.kuhn@h-da.de](mailto:michael.kuhn@h-da.de)

<sup>3</sup> CIT Cork Institute of Technology, Cork, Ireland

[BerndLudwig.Wenning@cit.ie](mailto:BerndLudwig.Wenning@cit.ie)

**Abstract.** Civil aviation surveillance is carried out on two radio channels and has seen a growing demand in the last decades. Not only that more and more applications have been added to the channels but the number of flights had been growing constantly. This was true until the corona crisis almost brought the air traffic over Europe to a standstill and by mid of April 2020 the number of flights dropped to around 12% compared to the year before [3]. In this paper the consequences for the civil aviation surveillance channels and the success rates for telegrams on those channels are discussed.

**Keywords:** Surveillance Radar · Mode-S · ADS-B · Civil Aviation · COVID-19 · Corona

## 1 Background

The world's first air traffic control tower on Croydon Airport is celebrating its 100<sup>th</sup> anniversary this year. From the very beginning, air traffic controllers needed to be aware of the positions of the surrounding aircraft in order to separate them safely. While the air traffic controller of Croydon Airport queried their pilots to state their positions, modern surveillance is nowadays able to retrieve a wide range of information from aircraft and can therefore deliver an accurate air situation with high update rates. This technical improvement allows more aircraft to manoeuvre in a dense airspace while maintaining a high level of safety. Radar stations send out interrogations on a 1030 MHz channel and receive replies from the aircraft that are sent back on a 1090 MHz channel. The most common setup is to have a grid of secondary surveillance radar stations that query surrounding aircraft for their altitude, their ID and further necessary information needed by the air traffic control. The position is calculated by the round-trip time of the signal and the angle of the radar-station under which the interrogation was carried out. Aircraft equipped with a transponder receive these requests and answer on the 1090 MHz channel with the queried data. Besides that, there are other applications that are using these channels, for example the Traffic Collision Avoidance System (TCAS) that is used by aircraft to query the surrounding planes in order to be aware of possible unintended approaches.

Another common application is Automatic Dependent Surveillance Broadcast (ADS-B) that is using the aircraft transponder to send out information about the aircraft in a pseudo-spontaneous manner without an interrogation on the 1030 MHz channel. ADS-B has seen a growing popularity in the last years and will soon be mandatory in the European airspace.

All these applications are carried out in parallel on the two surveillance radar channels that are used in civil aviation. Over the last years the need to measure the load of these channels has become more and more important for the air navigation service providers (ANSP) around the world. [10, 2, 7]

## 1.1 Channels

The communication is carried out on two channels, 1030 MHz and 1090 MHz, that each have their dedicated purpose:

**The 1030 MHz channel** is used for interrogations. These interrogations can be one of four possible protocols, that either query all surrounding aircraft or interrogate single aircraft for specific information. These queries are either sent out by secondary surveillance radars or TCAS equipped aircraft [1]. However, this report focuses on the the 1090 MHz channel only.

**The 1090 MHz channel** is used for the replies to interrogations sent through the 1030 MHz channel. It is usually more crowded than the 1030 MHz channel, since a single interrogation can result in multiple replies. Additionally, some applications like ADS-B only use the 1090 MHz channel. The information is transmitted using three types of telegrams:

*Mode A/C Reply* is the reply to an *A* or a *C* interrogation by a radar-station, respectively. It can either return the altitude (Mode-C) or a flight-id of the responding aircraft (Mode-A). It encodes a four-digit octal number using 12 Bit. Together with the necessary framing pulses it has a duration of  $20.3 \mu s$  on the channel.

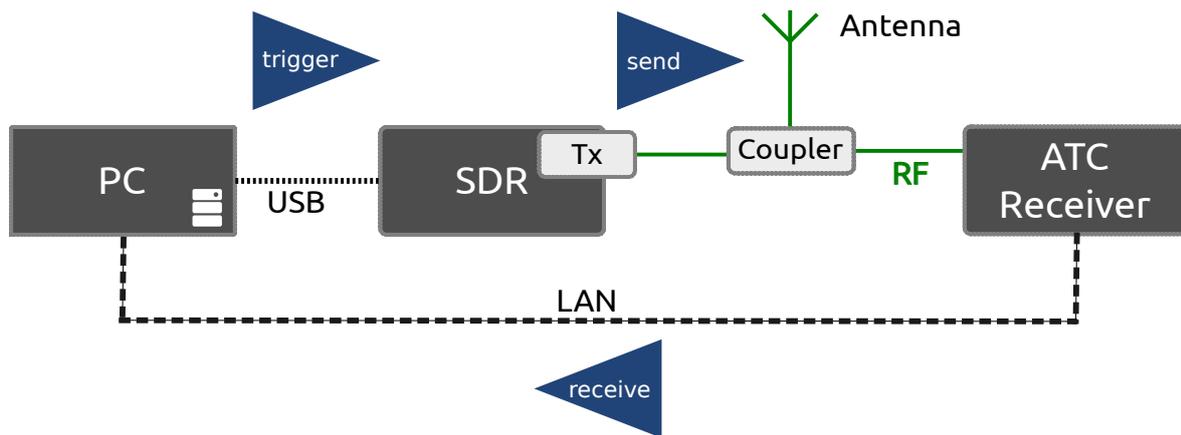
*Mode-S Short Reply* is one of the possible reply types for Mode-S interrogations. This telegram is also used by TCAS and transports small portions of data. It consists of a preamble, a data field and a checksum field. The short reply lasts  $64 \mu s$  on the channel and has a payload of 56 Bit.

*Mode-S Long Reply* is the second type of Mode-S replies. This Extended Length Message (ELM) is used to retrieve more detailed data from the aircraft. Its structure is the same as Mode-S Short Replies but it has a longer payload of 112 Bit and has a duration  $120 \mu s$  on the channel.

To measure on that channel, a measurement system had been created that uses affordable hardware and is able to monitor the receivable traffic on a time span of various days [12]. This system does now allow a comparisons of the channel loads, prior and during the corona Crisis

## 2 Measurement System and Setup

The system setup as described in our previous work [12] consists of a standard receiver that is used for multiple applications by ANSP. A software defined radio (SDR) sends well-known telegrams to the receiver by coupling the telegrams into the radio frequency (RF) channel on the input of the receiver.

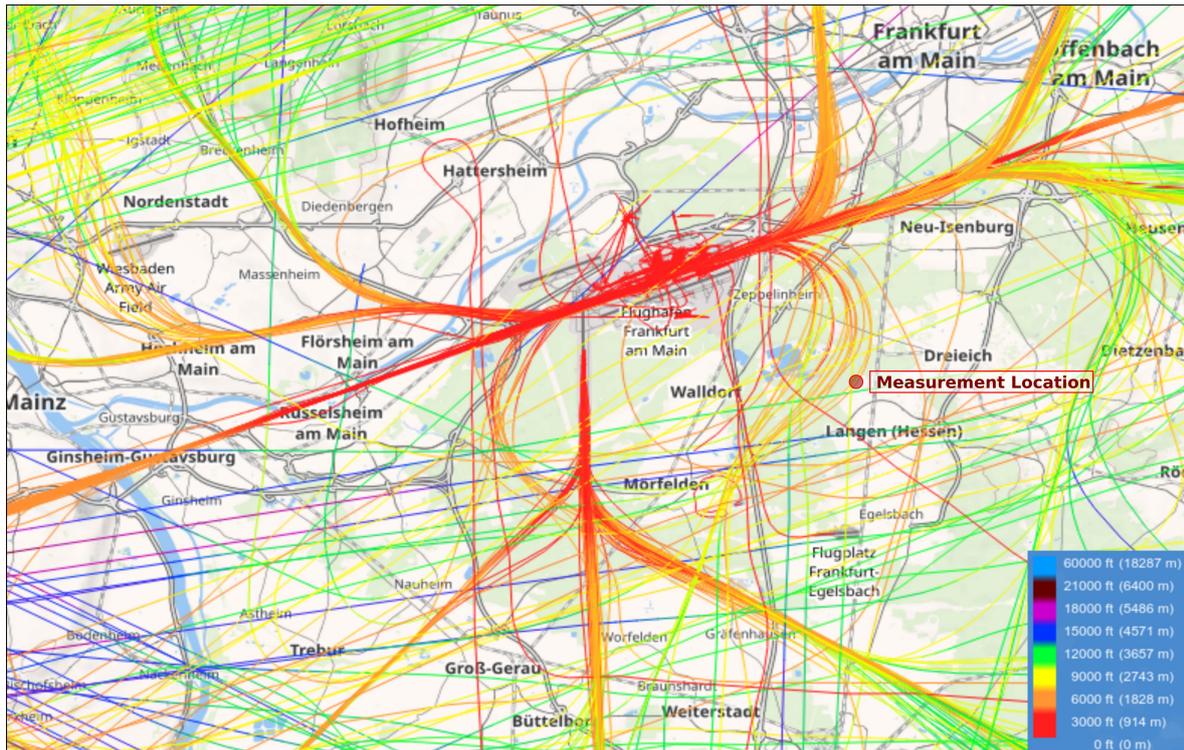


**Fig. 1.** The principle of the measurement system

A standard PC is used to trigger the SDR and receive the data from the mixture of the test telegrams and the *real* RF environment as shown in figure 1. This data is used to create a database of the success rates of the test telegrams in the current environment. The results are recorded in a continuous manner containing the rates of successfully decoded telegrams and the environment under which the data has been recorded. This approach has already been described in previous works [11, 15] but the advantage of using a SDR is to be able to adaptively change the constellation of telegram-types, rates and levels with respect to the current RF environment and therefore be able to gain knowledge about lesser known telegram constellations when environmental changes on the channel are detected.

This setup is under continuous improvement and will be expanded throughout the work on this research. Currently a team of students from HOCHSCHULE DARMSTADT (h\_da) is working on an even more cost-efficient approach which shall be achieved by the deployment of open-source tools in combination with cheaper hardware (see Section 4).

Figure 2 indicates the measurement location in Germany in the area of Frankfurt International Airport (EDDF). The system was using a 90° segment antenna heading north-west. It is one of the most crowded areas in the German airspace and thus features high channel loads.



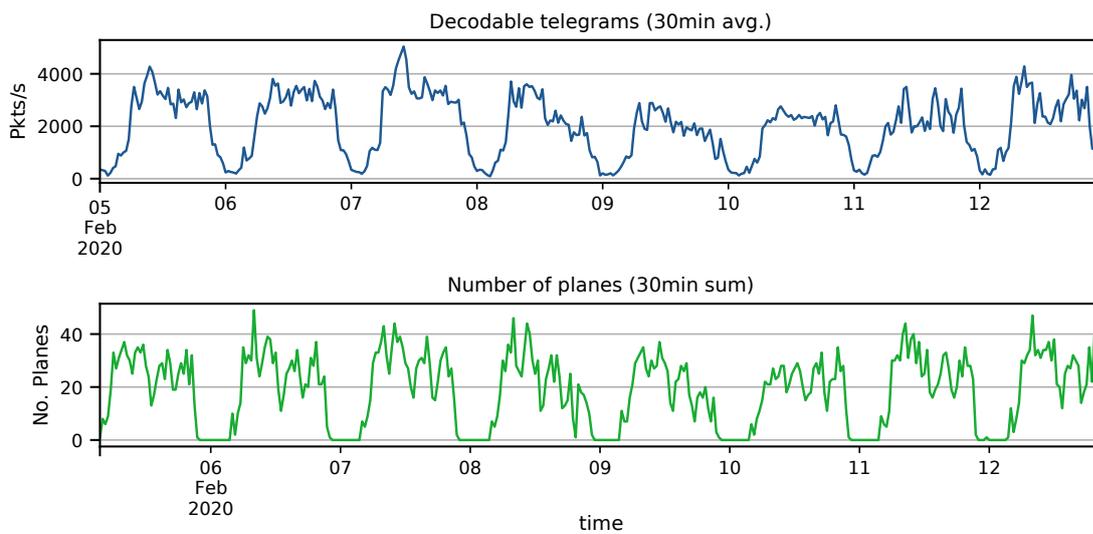
**Fig. 2.** An indication of the measurement location, and all flights between 14<sup>th</sup> and 21<sup>st</sup> April in the area of Frankfurt.

### 3 Results

The results shown in this paper have been recorded in two time windows. The first measurement session has been carried out from the 5<sup>th</sup> to the 12<sup>th</sup> of February 2020. The second session was carried out from the 15<sup>th</sup> to the 25<sup>th</sup> of April 2020. During this period the number of telegrams and the success rate have been recorded together with a wide range of additional information like the number of planes in view and the types of telegrams that were seen in each time frame.

#### 3.1 Time Variance

The number of receivable telegrams varies very strongly throughout each day and shows strong impacts on the channel. There is a clear recurring behaviour in Figure 3 over each day. The communication increases as soon as Frankfurt airport opens. At the beginning of each day there is a small peak that is followed by a little drop. That is when the intercontinental flights arrive in Frankfurt in the early morning. The second and largest peak of the day is around 6am to 8am local time where most of the departures are taking place. There is also a weekly dependency on the traffic, on Friday the 7<sup>th</sup> of February the "departure peak" is the largest in the entire record as this is usually one of the busiest days of the week.



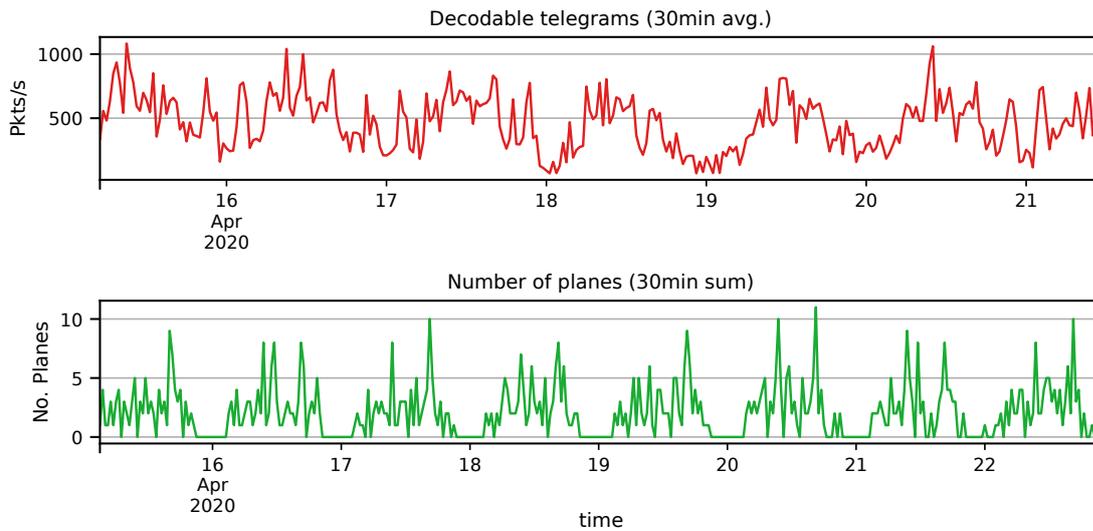
**Fig. 3.** The number of telegrams and planes over several days in February

### 3.2 Special Events and the Corona-Impact

In Figure 3 one can also see that special events do have a direct impact on the channel utilisation. There is a sharp drop in the traffic from noon on the 8<sup>th</sup> of February until the 11<sup>th</sup>. This drop was caused by storm Ciara that hit Germany during that time and led to a large number of cancelled flights. On the 12<sup>th</sup> of February the traffic is back to a normal level causing a equally normal amount of packets on the channel. This is one of the strongest declines since the outbreak of the Eyjafjallajökull in April 2010 [4, 8] where for a period of 4 days almost 100% of the flights were cancelled. However, this drop in flights was very soon surpassed by the lockdown due to the corona crisis [5].

Figure 4 shows the traffic during a week in April 2020, where most of the flights were cancelled due to the shutdowns associated with the corona virus. The figure displays the same axis as Figure 3 but with data recorded in April 2020. It shows that almost all of the characteristics have changed. Not only has the total amount of packets per second dropped significantly, but also the day and night periods differ only very slightly and due to the low number of aircraft the graph is very unstable. As shown in Table 1 the decrease of communication on the 1090 MHz channel is very obvious. While the minimum of packets per second has not changed significantly due to low numbers of flights in the night during the observation period in February, the maximum in April reaches only around 32% compared to February. The mean over the entire inspection period drops to around 23% compared to *pre-corona times*. The distribution of the telegrams has also changed completely as there is only a very small difference between daytime and nightly traffic which can be seen in the low standard deviation (std. dev.).

One can see the strong relationship between the actual number of flights and the telegrams that were received on that channel. The recording of our data is



**Fig. 4.** The number of telegrams and planes over several days in April

supported by the observations from the OPENSKY NETWORK, a research project that enables full exploration on MODE-S and ADS-B data [13].

The data from the network is gained from a distributed set of sensors that listen on the 1090 MHz channel and feed decodable telegrams to a central server. Due to ADS-B the network is able to monitor a certain amount of flights and track their movement in the airspace. Currently, more than 60% of all flights over Europe are equipped with ADS-B [14]. It will become mandatory for all aircraft with a maximum takeoff mass (MTOM) greater than 5700 kg or a maximum airspeed capability greater than 250 knots by end of October 2025 [6]. Overall, this already gives a valued estimate from a different source to prove the measurements from the described setup. The figures show a strong correlation of aircraft identified by the OPENSKY NETWORK from or to Frankfurt international airport (EDDF) to the number of telegrams detected by the measurement setup. While the number of telegrams increases by a factor of 90 between day and night, there are approximately 80 times more flights during the day and none during the night. Eventually, this may be caused by other aircraft not heading to nor departing from EDDF. However, this correlation needs to be investigated, as it could also imply a saturation on the channel not allowing to decode all telegrams during the day peak anymore.

**Table 1.** Comparison of the datasets

	February	April	Relative
min	54 Pkt/s	62 Pkt/s	114%
max	4863 Pkt/s	1575 Pkt/s	32%
<b>mean</b>	2223 Pkt/s	525 Pkt/s	23%
std. dev.	1262 Pkt/s	258 Pkt/s	20%

The corona lockdown gives the opportunity to gain new insights into the surveillance radar channels as only the number of planes has changed in this crisis whereas other parameters like the number of radar stations and interrogation pattern kept unchanged. The *corona dataset* from April allows a comparison of the impact of the number of aircraft in an otherwise unchanged environment.

The contribution of the planes to the decodable traffic on the channel can be calculated by the number of received telegrams divided by the number of planes which gives the rate *packets per aircraft*.

In February the maximum amount of packets per second ( $f_{pkt}$ ) as an average during one hour was  $f_{pkt} = 4863/s$ . In this hour 84 planes were heading to or departing from Frankfurt ( $f_{FRA} = 84/h$ ), while in April the maximum average of packets per second was  $f_{pkt} = 1575/s$  with  $f_{FRA} = 16/h$ . So, in February the system was able to detect  $\approx 2.08 \cdot 10^5 Pkt/Plane$  while in April a maximum of  $\approx 3.54 \cdot 10^5 Pkts/Plane$  where received.

In the high traffic environment the contribution to the number of decodable telegrams per plane is much lower than in the low traffic environment. However, this stands in strong contradiction to the fact that all planes are interrogated in the same manner. Additionally, the TCAS system that queries surrounding planes for their positions should add an increasing amount of telegrams per planes the denser the airspace is being used.

The number of packets per second is not growing linearly with an increasing number of aircraft on the surveillance radar channels.

### 3.3 Channel Load

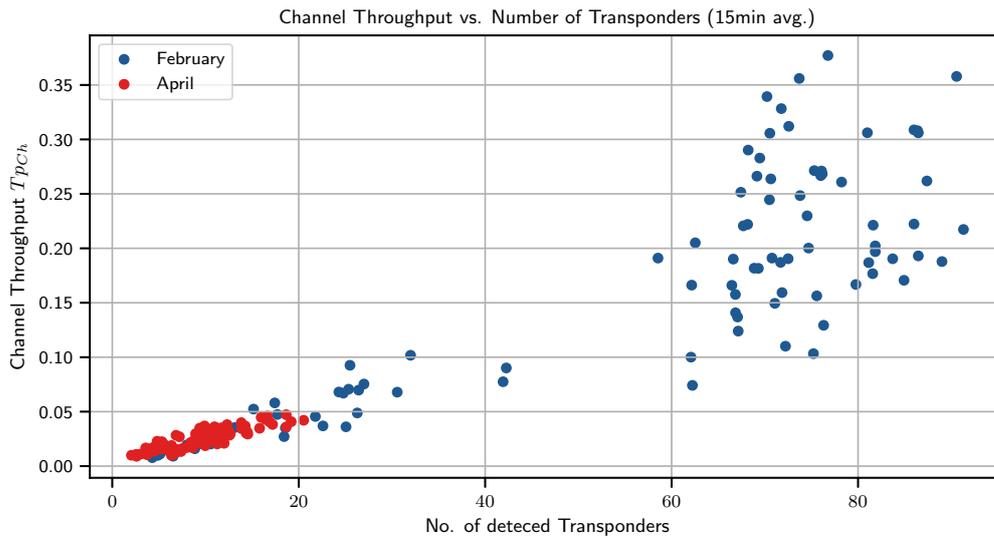
In order to calculate the impact of traffic on the 1090 MHz channel it is necessary to inspect the decoding probability during changing channel loads. However, the system is only able to see the channel throughput ( $Tp_{Ch}$ ) on the receiver side and does not know the *real* amount of telegrams that were transmitted on the channel.

The channel throughput equates to

$$Tp_{Ch} = \frac{1}{t_{\Delta}} \cdot (N_{AC} \cdot T_{AC} + N_{SS} \cdot T_{SS} + N_{SL} \cdot T_{SL}) \tag{1}$$

This rate is calculated by the number of decodeable telegrams  $N_x$  of each type of telegram  $x \in \{Mode A/C, Mode S Short, Mode S Long\}$  times the duration these telegrams occupy the channel  $T_x$  (see also section 1.1) divided by the time of observation  $t_{\Delta}$ . A value of  $Tp_{Ch} = 1$  would refer to a 100% channel usage where the channel is permanently occupied.

Figure 5 shows the channel throughput for changing numbers of aircraft that were in sight during the 6<sup>th</sup> of February and the 16<sup>th</sup> of April. Each point represents a 15 minute time interval throughout the day. With growing traffic, the channel usage is increasing very fast and varies between 0.10 and 0.35 for the majority of the time. During that time 50 to 90 aircraft were sight for the February data set. In this region it seems like there is already some saturation on

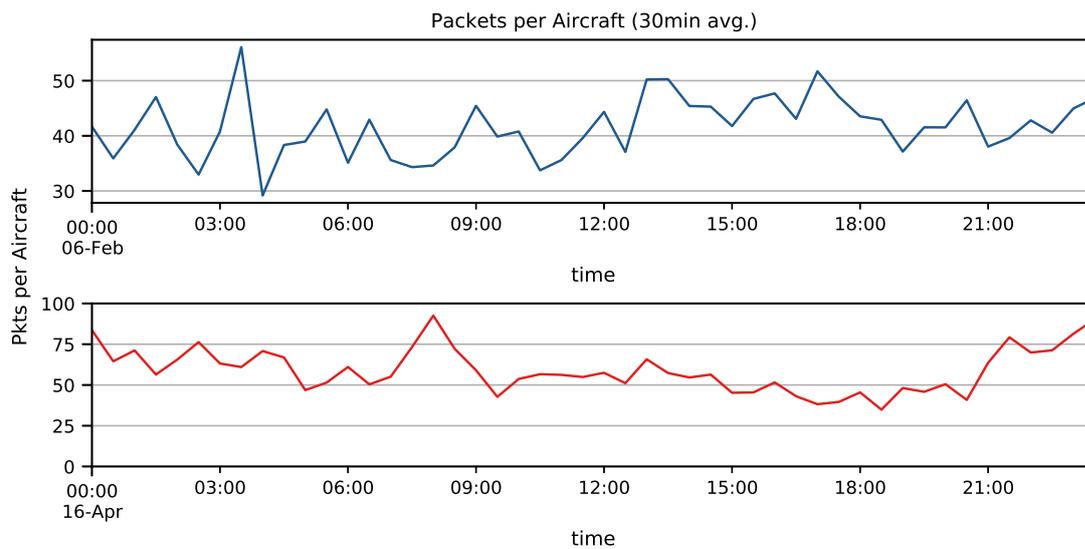


**Fig. 5.** The channel throughput on two days in February and April

the channel as the  $Tp_{Ch}$  increases faster than the number of planes. This topic needs further inspection in the future. However, this implies that the channel is already near or beyond its maximum throughput. The decodable channel throughput of around 0.30 is already more than the maximum that could for example be achieved with the ALOHA access protocol where the maximum is reached around 0.18 [9]. In contrast, the channel access on the 1090 MHz channel is a mixture of a pseudo-random access by individual aircraft (e.g. ADS-B and TCAS) plus a mixture of scheduled access like Mode-S and Mode A/C replies to multiple surrounding radar stations that can be scheduled and adapted in a coordinated manner by the air navigation service provider.

Each of these applications have measures to reduce the channel load: For example Mode-S radar stations can coordinate the interrogations and interchange data for individual aircraft in order to reduce interrogations of the aircraft. Additionally, there are measures to reduce the impact of the spontaneous transmitted packets. For example, TCAS is reducing the transmission power for the interrogations, if a high number of surrounding planes is detected thus reducing the range of interrogations and the number of replies. The higher decodable throughput rate shows that these measures have a valuable effect on the channel performance and lead to a channel load that is higher than a random access channel would be.

Figure 6 shows the number of telegrams per identified transponder for the February and April data set. This is calculated by the number of identified aircraft transponder IDs divided by the number of telegrams per second. This is done because it is not always possible to identify the sending aircraft only by listening to its replies. Therefore, only individual aircraft IDs have been used to calculate the amount of aircraft even though it might not have been possible to determine their position or altitude during the reception. Figure 6 clearly states



**Fig. 6.** The number of telegrams per plane.

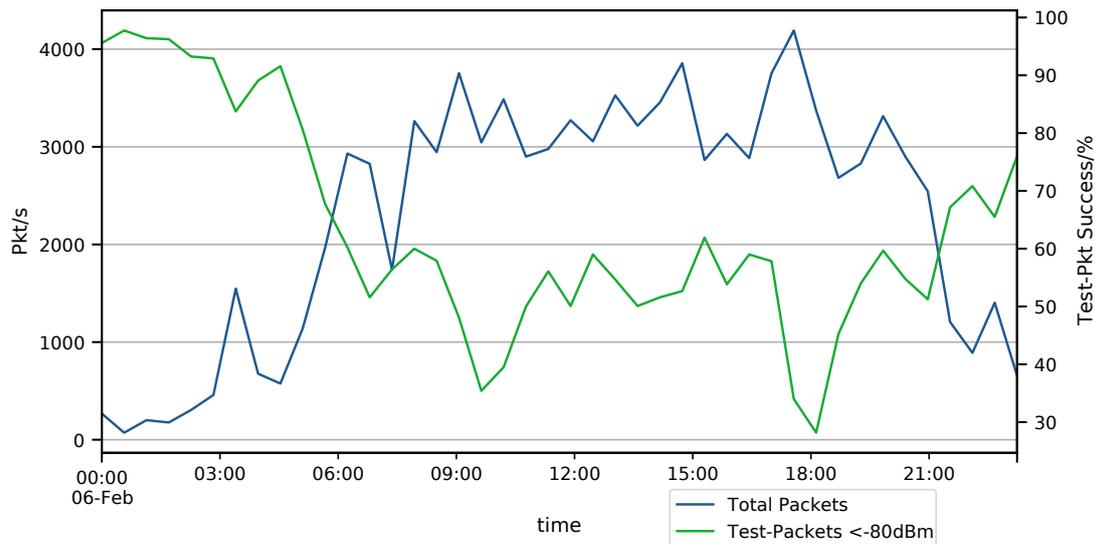
that in average there were less telegrams per plane received in February with a high amount of flights.

There are several possible reasons why this is the case. As of yet, the recorded data is not sufficient to answer this completely but together with the lost test telegrams it stands to reason that more and more telegrams are not decodeable as the channel is already close to its maximum throughput capability. To determine the reason for this behaviour further research is necessary. However, it can be observed that the number of decodable messages doesn't increase linearly with the number of planes. This implies that the channel is crowded and telegrams will be lost due to collisions. The more crowded the airspace the lower the rate of detected telegrams per aircraft. The details of these relations are scope of our future research where we will investigate the rate between successfully and falsely decoded telegrams with respect to the channel load.

### 3.4 Training samples

Training samples in the form of test telegrams play a key role in these measurements. We assume that the distortions, interference and other disruptions have the same effect on *natural* telegrams as they have on the test telegrams. So test telegrams are used in order to gain and improve the knowledge of the characteristics on that channel. The test telegrams are continuously injected into the channel with changing power levels (see section 2).

It is therefore possible to see the impact of changing radio environments on the test telegrams with changing levels. Figure 7 shows the median of the success rate for a single day in the February dataset. The figure clearly shows the strong inverse proportionality. The higher the number of telegrams per second the lower the chance for the test telegrams to survive. For this chart, only telegrams with



**Fig. 7.** Success rate and packet load on 6<sup>th</sup> of February

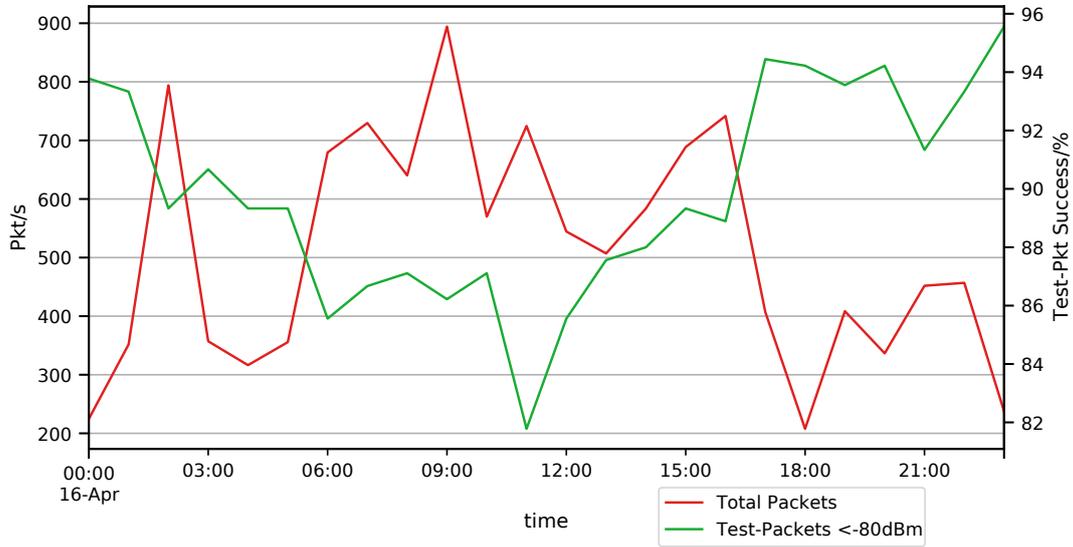
levels around -80 dBm have been taken into account which is the approximate average of all received telegrams.

Figure 8 shows a similar result for the April dataset but the overall success rate is much higher than in February. The success rates throughout the day are around the values that were achieved during the night in February.

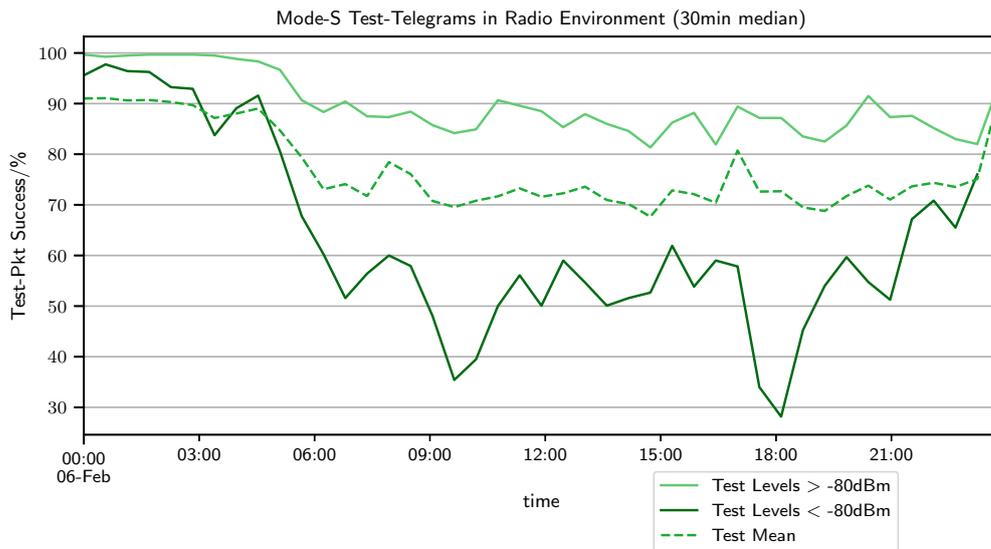
Figure 9 is showing the success rate for the 6<sup>th</sup> of February and indicates that the level of the test telegrams plays a crucial role in the success rate. The graph shows three lines throughout the day. The light-green line shows the average success rate for telegrams with high levels. It shows that the high levelled test telegrams hardly suffer from the increasing communication on the channel (refer to Figure 7) but even these high levelled telegrams suffer from that communication and drop from 100% to around 90% during the day. For lower levelled telegrams ( $s_l < -80$  dBm) the success rate is not only in general worse, but the impact of the radio environment is significantly stronger than for the high levelled telegrams.

## 4 Conclusion and Future improvements

As the described setup looks promising, a team of students from HOCHSCHULE DARMSTADT (h\_da) is currently working on an even more reduced measurement system. It uses a Realtek SDR RTL2832U based system that was designed for Digital Video Broadcasting–Terrestrial (DVB-T) reception. It is able to decode the 1090 MHz packets and transport them to a PC via USB Interface. A wide range of Open-Source software is already available for that. The test data generator is realised with an ADALM PLUTO<sup>®</sup> SDR by Analog Devices. This SDR evaluation board is able to send on the 1090 MHz channel and synthesise the



**Fig. 8.** Success rate and packet load on 16<sup>th</sup> of April



**Fig. 9.** Success rate for different levels on 6<sup>th</sup> of February

needed test telegrams. These devices are very affordable and can be used out of the box to generate the test data on the 1090 MHz channel. Together with a standard PC this setup should be sufficient to measure on the channels. As the losses of the receiver shall be estimated by machine learning, the impact of decoding capabilities are expected to be relatively low as long as a reasonable rate of the traffic can be received. The quantification of the losses and in a second step the extrapolation to the *real* amount of data should be possible with this equipment just with the professional Air Traffic Control receiver.

The comparison between the two approaches will be further investigated. If it turns out that the RTL2832U and the ADALM PLUTO® can be used to create a semi-supervised data generator for a machine learning based correction, the system can be set up at multiple locations over Germany in order to create a wide range of data with changing characteristics. This might also lead to the conclusion, whether there is a need for detailed monitoring with a high spatial resolution or if there are little advantages for such a distributed measurement network.

This data could then be joined with the data from the OPENSKY-NETWORK and used to create a large scale radio field monitor. Overall, the entire communication structure on the channel has changed since the corona lock-down stopped the majority of flights over Europe. This involuntarily allows a view on the surveillance radar channels and the communication that could not be expected before.

The described measurement setup delivers results that strongly relate to the results of the OPENSKY NETWORK and the reported decline of flights by the European authorities [5] but also allows new insights on the number of telegrams per aircraft and the saturation on the channels. These new insights do also raise new questions that need to be investigated in future works.

The low number of aircraft have a significantly higher probability to successfully transmit their results. As the number of telegrams dropped the channel is less occupied. This shows that the effects of interference and disruptions decline as the number of aircraft in the surrounding airspace is reduced.

The measurement system is able to show the effects on the surveillance radar channel in total and can give new and deeper insights into the communication on this channel. It has been shown that the number of aircraft does not linearly correspond to the traffic on the surveillance radar channels, thus a measurement and observation system is necessary to assess the channels' states and the remaining capacity. This is one crucial part to maintain a secure surveillance in civil aviation that will hopefully recover to its old strength very soon.

## Acknowledgement

The authors would like to thank Deutsche Flugsicherung GmbH for supporting this work.

## References

1. Administration, F.A.: Introduction to TCAS II. Advisory Circular pp. 1–50 (2011)
2. Bredemeyer, D.I.J., Erik J. Wischmann, Gottstein, J.: MOSTDONT - Final Report - Release 1.2. Tech. rep., European Aviation Safety Agency (EASA) (2010)
3. Deutsche Flugsicherung: LIZ Bulletin 16/2020 IFR movements (entries, exits, overflights and domestics) in all control sectors with a German designator. Tech. rep. (2020)
4. Eurocontrol: Ash cloud of April and May 2010: Impact on Air Traffic p. 45 (2010)
5. Eurocontrol: COVID19 Impact on European Air Traffic EUROCONTROL Comprehensive Assessment Air Traffic situation: Tue 21 April & mid-Week 17 (20-21 April) (compared with equivalent period in 2019) (2020)
6. Eurocontrol: Eurocontrol ADS-B (may 2020), <https://ads-b-europe.eu/>
7. European Commission, T.: Laying down requirements for the performance and the interoperability of surveillance for the single European sky. Official Journal of the European Union (1207), 35–52 (2011)
8. Gislason, S.R., Alfredsson, H.A., Eiriksdottir, E.S., Hassenkam, T., Stipp, S.L.: Volcanic ash from the 2010 Eyjafjallajökull eruption. Applied Geochemistry (2011)
9. Kuo, F.F.: ALOHA system. Computer Communication Review (1), 41–44 (1995)
10. Otsuyama, T., Honda, J., Naganawa, J., Miyazaki, H.: Analysis of signal environment on 1030/1090MHz aeronautical surveillance systems. 2018 IEEE International Symposium on Electromagnetic Compatibility and 2018 IEEE Asia-Pacific Symposium on Electromagnetic Compatibility, EMC/APEMC 2018 p. 71 (2018)
11. Raekow, R.: Funkfeldüberwachung von Sekundärradarkanälen (german) Diplomarbeit, Hochschule Darmstadt (2010)
12. Raekow, R., Kuhn, M., Wenning, B.L.: A novel approach to emulate and detect packet loss on surveillance radar channels. In: IEEE VTC Spring 2020 Conference Proceedings (2020)
13. Schafer, M., Strohmeier, M., Smith, M., Fuchs, M., Pinheiro, R., Lenders, V., Martinovic, I.: OpenSky report 2016: Facts and figures on SSR mode S and ADS-B usage. AIAA/IEEE Digital Avionics Systems Conference - Proceedings (2016)
14. Stanzel, S.: DFS ADS-B Implementation in High Density Radar Controlled Airspace-Experiences and Challenges. Tech. rep. (2019)
15. Walberer, A., Marquard, S., Herber, A., Fischer, H.: SSR Radio Field Simulation and Monitoring-Motivation, Methods and Results. Tech. rep. (2013)

# Wearable Gait Analysis – stepping towards the mainstream

Philip Morrow, Huiru Zheng, Graham McCalmont, Haiying Wang, and Sally McClean

School of Computing, Ulster University,  
Jordanstown Campus, Shore Road,  
Newtownabbey, Co. Antrim, BT37 0QB, Northern Ireland,  
{morrow-p4, h.zheng, mccalmont-g, hy.wang, si.mcclean}@ulster.ac.uk

**Abstract.** Wearable technologies have transformed the accessibility of gait analysis, offering the opportunity to venture outside of the laboratory and into everyday life. This research article is concerned with investigating the progress that has been made, and the steps that remain in gait analysis becoming a mainstream activity. The evidence for the effectiveness of wearable gait analysis technologies was reviewed, indicating that these devices are capable of supporting gait analysis in a ‘real-world’ environment. Research into the application of wearable technology for gait analysis was found to be limited in terms of scope, with progress still to be made in improving the perception of these devices. Challenges to be addressed within this field of research were identified: (1) Large scale data collection; (2) Broader scope of wearable gait analysis; (3) ‘Real-world’ gait analysis; (4) Case study research approach (5) Gait analysis as a service; (6) User testing/evaluation.

The development of wearable gait analysis systems, and the underlying research that supports their application, should be cognizant of how mainstream acceptance is contingent upon meeting these challenges. The path towards addressing them is considered in the context of the eZiGait portable gait analysis system, highlighting the value of collaboration with industry.

**Keywords:** Gait analysis, wearable devices, inertial measurement unit, smart insoles, clinical application, gait rehabilitation, industry perception.

## 1 Introduction

Gait Analysis is concerned with the study of human motion in order to develop an understanding of an individual’s ability to walk. A normal gait pattern has a positive effect on quality of life by enabling a person to perform their everyday activities uninhibited [1]. This ability is assessed by analysis of specific characteristics that constitute their walk pattern. The analysis of the gait pattern of an individual involves the extraction of specific parameters, such as spatiotemporal parameters, that are associ-

ated with different characteristics of walking. The different phases of walking constitute an overall gait cycle that represents the activity between one foot contacting the ground and that same foot again contacting the ground. Depending on the level of granularity used, the gait cycle can be composed of up to eight phases [2]. At the coarsest level of granularity, the two main phases are the stance phase, where the foot remains in contact with the ground, and the swing phase where the foot is lifted off the ground. Identifying different phases of the gait cycle enables spatiotemporal parameters to be extracted e.g. walking speed, step time, step length etc. The use of pressure and force sensors enables kinetic parameters such as centre of mass and centre of pressure to be determined [3].

The development of gait analysis methods has largely been concerned with healthcare and rehabilitation. The role of analysing the gait of an individual in providing insight into their general health has long been established [4]. Conditions that have been determined to result in gait disorders include Cerebral Palsy [5], Parkinson's Disease [6] and Alzheimer's Disease [7]. For rehabilitation, gait analysis can contribute to improving an individual's limb movements to enable them to regain a normal walking function. The applications of gait analysis within sports include improving performance [8] and avoiding injury [9]. For security purposes, individual gait profiles have been demonstrated to provide the basis for identifying a person [10]. The range of applications continues to expand, with artificial gait for humanoid robots and more realistic computer-generated models for the entertainment industry among recent developments [11].

Traditionally gait analysis took the form of human observation of an individual subject walking. This observation was performed by a trained professional and could be complemented with self-reporting by the subject. General quantitative features such as walking speed, as well severe gait disorders, can be discerned through this approach. These judgements are inherently subjective in nature and thus limit their reliability in informing a medical diagnosis or treatment decision. More significantly, subtle deviations from 'normal' gait cannot be discerned from human observation limiting the insight that can be provided by this analysis. More sophisticated gait analysis approaches were developed to facilitate more objective analysis of gait. For example, motion laboratories incorporating such techniques as video imaging to record movements enable analysis of kinematic information. This generally involves the placement of markers on the body of the test subject, with marker-less systems yet to gain as widespread use [12]. The use of single or multiple cameras determines whether 2D or 3D motion analysis can be performed [13]. Force plates measure vertical ground forces exerted by an individual to enable analysis of kinetic information [14]. The primary downside of such approaches was the considerable cost incurred in setting up an appropriate laboratory to conduct the analysis. The prohibitive cost in turn ensured that access to this form of gait analysis was limited to specialist locations. The advent of wearable technology greatly reduced the barriers to use present with conventional technology, thus opening the possibility of widening access to gait analysis. Despite the breakthrough provided by these wearable devices mainstream acceptance of gait analysis remains a goal yet to be achieved. This research article considers the developments made within the field of wearable gait analysis and the barriers

ers that remain in fully realizing its potential. Potential approaches that may be adopted to move towards mainstream acceptance are then discussed in the context of the eZiGait mobile gait analysis system that is currently being developed.

## **2 Review of Wearable Gait Analysis Research/Development**

The review of developments within this area commences with a comparative overview of the main types of gait analysis technologies that have been developed thus far. The evidence for their effectiveness of wearable devices for enabling gait analysis is then considered. The contribution that these devices can make is then examined in terms of the scope of previous research and the potential users of such technology.

### **2.1 Comparative Overview of Gait Analysis Technologies**

Table 1 provides an overview of the main types of technologies that are used to perform gait analysis. Optical Motion Capture systems represent the most accurate technology for capturing Kinematic information, but correspondingly incur the highest costs in terms of initial setup and subsequent operation. Force plates systems, that measure Vertical Ground Forces, can be implemented at varying levels of cost. Both systems are limited to use within a laboratory environment and only cater for recording data over limited durations. The application of gait analysis is therefore limited to controlled exercises that may not be reflective of the ‘normal’ walking patterns exhibited by an individual person. Gait Pressure mats have been developed as a more practical alternative, with medium costs incurred and some degree of portability. The relative bulk of such systems still inhibits their accessibility to some extent, and they remain suitable for indoor use only. The limited length of the mats means that they are unsuitable for extended recording durations. The development of wearable technologies represents the most significant breakthrough in terms of accessibility. The main types of inertial sensors provide measurements in up to 3 axes of motion and can be incorporated together in the form of an Inertial Measure Unit (IMU). Accelerometers provide measurements of the rate of change of velocity for an individual in their own real frame. Gyroscopes provide measurements of orientation and angular velocity. Magnetometers measure the direction and strength or relative change of a magnetic field at a location. These inertial sensors may be utilised within a standalone device or within other technology such as a smartphone. The presence of inertial sensors within smartphones provides the greatest degree of availability, although standalone IMU devices provide greater flexibility e.g. can be worn on different parts of the body. The location of the IMU device impacts both the measurement accuracy achievable and the level of comfort for the user. The costs of such devices can vary according to how they are implemented, but a significant reduction can be achieved relative to lab-based systems. The suitability for outdoor use and capacity for captur-

ing sensor data over an extended duration provides a significant advantage in terms of measuring ‘real-world’ data.

**Table 1.** Comparison of Main Technology Types

<b>Technology Type</b>	<b>Key Characteristics</b>	<b>Practicality</b>
Optical Motion Capture System	<ul style="list-style-type: none"> <li>- High Measurement Precision</li> <li>- High Technology Costs</li> <li>- High Computational Costs</li> </ul>	<ul style="list-style-type: none"> <li>- Lab-based System</li> <li>- Requires expert operation</li> <li>- Limited recording duration</li> </ul>
Force Plates	<ul style="list-style-type: none"> <li>- Low Measurement Precision</li> <li>- Low to High Technology Costs</li> <li>- Low Computational Costs</li> </ul>	<ul style="list-style-type: none"> <li>- Limited Portability (due to weight) and used mainly within Lab environment</li> <li>- Limited recording duration</li> </ul>
Gait/Pressure Mat	<ul style="list-style-type: none"> <li>- Variable Measurement Precision</li> <li>- Medium Technology Costs</li> <li>- Low Computational Costs</li> </ul>	<ul style="list-style-type: none"> <li>- Portable but suitable for indoor use only</li> <li>- Requires expert operation</li> <li>- Limited recording duration</li> </ul>
Inertial Measurement Unit (Accelerometer, Gyroscope, Magnetometer)	<ul style="list-style-type: none"> <li>- Low Measurement Precision</li> <li>- Low to Medium Technology Costs</li> <li>- Low Computational Costs</li> </ul>	<ul style="list-style-type: none"> <li>- Portable use both indoors and outdoors</li> <li>- Can be worn and operated by user</li> <li>- Level of comfort dependent upon design</li> <li>- Extended recording duration</li> </ul>
Insole Pressure Sensor System	<ul style="list-style-type: none"> <li>- Low Measurement Precision</li> <li>- Low to Medium Technology Costs</li> <li>- Low Computational Costs</li> </ul>	<ul style="list-style-type: none"> <li>- Portable use both indoors and outdoors</li> <li>- Can be worn and operated by user</li> <li>- Extended recording duration</li> </ul>

Insole Pressure Sensor systems provide a more accessible alternative to Force Plates and at a significantly lower cost. These insoles deploy an array of sensors to record pressure at several locations on the sole of each foot. The number of sensors embedded into an insole can range from a few to over 40 depending on the design.

The level of availability of Insole Pressure Sensor systems is not as widespread as that of smartphone-based inertial sensors, but they otherwise retain the same benefits in terms of measuring ‘real-world’ data. Smart insole systems may combine the use of inertial sensors with pressure sensors, enabling a wider variety of application for any recorded data as well as the potential for fusion of the different types of data. In these systems, the inertial sensors may be attached to the insoles or embedded directly into the insoles. The development of these smart insole systems has addressed many of the shortcomings of previous gait analysis systems.

## 2.2 Clinical Application of Gait Analysis

Jarchi et al. [15] reviewed gait analysis involving accelerometry with a focus on how it is applied to clinical applications. The review included 159 research papers, starting from the year 2000, and defined three classifications for these studies. The most common category, at 38% of research papers, was the validation of gait parameters against more established approaches such as video analysis, force-plates etc. The other two categories were focused on applying gait analysis to clinical applications, with 32% of papers utilising an accelerometer and 30% of studies incorporating single or multiple configurations of sensors.

Baker et al. [16] described four main reasons for performing clinical gait analysis. The first reason was the diagnosis of a specific disease, followed by the assessment of the severity of that disease or an injury. The other two reasons were concerned with either monitoring or predicting the progress of the disease or injury, with or without an intervention taking place on the patient. The success of clinical gait analysis can be judged according to the extent that it supports the desired outcomes for these intended uses i.e. its efficacy.

Table 2 shows how research involving the clinical efficacy of gait analysis has progressed over the past two decades. This research was focused on conventional three-dimensional instrumented gait analysis performed using laboratory methods such as motion capture and force plates. Two systematic reviews, conducted by the same authors [17] - [18], categorised the research studies published during their respective preceding decade according to the efficacy type that was addressed. For the second systematic review, a new efficacy type (2b) was introduced to reflect how the focus of research had evolved. This type 2b research was concerned with an evaluation of the efficacy of treatment at a group level, in contrast to individual patient outcomes addressed at type 5. In some cases, the research conducted in an individual study incorporated multiple adjacent efficacy types. The overall number of research studies had risen significantly in the latter decade, from 210 to 2712 papers, indicating the increased significance of this field of research. The level of efficacy that has been investigated most has also changed from type 1 to type 2, including the new type 2b. This may reflect the progress beyond evaluating the technical performance of gait analysis methods towards their clinical application. The efficacy type 2b research indicated the significance of being able to diagnose patients as being part of a group

and consequently predicting treatment outcomes for that type of patient. The absence of any increase in research studies at higher efficacy types indicates the limitations on the understanding of how gait analysis can contribute at these levels. In particular, the question to be asked is whether the limitations on accessibility associated with laboratory gait analysis have themselves impacted the progress of research in this regard.

**Table 2.** Number of Research Studies Grouped by Efficacy Type

<b>Efficacy Type</b>	<b>Number of Research Studies (Jan 2000 to Sept 2009)</b>	<b>Number of Research Studies (Sept 2009 to Oct 2019)</b>
1 - Technical	116	313
2 - Diagnostic Accuracy	89	1466
2b - Outcome Prediction	-	927
3/4 - Diagnostic Thinking & Treatment	11	6
5 - Patient Outcome	7	3
6- Societal	1	0

The highest level of efficacy (type 6) was concerned with the cost effectiveness of gait analysis in terms of its impact on society. The need for such research was highlighted in a conference held amongst professionals within this field to establish a consensus on the general progress of gait analysis [19]. The absence of cost effectiveness studies on the use of motion laboratories was broadly agreed to be a significant problem. The absence of standardised procedures for gait analysis and publicly available data on ‘normal’ gait profiles was also identified as a problem to be addressed. The scientific evidence supporting the use of gait analysis was generally considered limited to a few medical conditions. The overriding conclusion of the conference was that the value of gait analysis as a research tool has thus far exceeded its value for clinical applications.

### 2.3 Overall Effectiveness of Wearable Gait Analysis Devices

Numerous research studies have examined the performance of wearable devices in the context of Gait Analysis. These studies have sought to demonstrate the feasibility of utilising wearable sensors for gait analysis and provide a means of evaluating them as a substitute for more conventional laboratory approaches.

Kobsar et al. [20] conducted a systematic review of studies on the validity and reliability of wearable inertial sensors. The context of this review was on healthy adults walking as opposed to individuals with any underlying medical conditions. The review covered research papers from 1998 to 2019 in terms of set parameters, rating each parameter according the quantity, quality and consistency of results across studies that addressed the parameter. For the step time and stride time parameters the results were rated as excellent. For step length, stride length, swing time and stance

time the results were rated as good to excellent. For gait variability and gait symmetry parameters the results were only rated as poor to moderate. The issue with the lower performing parameters was more an issue of the limited number of studies, or their design, rather than the performance of the sensors themselves. The main weaknesses of the research studies were from a statistical perspective e.g. underpowered results, unjustified sample sizes, and inadequate statistical analysis in support of the evaluation of results. The meta-analysis undertaken within this review highlighted the difficulty in comparing results across studies. The large number of studies were completed without a standardised protocol and addressed different subsets of gait parameters. For each gait parameter there were typically between three to five studies that could be compared. Establishing a comprehensive appraisal of the validity and reliability of wearable inertial sensors was concluded to more a question of quality than of quantity in terms of future research studies.

Direct evidence for the performance of inertial sensor-based systems relative to that of conventional gait analysis approaches was reviewed for the time period from 2005 to 2017 [21]. In this systematic review, a total of 16 research articles were selected that compared gait parameters obtained using wearable or inertial sensors against those obtained using a motion laboratory. Meta-analysis was performed on seven different gait parameters addressed by multiple studies. Table 3 shows a summary of the results of this meta-analysis.

**Table 3.** Meta-Analysis of Gait Parameter Results

<b>Gait Parameter</b>	<b>Number of Studies</b>	<b>Number of Subjects</b>	<b>Standardised Mean Difference</b>	<b>95% Confidence Interval (Leftmost value, Rightmost value)</b>
Gait Speed	5	149	0.11	(-0.10, 0.33)
Step Length	6	245	0.14	(-0.14, 0.43)
Step Time	5	234	-0.03	(-0.24, 0.18)
Stance Time	3	140	0.50	(0.24, 0.76)
Stride Time	3	110	-0.03	(-0.29, 0.23)
Cadence	4	138	0.46	(-0.41, 1.34)
Swing Time	3	140	-0.63	(-1.32, 0.05)

Standardised Mean Difference (SMD) refers to the difference in the mean value of each group (IMU and motion laboratory) divided by the standard deviation. In this case, a positive SMD indicates that the IMU group had the greater mean value and a negative SMD indicates that the other group had the greater mean value. The leftmost and rightmost values of the 95% confidence interval for the SMD value provides an indication of the level of uncertainty associated with the value. Statistically significant differences between the groups are indicated when the 95% confidence interval does not contain zero. The results indicate that for gait speed, step length, step time and stride time there was a good level of agreement between the two groups. Only stance time demonstrated a statistically significant difference between the two groups.

Cadence and Swing Time had the greatest level of uncertainty regarding the SMD value, so caution is necessary in drawing any conclusions for those parameters.

The value of being able to obtain the gait parameters using relatively low-cost wearable devices in short walking exercises was noted by the authors of the review. Limitations on the understanding of the effectiveness of wearable inertial sensors that can be derived from this review were, however, identified. These included the different devices, algorithms and test procedures used by the studies. The use of wearable devices was not considered to be a substitute for motion laboratory gait analysis as the latter approach is necessary for identifying the locomotor strategy used by patients.

The use of Artificial Intelligence (AI) approaches such as machine learning for analysis of gait data has been a common research theme. The effectiveness of AI approaches to assist with gait analysis methods was considered by a systematic review of research studies focused on inertial sensors and adaptive algorithms [22]. The review, considering only Journal Papers, looked at research studies published between 1968 and 2016 that used IMUs and adaptive AI algorithms to classify gait events. The quality of results obtained from adaptive AI algorithms was reported to be above average, leading the authors to conclude that they are suitable for use in gait analysis. The lack of standardization in terms of the methods used was again highlighted as an issue moving forward. As most of the studies utilised healthy test subjects, the potential for improvements with patients whose gait is impacted by a medical condition was not discernible from the research.

The use of inertial sensors for gait recognition was reviewed by Sprager and Juric [23]. The performance of inertial sensor-based gait recognition was found to be more effective for large datasets, in the region of 200 people, with error rates of around 5% reported. These research studies generally used very short walking exercises, typically involving only a few steps, thus limiting the extent to which they would be replicated in practice. For those research studies that allowed for walking in an uncontrolled environment over a longer period the reported accuracy rates ranged from under 70% to above 90%.

## 2.4 Scope of Wearable Gait Analysis Devices

Chen et al. [24] conducted a systematic review of gait analysis research studies involving wearable sensors, examining how pervasive the technology could become. The authors asserted that only 0.2% of gait analysis papers both involved the use of inertial sensors and were conducted in real clinical settings. There was no indication that the total number of research papers accounted for those published before inertial sensors were developed so this relative proportion may be understated. The review included 2906 papers focused on applying gait analysis to medical conditions. The distribution of these studies was found to strongly favour some medical conditions with Parkinson's Disease (29%), Cerebral Palsy (17%), Orthoses (13%), Lower Limb Osteoarthritis (6%) and Post Stroke (6%) accounting for over 70% of the research papers. The remaining approx. 30% of research was distributed across 13 categories suggesting that research gaps remain to be addressed. Possible reasons suggested for this pattern of research were limited availability of affordable technology or a need to

align technology advancements with the medical knowledge within those domains. A systematic review of the use of smartphone systems for physical rehabilitation revealed a similar narrow focus [25]. From the 74 research studies that were reviewed, the diseases that dominated the research were stroke, cardiac disease, balance impairment and joint/limb rehabilitation.

In order to provide the greater accessibility within standardised medicine that wearable sensors are intended to provide it is necessary to gain acceptance within the medical community. In terms of research, using established gait analysis procedures ensures generalisability for the results. Numerous research studies have shown that standard clinical tests such as the Timed Up and Go (TUG) test and the 6-Minute Walk test, can be successfully administered using low cost sensors in mobile or wearable devices [26], [27]. The Standardised procedure outlined for the TUG test has also been successfully adapted towards use with wearable devices [28]. The usage of wearable devices for standard tests of a longer duration, as with the 6-Minute Walk test, has also enabled a more detailed analysis of gait features to be performed. For example, sensor data obtained from turning during the test provides a clearer indication of impaired gait than may be obtained from regular walking in a straight line.

Research into the use of wearable devices for gait analysis outside of the laboratory environment was reviewed by Benson et al. [29]. The research papers were comprised of 43 walking studies, 13 running studies and a single study incorporating both forms of exercise. While the outcomes of the studies were assessed as being sufficiently reported they generally did not adequately describe the statistical power of their results. There was some progress towards using larger number of participants within walking studies, but it was recommended further improving this by monitoring gait over longer periods of time and in natural environments. The need to address the usability of wearable sensors was also highlighted. For long term use, the location the device is to be worn and its size/weight need to be as unobtrusive as possible without compromising the validity of the measurements that can be obtained.

### **3 Adoption of Wearable Gait Analysis Devices**

Investigating the perception of wearable devices provides insight into the progress that has been made towards achieving mainstream acceptance. The development of wearable technology has been largely driven by potential medical applications, but in terms of impact the main market has been the general fitness industry [30]. The medical field represents a higher barrier to entry in terms of performance in comparison to the general fitness market. The level of accuracy acceptable for a consumer-oriented device would not generally be considered sufficient for supporting medical decisions. In contrast, familiarity with gait analysis is much less prevalent within general consumers.

Issues affecting the adoption of wearable gait analysis tools are discussed in the following sections. This considers the perception of wearable devices amongst both professional users, interested in adopting them into their work practices, and individu-

als that would have their gait characteristics analysed. An example of how gait analysis can be integrated into the wider healthcare system is then provided.

### 3.1 Perception of Wearable Devices

In a survey of the adoption of wearable sensors in the workplace [31], 90.4% of respondents that wore a device at work did so for monitoring general physical activity. The most popular devices that respondents expressed confidence in, such as “Fitbit”, were typically providing easy to understand data such as step counts. Doubts were commonly expressed by respondents about the validity and efficacy of more sophisticated uses of wearable technology.

Mobile phone health apps have experienced substantial growth, but a systematic review of the scientific evidence behind their diagnostic performance was underwhelming [32]. Even when including research focused on symptom monitoring, for supporting diagnosis, it was found that studies were lacking both in terms of quantity and quality.

The need for a patient-centred focus has been promoted as a response to the increasingly complex delivery of healthcare. In the context of physical rehabilitation, several themes have been identified within research [33]. It is important that treatment is individualized for the patient and that they have a good understanding of their symptoms and treatment path. Goal setting and establishing a feeling of empowerment for the patient were reported as having a positive effect in helping them cope with their condition. The use of wearable devices offers the potential for providing patients with this patient-centred treatment both in and outside of physiotherapy sessions.

Morris et al [34] conducted a survey of clinicians concerning their perspectives on the use of mobile health and rehabilitation applications. Over 500 clinicians were surveyed, drawn from a range of professions such as physical, occupational and psychological therapists. While a large proportion of the respondents reported prescribing specific out of clinic exercises and interventions to patients, it was clear that acceptance of mobile solutions continues to be a challenge. Only 51% of respondents expressed comfort in integrating mobile technology into their clinical practice. In addition, only 23% of respondents considered themselves to be knowledgeable about the available technology. The top barrier to their use, identified by 72% of respondents, was the inability of patients to learn and correctly use mobile technology.

The perspectives of physiotherapists on the use of wearable or mobile health technology were investigated by Blumenthal et al. [35] In this study a simplified version of the popular Technology Acceptance Model (TAM) framework [36] was used to investigate the willingness of 76 participating therapists and students to implement these types of technology into their clinical practice. The primary motivation of the research was to investigate why the usage of this technology remained low within the physiotherapy industry. The study found no evidence that early adopter behaviour was influenced by age or previous experience with technology. It was suggested that the perceived usefulness of the technology was an important determinant of early adoption. Encouraging physiotherapists to invest their time and resources in imple-

menting this technology requires a clear demonstration of how it would add value to their practice. Increasing patient engagement and improving how progress is communicated were rated as highly important by the study participants. The importance of the user experience was therefore proposed as an important design consideration for mobile health technologies.

Guillen-Gamez and Fernández [37] investigated the perceptions of the subjects involved in wearable technology research. The study included a total of 606 patients and relatives, comprised of 60.2% female and 39.8% male participants. The attitude of the participants towards the usefulness of wearable devices was reported as being medium high, with a slightly higher level reported for males. In particular, the participants viewed the devices as being more useful for caring for the health of elderly people than for themselves. While men were generally more accepting of the use of wearable devices there was a greater contrast across age for female participants. Women under 30 owned more wearable devices than any other group, but women over 45 had the lowest level of acceptance for the use of such devices. In terms of the acceptance expressed for the location of the wearable device, on the wrist had the highest level (Male 52.7%, Female 50.19%) out of five options. The option of wearable devices placed in shoes, relevant for Smart Insole technology, was ranked third for Male (48.77%) and fourth for Female (45.30 %) participants.

Evaluating usability and accessibility of mobile health apps has been identified as an important consideration for achieving a patient-centred focus in improving rehabilitation outcomes [38]. In this review, usability was reported to be high for those apps that were educative and supported self-reporting of symptoms. For apps focusing on intervention, the most positive effects were found with functional outcomes such as gait and self-management skills. Positive effects were also found with health outcomes such as pain and quality of life. Evaluating the impact of mobile health systems requires consideration of both types of outcomes. Investigating functional outcomes from using these systems outside of the controlled laboratory environments is therefore as important as investigating their accuracy.

### **3.2 Integration of gait analysis into clinical practice**

The acceptance of gait analysis within mainstream clinical practice is contingent upon successful integration into the wider healthcare system. Research on how to achieve this integration has been limited, but an in-depth study on the design of a gait test within clinical rehabilitation provided instructive guidelines [39]. This study adopted a service-oriented approach to design in contrast to viewing gait analysis as a technological challenge. The multiple stakeholders in this service include patients (and their relatives), doctors, and therapists. Each of these stakeholders present different needs that must be met and therefore they must be considered in the design process. The design in this study was comprised of three main phases. Firstly, 'User-product proximity effect' where the gait test is performed in a motion laboratory and the effects on the various stakeholders are observed. Secondly, 'Effect and value in the service' where an overview is obtained of the path followed by the patient through the service

e.g. diagnosis, care and treatment decisions. Thirdly, ‘User interactions’ that define the information flows between stakeholders that are necessary to facilitate implementation of the gait test. The design recognised the need to treat the patient as an individual that will respond in their own way to treatment. Providing guidance to the patient at each stage of the service being provided is essential in enabling them to view the gait test as a motivational tool within their overall treatment.

## 4 Towards Mainstream Adoption of Wearable Gait Analysis

The review conducted in the preceding sections enabled the identification of the challenges to be addressed for wearable devices. The eZiGait system is introduced as an example of a wearable gait analysis system, with consideration then given to how future development of this system can be directed towards addressing the challenges that have been identified.

### 4.1 Future Challenges for the adoption of Wearable Gait Analysis Devices

The issues to be addressed on the path towards mainstream acceptance of Wearable Devices for gait analysis can be summarised as follows:

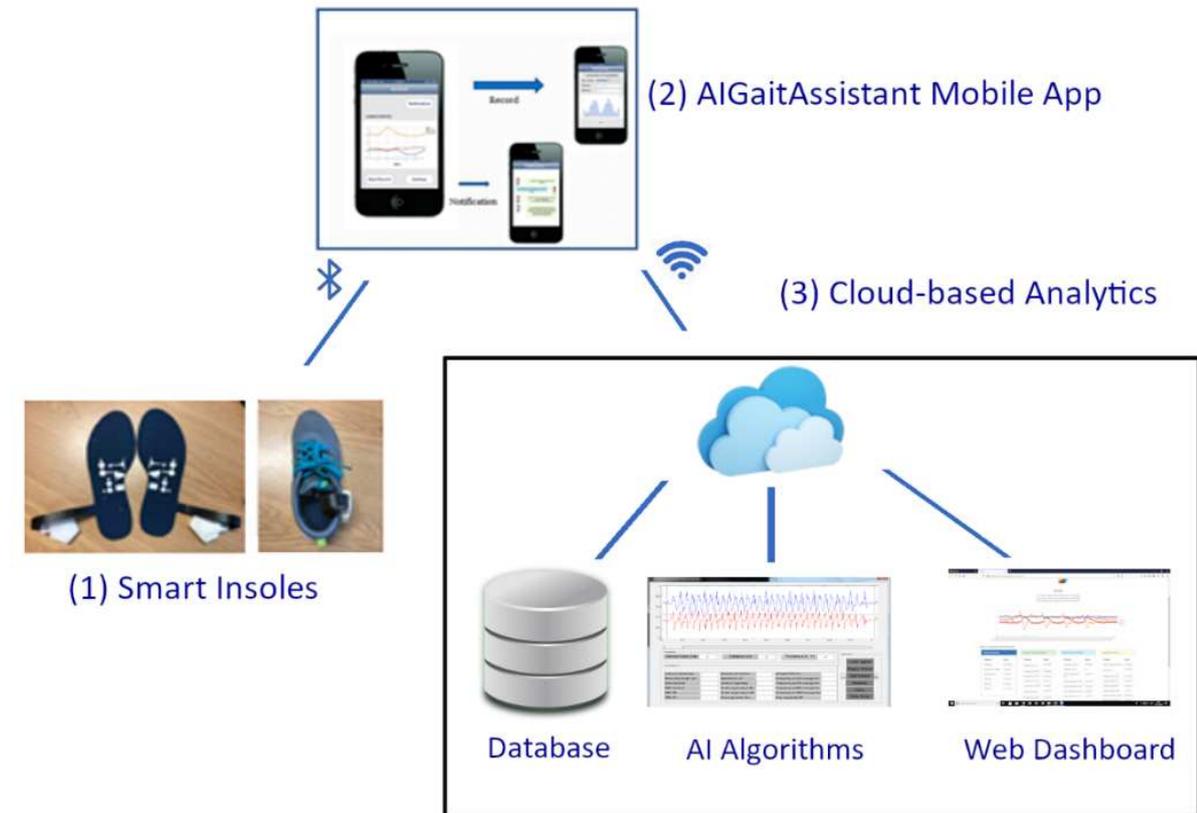
1. **Large scale data collection:** in order to achieve a greater understanding of normal and abnormal gait profiles it is necessary significantly increase the size of datasets. The accessibility and cost of wearable sensor systems needs to be considered if they are to enable the establishment of datasets on a large scale. The increase in the adoption of these systems could potentially decentralise the establishment of these datasets from researchers to the users.
2. **Broader scope of wearable gait analysis:** to facilitate the acceptance of gait analysis it is necessary to broaden the nature of the data collected. Instead of focusing on a limited selection of pathological conditions research should explore whether an individual’s gait profile is affected by other conditions. In terms of the sports and leisure industries it should be investigated whether gait analysis can benefit the more ‘casual’ athlete rather than being restricted to professional athletes. The impact of wearable gait analysis should be to move gait analysis beyond being restricted to relatively niche markets.
3. **‘Real-world’ gait analysis:** the development of wearable sensor devices has enabled the collection of subject data in ‘real-world’ conditions. It is imperative for research to increase efforts to engage in the collection of data in these conditions. Wearable sensor technology provides the basis for users to record data in their own time and location.
4. **Case study research approach:** increasing the quality of research includes the adoption of in-depth research studies. Case studies involving the progress of patients through the treatment/rehabilitation process are necessary to understand both how gait analysis can inform decisions made at each stage and how it impacts recovery for a patient.

5. **Gait analysis as a service:** the value of wearable sensor systems is to be derived from the service they provide to the various user groups. The objective is therefore to identify the needs of each user group, such as patients, medical professionals, athletes etc., and incorporate them into the services provided by the system. Achieving this objective requires including each user group as stakeholders and having them guide the direction of development of these services.
6. **User testing/evaluation:** to ensure the suitability of wearable sensor systems to fulfil their intended role, they must be subjected to appropriate evaluation by the intended users so that their design can be refined accordingly. For the wearable devices, it is necessary to ensure that the size/weight and location of the technology remains comfortable for the patient. For the supporting software, the gait analysis reports produced must be both understandable to the users and convey information that is of value to them.

The path towards addressing these challenges would include collaboration with organisations with a vested interest in the growth of wearable gait analysis, such as medical and rehabilitation organisations. The patients/customers of such organisations represent prospective users that need to be reached in order to achieve the goal of wearable technology i.e. to improve their quality of life.

## 4.2 eZiGait System Overview

The architecture of the eZiGait system is shown in Fig 1. The main components of the system are the Smart Insoles, the AI Gait Assistant mobile application, and the Cloud-based Analytics. The Smart Insoles used in previous research studies [40] - [41] were provided by TreeHouse Technology Ltd in China. There are eight separate pressure sensors built into each insole, providing readings of Vertical Ground Forces for eight locations within each foot. A pressure map is thus obtained for each foot, charting the distribution of weight exerted across each foot during the Gait cycle. An attached electronic device, that is worn around the ankle, is equipped with an IMU. This IMU provides a 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer. By simultaneously sampling data in three axes, the test subject's movements are effectively monitored.



**Fig. 1.** Overview of eZiGait System

The electronic device is equipped with a rechargeable battery that allows sensor data to be recorded over prolonged exercise sessions. The battery for this device is charged using any standard USB connection. The sensor readings are streamed in real-time and transmitted via Bluetooth to any paired device, such as a smartphone. An API is provided for the Smart Insoles that facilitates communication with other applications. A mobile app, AIGaitAssistant, has been developed for Android based smartphones. This app serves as the intermediary between the Smart Insoles and the Cloud-based Analytics, analysing the gait characteristics of the test subject. The following main functionality is included within the smartphone app:

- Connect to Bluetooth paired Smart Insoles to receive real-time sensor data on test subject movements.
- Display of real-time sensor data from each of the sensors within the GUI. This enables the user to verify that each of the sensors is functioning correctly prior to performing any exercise.
- Record sensor data for a specified exercise session. Each session is defined by the user in terms of a label to identify that session, the type of exercise to be completed, automatic or manual upload and generation of a gait report. The recorded session data is automatically saved locally on the smartphone as a CSV file.
- Upload the sensor data to a web server via a Wi-Fi connection for subsequent cloud-based analysis.

- Generation of a Gait Report for the completed exercise session. This report is comprised of the following detail
  - Overall summary of the gait session e.g. time, step count etc.
  - Gait Phase Distribution for each foot, with the overall gait cycle divided into Loading, Foot-Flat, Pushing and Swing Phases
  - Temporal Symmetry between each foot for Step Time, Stance Time, Swing Time and Stride Time
  - Pressure Symmetry between each foot for Heel Region, Mid-foot Region, Toe Region and Overall Foot.

The Cloud-based Analytics are hosted on a web portal that enables further examination of the recorded sessions by healthcare/rehabilitation professionals. For example, the progress of an individual in terms of their gait characteristics can be tracked over the course of their uploaded sessions. The analysis provided by this web portal continues to be developed with the aim of applying advanced machine learning approaches to establish a greater understanding of an individual gait.

### 4.3 eZiGait and the Future of Wearable Gait Analysis

The several challenges to be addressed for wearable gait analysis to gain more mainstream acceptance are being considered during the development of eZiGait.

eZiGait is being developed in conjunction with related research projects concerned with gait analysis. It has been utilised within previous research studies [40] - [41] in addition to internal data collection and validation. This provides a mechanism for conducting user testing and refining both functional and non-functional user requirements. It also presents opportunities for collaborating with external parties in ‘real-world’ environments. For example, the Smart Insoles and smartphone app have been utilised during rehabilitation sessions with a local physiotherapist. This provided the opportunity to learn from prospective users of the system and refine the user experience based on the feedback obtained. This ongoing collaboration will provide the opportunity to collect data from both healthy subjects and rehabilitation subjects. The healthy subjects include individuals participating in regular sports exercises within a gym environment. This enables data related to ‘normal’ gait to be collected in a ‘real-world’ situation in contrast to a lab-controlled environment with specific exercises. The access to rehabilitation subjects presents a learning opportunity that addresses multiple challenges.

- The physiotherapist is one the few specialists available in Ireland that provides the use of an exoskeleton as part of the rehabilitation process. The range of conditions affecting patients undergoing rehabilitation therefore presents the opportunity to collect data outside of the common conditions more regularly addressed within research. This will enable an understanding to be developed of how these conditions impact the gait profile of the subjects.
- The collection of data at successive stages of the rehabilitation process will enable the subsequent effect of the treatments on the subject’s gait profile to be assessed.

The in-depth case-study approach presents the opportunity to learn how gait analysis can form part of the rehabilitation process.

- The input from the physiotherapist throughout the research can guide the development of the gait analysis system by communicating which information is of use to the rehabilitation process and identifying potential new aspects of gait to be investigated.
- The needs of both the physiotherapist and the rehabilitation subjects, in terms of the service that is being provided, can be analysed and subsequently specified. This includes an understanding of the expectations held by these users prior to each gait analysis session and the level of satisfaction achieved upon completion of the session. Designing the service is as important to its success as the performance of the technology.
- The deployment of the system in a ‘real-world’ environment enables user testing and evaluation of the system to be performed. This will enable the system to be refined in order to enhance the accessibility of software. The presentation of information is integral to user understanding and subsequently them benefitting from their experience with the system.
- The benefits and costs of using the system can be investigated. The impact upon the rehabilitation process in terms of the extent of the recovery and the timeframe involved can establish what benefits, if any, that gait analysis provides. The cost incurred, such as the provision of the technology and the learning time required, can also be established, enabling an assessment to be made on the value that is provided.

## 5 Conclusion

Wearable gait analysis has provided a significant breakthrough in terms of the accessibility and affordability of such technology. This technology provides the potential for enabling gait analysis to become commonplace within mainstream healthcare systems and both the sports and leisure industries. Realising this potential is dependent upon directing the technology towards the needs of these users and demonstrating clear and cost-effective benefits. Research within this field should include investigating wearable gait analysis systems within the context of those industries that it aspires to benefit in order to further understand the role that this technology can fill. Bringing wearable gait analysis research out of the lab environment should enable the technology to follow suit and step forward into the mainstream.

## 6 Acknowledgement

This research is supported by the Invest NI Proof of Concept (PoC) programme (Project ID: PcC809).

## 7 References

1. Nutt, J.G., Marsden, C.D., Thompson, P.D.: Human walking and higher-level gait disorders, particularly in the elderly. *Neurology*. 43(2), 268-279. doi:10.1212/wnl.43.2.268, (1993)
2. Liu, T., Inoue, Y., Shibata, K.: Development of a wearable sensor system for quantitative gait analysis. *Measurement*. 42, 978–988. doi: 10.1016/j.measurement.2009.02.002. (2009)
3. Muro-de-la-Herran, A, Garcia-Zapirain, B, Mendez-Zorrilla, A.: Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors (Basel)*. 14(2), 3362-3394. Published 2014 Feb 19. doi:10.3390/s140203362 (2014)
4. Fish, D.J., Nielsen, J.P., Clinical assessment of human gait, *Journal of Prosthetics and Orthotics* 5(2), 39–48, (1993)
5. Tugui, R.D., Antonescu, D.: Cerebral palsy gait, clinical importance. *Maedica (Buchar)*. 8(4), 388-393 (2013)
6. Shichkina, Y, Stanevich, E, Irishina, Y.: Assessment of the Status of Patients with Parkinson's Disease Using Neural Networks and Mobile Phone Sensors. *Diagnostics (Basel)*. 10(4), 214. Published 2020 Apr 12. doi:10.3390/diagnostics10040214 (2020)
7. Higuma, M, Sanjo, N, Mitoma, H, Yoneyama, M, Yokota, T.: Whole-Day Gait Monitoring in Patients with Alzheimer's Disease: A Relationship between Attention and Gait Cycle. *Journal of Alzheimer's Disease Rep*. 1(1) 1-8. doi:10.3233/ADR-170001 (2017)
8. Wahab, Y & Abu Bakar, N.: Gait analysis measurement for sport application based on ultrasonic system. *Proceedings of the International Symposium on Consumer Electronics, ISCE*. 10.1109/ISCE.2011.5973775. (2011)
9. Silva, B., Natarajan, A., Motani, M. & Chua, K.: A Real-Time Exercise Feedback Utility with Body Sensor Networks. 49 - 52. 10.1109/ISSMDBS.2008.4575013. (2008)
10. Chen, J.: Gait correlation analysis based human identification. *Scientific World Journal*. 168275.. doi:10.1155/2014/168275 (2014)
11. Prakash, C., Kumar, R. & Mittal, N.: Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review*. 49, 1-40. 10.1007/s10462-016-9514-6. (2018).
12. Colyer, S.L, Evans, M, Cosker, D.P., Salo, A.I.T.: A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Medicine Open*. 4(1), 24. doi:10.1186/s40798-018-0139-y (2018)
13. Klette, R., Tee, G.: Understanding Human Motion: A Historic Review. In: Rosenhahn, B., Klette, R., Metaxas, D. (eds) *Human Motion. Computational Imaging and Vision*, vol 36. Springer, Dordrech doi:10.1007/978-1-4020-6693-1\_1 (2008)
14. Sutherland, D.H.: The evolution of clinical gait analysis part III--kinetics and energy assessment. *Gait Posture*. 21(4), 447-461. doi: 10.1016/j.gaitpost.2004.07.008 (2005)
15. Jarchi, D., Pope, J., Lee, T.K.M., Tamjidi, L., Mirzaei, A. and Sanei, S.: "A Review on Accelerometry-Based Gait Analysis and Emerging Clinical Applications," in *IEEE Reviews in Biomedical Engineering*, 11, 177-194, doi: 10.1109/RBME.2018.2807182. (2018)
16. Baker, R., Esquenazi, A., Benedetti, M.G., Desloovere, K.: Gait analysis: clinical facts. *European Journal of Physical and Rehabilitation Medicine* 52(4), 560-574. (2016)
17. Wren, T.A., Gorton, G.E. 3rd, Ounpuu, S., Tucker, C.A.: Efficacy of clinical gait analysis: A systematic review. *Gait Posture*. 34(2), 149-153. doi: 10.1016/j.gaitpost.2011.03.027 (2011)

18. Wren, T., Tucker, C., Rethlefsen, S., Gorton, G. & Ounpuu, S.: Clinical efficacy of instrumented gait analysis: Systematic review 2020 update. *Gait & Posture*. 80. 10.1016/j.gaitpost.2020.05.031. (2020)
19. Benedetti, M.G., Beghi, E., De Tanti, A., et al.: SIAMOC position paper on gait analysis in clinical practice: General requirements, methods and appropriateness. Results of an Italian consensus conference. *Gait Posture*. 258, 252-260. doi: 10.1016/j.gaitpost.2017.08.003 (2017)
20. Kobsar, D., Charlton, J.M., Tse, C.T.F., et al.: Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. *Journal of Neuro Engineering Rehabilitation*.; 17(1), 62. doi:10.1186/s12984-020-00685-3 (2020)
21. Petraglia, F., Scarcella, L., Pedrazzi, G., Brancato, L., Puers, R., Costantino, C.: Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine*, 55(2), 265-280. doi:10.23736/S1973-9087.18.05306-6 (2019)
22. Caldas, R., Mundt, M., Potthast, W., Buarque de Lima Neto, F., Markert, B.: A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. *Gait Posture*. 57, 204-210. doi: 10.1016/j.gaitpost.2017.06.019 (2017)
23. Sprager, S., Juric, M.B.: Inertial Sensor-Based Gait Recognition: A Review. *Sensors (Basel)*. 15(9), 22089-22127. Published 2015 Sep 2. doi:10.3390/s150922089 (2015)
24. Chen, S., Lach, J., Lo, B. and Yang, G.: Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review, *IEEE Journal of Biomedical and Health Informatics*, vol. 20(6), 1521-1537, doi: 10.1109/JBHI.2016.2608720. (2016)
25. Moral-Munoz, J.A., Zhang, W., Cobo, M.J., Herrera-Viedma, E., Kaber, D.B.: Smartphone-based systems for physical rehabilitation applications: A systematic review, *Assistive Technology*. 1-14. doi:10.1080/10400435.2019.1611676 (2019)
26. Ponciano, V., Pires, I., Ribeiro, F., Marques, G., Garcia, N., Pombo, N., Spinsante, S., & Zdravevski, E.: Is The Timed-Up and Go Test Feasible in Mobile Devices? A Systematic Review. *Electronics (Basel)*. 9(3). 10.3390/electronics9030528. (2020)
27. Storm, F.A., Cesareo, A., Reni, G., Biffi, E.: Wearable Inertial Sensors to Assess Gait during the 6-Minute Walk Test: A Systematic Review. *Sensors (Basel)*. 20(9), 2660. doi:10.3390/s20092660 (2020)
28. Salarian, A., Horak, F.B, Zampieri, C., Carlson-Kuhta, P., Nuttand J.G., Aminian, K.: iTUG, a sensitive and reliable measure of mobility, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18(3), 303–310, (2010)
29. Benson, L.C., Clermont, C.A., Bošnjak, E., Ferber, R.: The use of wearable devices for walking and running gait analysis outside of the lab: A systematic review. *Gait Posture*. 63, 124-138. doi: 10.1016/j.gaitpost.2018.04.047 (2018)
30. Seneviratne, S., Hu, Y., Nguyen, T., Lan, G., Khalifa, S., Thilakarathna, K., Hassan, M. & Seneviratne, A.: A Survey of Wearable Devices and Challenges. *IEEE Communications Surveys & Tutorials*. 1-1. 10.1109/COMST.2017.2731979. (2017)
31. Schall, M.C. Jr., Sesek, R.F., Cavuoto, L.A.: Barriers to the Adoption of Wearable Sensors in the Workplace: A Survey of Occupational Safety and Health Professionals. *Hum Factors*. 60(3), 351-362. doi:10.1177/0018720817753907 (2018)
32. Buechi, R., Faes, L., Bachmann, L.M, et al.: Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis. *BMJ Open*. 7(12), e018280. doi:10.1136/bmjopen-2017-018280 (2017)
33. Wijma, A.J., Bletterman, A.N., Clark, J.R, et al.: Patient-centeredness in physiotherapy: What does it entail? A systematic review of qualitative studies. *Physiotherapy Theory and Practice*. 33(11), 825-840. doi:10.1080/09593985.2017.1357151 (2017)

34. Morris, J., Jones, M., Thompson, N., Wallace, T., DeRuyter, F.: Clinician Perspectives on mRehab Interventions and Technologies for People with Disabilities in the United States: A National Survey. *International Journal of Environmental Research and Public Health*. 16(21), 4220. doi:10.3390/ijerph16214220 (2019)
35. Blumenthal, J., Wilkinson, A., Chignell, M.: Physiotherapists' and Physiotherapy Students' Perspectives on the Use of Mobile or Wearable Technology in Their Practice. *Physiotherapy Canada*. 70(3), 251-261. doi:10.3138/ptc.2016-100.e (2018)
36. Davis, F.D.: Perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*. 13(3), 319–40. <https://doi.org/10.2307/249008> (1989)
37. Guillen-Gamez, F. & Fernández, M.J.: Empirical Study Based on the Perceptions of Patients and Relatives about the Acceptance of Wearable Devices to Improve Their Health and Prevent Possible Diseases. *Mobile Information Systems*. 10.1155/2019/4731048. (2019)
38. Nussbaum, R., Kelly, C., Quinby, E., Mac, A., Parmanto, B., Dicianno, B.E.: Systematic Review of Mobile Health Applications in Rehabilitation. *Archives of Physical Medicine and Rehabilitation*. 100(1), 115-127. doi: 10.1016/j.apmr.2018.07.439 (2019)
39. Marín, J., Blanco, T., Marín, J.J., Moreno, A., Martitegui, E., Aragiés, J.C.: Integrating a gait analysis test in hospital rehabilitation: A service design approach. *PLoS One*. 14(10),: e0224409. doi: 10.1371/journal.pone.0224409 (2019)
40. McCalmont, G et al.: eZiGait: Toward an AI Gait Analysis and Assistant System, *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, 2018, pp. 2280-2286, doi: 10.1109/BIBM.2018.8621176. (2018)
41. An, G et al.: Feasibility study of a smart insole based AI Gait Assistant system, *BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering*, Hangzhou, China, 1-5. (2019)

# TeleStroke System (TSS)

Rishabh Chandaliya, Praveen Joshi, and Haithem Afli

ADAPT Centre,  
Cork Institute of Technology, Cork, Ireland

**Abstract.** According to the World Health Organisation(WHO), cardiovascular diseases (CVDs) are the number 1 cause of death worldwide, claiming nearly 17.9 million lives yearly [1]. Eighty-five percent of all CVD deaths are from heart attacks and strokes. This rising crisis has received relatively small coverage up to date, given its massive economic development effect in countries. However, Early detection of stroke and treatment of it is essential for an excellent outcome. This research aims to discuss the challenge of providing early detection of stroke through drooping mouth detection on Client-Server based architecture, a Tele-Stroke System (TSS), which is helpful for initial treatments. In this paper, the input data is from the Kaggle and YouTube Facial Palsy (YFP) database. We can find that simple Machine learning gives better accuracy than other deep learning models from the experimental results. It means that we can effectively assist doctors in early detection and diagnosing of stroke.

**Keywords:** Stroke Detection · Data Augmentation · face detection · facial palsy · droopy mouth detection · Machine Learning.

## 1 Introduction

Stroke is commonly described as a clinical syndrome of presumed vascular origin, characterized by rapid signs of local or global cerebral dysfunction that usually lasts for more than 24 hours or even death [18]. A stroke is said to occur when a blood vessel carrying nutrients and oxygen into the brain is either blocked, bursts, or ruptured by a clot. This results in decreased oxygen and blood supply in the brain, causing behavioral problems, memory loss, thinking habits, impairment, permanent brain damage, and even death [2].

Generally, there are three types of stroke. The first type of stroke is called Ischemic stroke. It occurs when there is a blockage in arteries either by gradual plaque build-up, blood clots, and other fat deposits. The second form is the hemorrhagic stroke that happens when a blood vessel located in the brain splits and induces blood to leak into the brain. This form of stroke is responsible for 13 percent of all strokes, and it accounts for over 30 percent of all stroke-related deaths. Lastly, an Embolic stroke occurs when a clot breaks off from the artery wall and becomes an embolus that can travel further down the bloodstream, resulting in the smaller artery's blockage. Embolic typically originates from the heart, where uncommon diseases may cause clot formation.

Depending on the latest National Stroke Association statistics[3], stroke is now the fifth leading cause of death in the United States alone and is a significant cause of adult disability. Every year, 0.8 million people in the U.S. are believed to have a stroke, and 0.13 million of them died from a stroke, meaning, on average, one American dies from stroke every 4 minutes. The lack of understanding of stroke signs is also one of the contributing factors for the unprecedented increase in death associated with the stroke.

## 2 Related Work

When you deal with stroke, it's always crucial to know what you want to identify. It is safer to provide a standardized program that addresses specific common stroke symptoms. One of the signs of stroke is facial paralysis, meaning, on one side of the face, there is facial sagging, which is known as face drooping. The droopy mouth is not only a symptom of stroke but also a sign of a brain and nerve disorder known as Bell's palsy. Bell's palsy, in short, is muscle weakness or paralysis on one side of the neck. As a result, facial nerve damage that controls the muscles on one side of the face causes that other side of the face to drop[4].

A Vision-based interface is provided where the user's hand, foot, or head movement are identified and monitored. The framework is designed for child-to-computer interaction. They base the device assessment on the concept of HAAT (Human Activity Assistive Technology) [5].

It projects the system to be less intrusive and cheaper than other solutions, in which they built up a video-based prediction process. This function: first, the movements of different parts of the body are isolated, then the motion characteristics are extracted and used to identify persons as stable or impaired. The results show that visually collected motion data allows the Identification of Bell palsy and Pseudoperipheral palsy as the state-of-the-art detection using data from electromagnetic sensors [6].

" Camera Mouse" was introduced in [7] to provide computer access for people with severe disabilities. With a video camera, the device monitors the user's gestures and converts them into the mouse pointer, which moves on the screen.

They built a visual monitoring system in [8] that passively observes moving objects in a site and learns from these observations patterns of behavior. The system is based on motion tracking, camera synchronization, recognition of operation, and events. The method is useful for sequence classification as well as for individual site operation instances.

In [9], they have converted a video into a sequence of frames used by correlating structures to predict human behavior. The machine examines the faces and identifies them to test the expression ROI (Region of Interest). It also utilizes Viola-Jones patterns, and the AdaBoost process to filter colors to identify the ears, noses, eyes, mouth, or upper body of people. These sections have higher entropy to detect emotions.

A method was proposed that recognizes three facial expressions in [10]; they use the geometric feature approach for extraction of features; the system uses the

neural network algorithm Multilayer Perceptron (MLP) with backpropagation for classification. The facial expressions to be identified are optimistic, happy, and surprised. The overall identification rate is 93.33 percent.

They provide a system for recognizing facial emotions in [11]; it is based on ASM (Active Shape Model) for facial feature extraction and Radial Basis Function Network (RBFN) to evaluate the symmetry of the face in different behavior patterns with accuracy 90.73

In [12], they developed a computer vision system that uses hidden Markov models (HMMs) to automatically identify individual action units or combinations of action units. Researchers used three methods to extract information on facial expression:

- Tracking point
- Complex flow monitoring with principal component analysis (PCA)
- Identification of high gradient components.

They present a system for face recognition and face emotion detection in [13], using the Open Source Computer Vision (OpenCV) library and python machine learning. Machine learning algorithms are used to recognize and classify different facial expressions and body movements. (OPPOSE)

The research focuses on image processing to obtain facial features/landmarks, mainly mouth corners, determine the droopy mouth threshold value and deployment using Android Studio. Consequently, the mobile application will address certain user groups such as neurosurgeons and emergency medical services. However, the mobile application could also be useful for patients, potential patients with strokes, and generally any user who is health-conscious anywhere and any-time[14].

## 3 Material

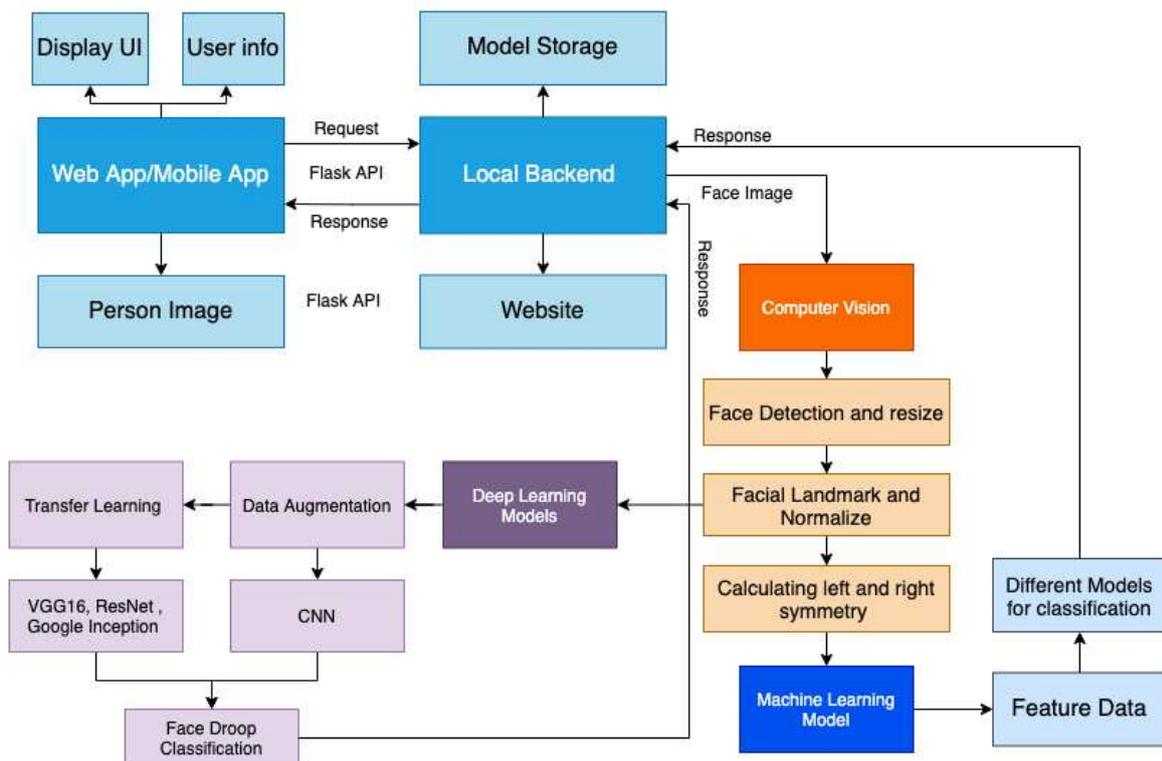
### 3.1 DATASET

- **Stroke Faces on Kaggle[15]**: This dataset consists of 1000 droopy faces curated from Google by Kaitav Mehta.
- **YouTube-Facial-Palsy-Database[16]**: The dataset in this paper contains 32 video clips collected from YouTube of 22 patients with facial paralysis. It is the first public database made available for the visual inspection study of facial palsy symptoms. Gee-Sern Jison Hsu, Jiunn-Horng Kang, Wen-Fong Huang[16] convert the videos into image sequences and have the images labeled by clinicians with facial palsy. As the number of patients is small, we have removed the patient's multiple pictures from the video.
- **Yale Face Database B[17]**: The collection comprises 5760 single-source light photographs of 10 subjects seen under 576 viewing conditions each. The subjects are healthy and do not have any symptoms of a stroke. A single image of subjects from the dataset is used for regular face input in CNN feed.

Number	Stroke	No Stroke
Training Images	900	900
Testing Images	100	100
<b>Total</b>	1000	1000

## 4 Method

This segment addresses the proposed TeleStroke system for Eye area and the mouth region. TeleStroke consists of client-server stages: the client-side relates to video pre-processing using face detection and landmark position to locate each sequence frame's faces. In this phase, the Integrated Deep Model is used. The face images detected are then cropped to the face, and the number of frames per sequence is normalized to a fixed length. The server side consists of various models, such as the ML model, FCNN, CNN, VGG16(Transfer Learning), one for each of the face analysis tasks. The proposed structure for Telestroke is given in figure 1.

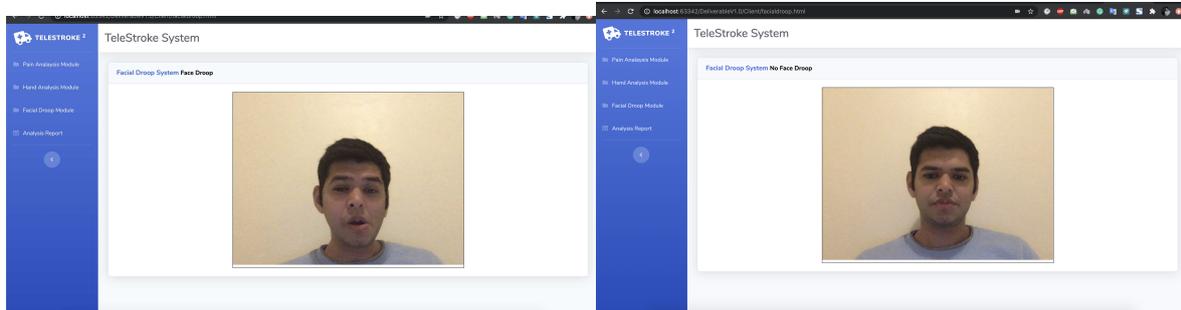


**Fig. 1.** Deep Client-Server Architecture

### 4.1 Front End

The front end is a web application comprised of HTML, CSS, JavaScript, AJAX, which ask the user's permission to grant for camera access that takes live stream

image of the user i.e., a WebSocket connection is established. From this stream, a picture is taken out, and the build model is applied. Ajax call is made to get the response and request faster.



**Fig. 2.** Face - Droop Image

**Fig. 3.** No Face - Droop Image

## 5 Experimental Evaluation

This section provides a detailed experimental evaluation of the Stroke recognition system using the TeleStroke Framework. All experiments are conducted on Google Colab with GPU using the Tensorflow 2.0.

The dataset is combined and viewed as one group of droopy as 0 and not droopy faces as 1. This dataset was transferred to different steps and tested on various algorithmic models.

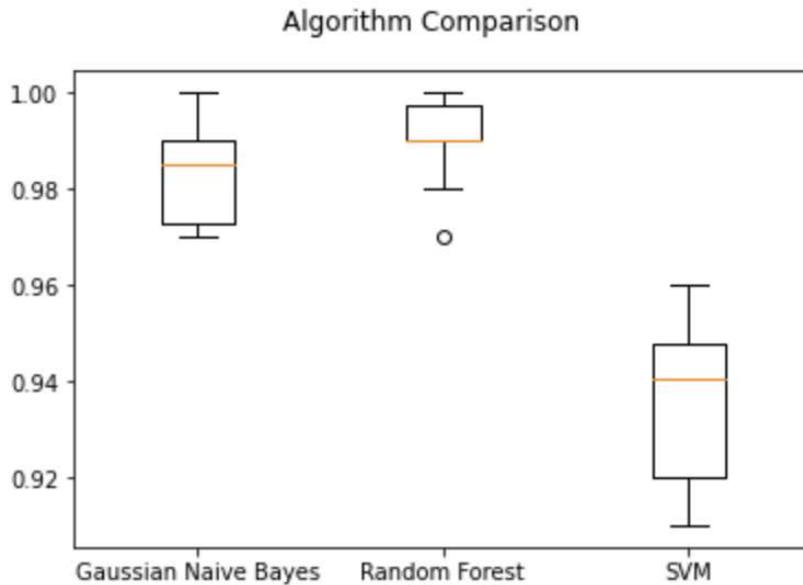
The evaluation methodologies are as follows:

### 5.1 Simple ML model

In this part, the Area ratio of the eye and Slope calculation of mouth plus eye and mouth coordinates are saved in CSV file, and the droopy faces are labeled as 0, and Normal Subjects are marked as 1 which will be used for classification.

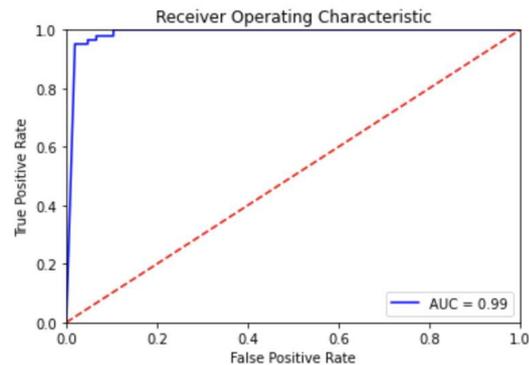
The Fig 4. shows the 10-fold cross-validation technique is performed to evaluate each algorithm on training data, which is critically designed with the same random seed to ensure that the equal splits are presented to the training data. Each algorithm is evaluated precisely the same way.

Fig 5 and 6 is the precision, recall, f1 score, and ROC Curve for the Naive Bayes as it was only tested with the testing data. It would indicate from these results of 10 cross-fold validation both Naive Bayes and Random Forest are perhaps worthy of further study on this topic.



**Fig. 4.** Algorithmic Comparison

	precision	recall	f1-score	support
0	0.95	0.93	0.94	109
1	0.94	0.96	0.95	142
accuracy			0.95	251
macro avg	0.95	0.95	0.95	251
weighted avg	0.95	0.95	0.95	251



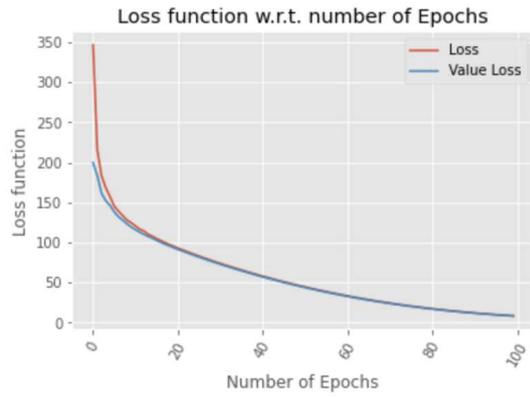
**Fig. 5.** Precision, Recall, F1-Score

**Fig. 6.** ROC Curve

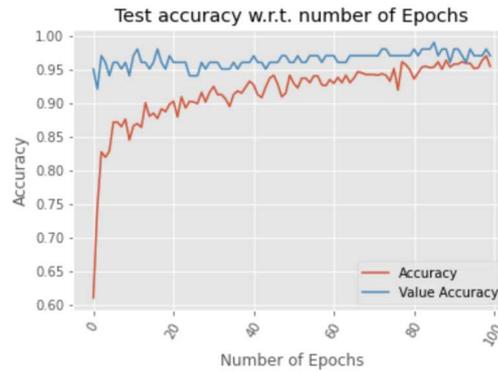
## 5.2 Fully Connected Neural Network.

The first layer has 600 neurons, the second layer has 400 neurons, and the third layer comprises of 200 neurons as shown in Fig 4.7 The lower the loss, the better a model will be (unless the model is overfitted to the training data). The loss is based on training and validation, and its definition for these two sets is how well the model is doing. Contrary to accuracy, the loss is not a number. It is a summation of the errors in training or validation sets that were made for each example.

There are also other subtleties when raising the value of the loss. For example, we may run into the problem of over-fitting, in which the model "memorizes" the patterns of training and becomes ineffective for the test set. We use Dropout for that reason.



**Fig. 7.** Loss Function

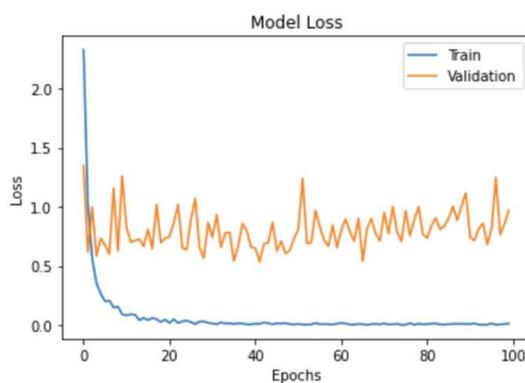


**Fig. 8.** Accuracy

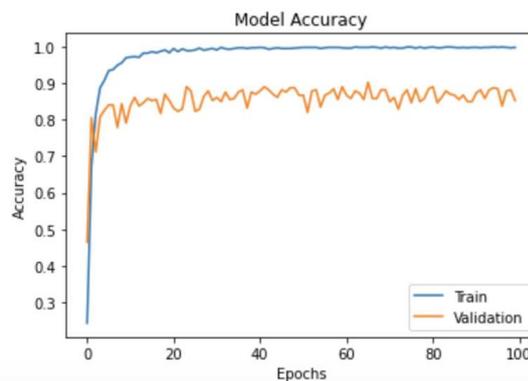
Name	Epoch	Accuracy	Loss
FCNN	100	0.95-0.97	0-2

### 5.3 CNN and VGG16 model

The amount of "wiggle" in the loss is proportional to the size of the batch. The wiggle should be reasonably large when the batch size is 1. If the batch size is the maximum data-set, the wiggle should be small, as each gradient update will monotonically increase the loss function (unless the learning rate is set too high).



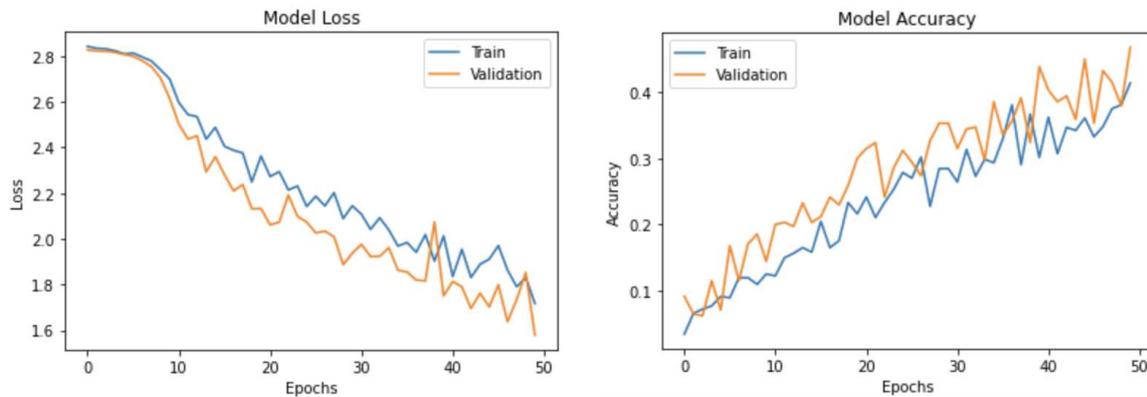
**Fig. 9.** Loss without augmentation



**Fig. 10.** Training loss without augmentation

The difference between the accuracy of the training and evaluation indicates how much overfitting. In Fig 7 and 8, we can observe that too much wiggle is formed, and the model starts to overfit at 0-20 range. The data is not augmented.

The yellow validation curve shows very low validation accuracy relative to the training's accuracy, suggesting heavy overfitting (note that validation accuracy can also begin to go down after some point). Compared to the above Fig 7 and 8



**Fig. 11.** Loss with augmentation

**Fig. 12.** Training loss with augmentation

Fig 9 and 10 is better and the wiggling is less and there is no sign of over fitting.

## 6 Conclusion

The proposed Web application requires incorporating four critical processes: image acquisition, key points extraction using OpenCV, mathematical measurement, and droopy mouth detection. This paper uses a Simple ML, Fully Connected NN, and CNN architecture with ResNet backbone to present a fully end-to-end framework named Telestroke System for droopy mouth detection. We tested the design using a combination of 3 datasets to get an F1 score of 94 percent, respectively, for the droopy mouth.

## 7 Discussion and Future Work

In this paper, we took on a novel and challenging task of assessing the different models with our proposed model. Although our model performed very well on the selected data sets, there are many areas where it is possible to pursue this research. Firstly, it is possible to achieve hyperparameter tuning and optimization by adding some complex feature vector representations; there is also the potential to investigate more complex backbone and deeper networks, which are limited in this work due to computational overhead 3D CNNs. It opens up a research opportunity for simplifying droopy mouth detection & recognition in Mobile Application.

The current dataset should be expanded to a more consistent one study by including a more substantial number of face photos labeled with demographic details, including age, gender, and ethnicity. As the data collection expands, additional knowledge will become more accurate, and we will consider studying the impact of original populations on paralysis disease and its ranking. Besides, the performance comparison of different pre-trained CNN models can be improved by using more precise facial landmark detection techniques and enabling the recognition of a droopy mouth from tilted or rotated images. Besides, the proposed model could be enhanced to achieve faster identification of droopy mouth and better recognition performance. Will be seen as a potential problem for research.

## 8 Acknowledgments

This research was conducted with the financial support of ADAPT Core (Platform & Spokes) under Grant Agreement No. 13/RC/2106 and at the ADAPT SFI Research Centre at Cork Institute Of Technology. Science Foundation Ireland funds the ADAPT SFI Centre for Digital Media Technology through the SFI Research Centres Programme. It is co-funded under the European Regional Development Fund (ERDF) through Grant 13/RC/2106.

## References

1. Stroke: a global response is needed <https://www.who.int/bulletin/volumes/94/9/16-181636.pdf>
2. MayField Clinic. 2013. What Is Stroke? [http://www.mayfieldclinic.com/pe-stroke.htm#.Va0rm\\_mqqko](http://www.mayfieldclinic.com/pe-stroke.htm#.Va0rm_mqqko)
3. National Stroke Association. 2016. Signs and Symptoms of Stroke
4. The Nature of Bell's Palsy. (Laryngoscope. 1949;59:228-235), R. Gacek, "Hilger
5. Vision based interface: an alternative tool for children with cerebral palsy, Magdalena Gonzalez, Debora Mulet, Elisa Perez, Carlos Soria, and Vicente Mut.
6. Video-based early cerebral palsy prediction using motion segmentation, Hodjat Rahmati, Ole Morten Aamo, Øyvind Stavdahl, Ralf Dragon, and Lars Adde
7. The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities, Margrit Betke, James Gips, and Peter Fleming
8. Learning patterns of activity using real-time tracking, Chris Stauffer and W. Eric L. Grimson
9. Human behavior prediction using facial expression analysis, Subarna Shakya, Suman Sharma, and Abinash Basnet
10. Real time facial expression recognition using realsense camera and ann, Jayashree V Patil and Preeti Bailke
11. Facial emotional expressions recognition based on active shape model and radial basis function network, Endang Setyati, Yoyon K Suprpto, and Mauridhi Hery Purnomo
12. Automated facial expression recognition based on face action units, James J Lien, Takeo Kanade, Jeffrey F Cohn, and Ching-Chung Li

13. A robust method for face recognition and face emotion detection system using support vector machines, KM Rajesh and M Naveenkumar
14. Mobile Health Awareness in Pre-Detection of Mild Stroke Symptoms, O.-M. Foong, J.-M. Yong, S. Sulaiman and D. Rambli
15. Dataset for Facial Palsy Kaitav Mehta =<https://www.kaggle.com/kaitavmehta/facial-droop-and-facial-paralysis-image>,
16. Deep Hierarchical Network With Line Segment Learning for Quantitative Analysis of Facial Palsy
17. Acquiring Linear Sub spaces for Face Recognition under Variable Lighting
18. World Health Organisation. 2014. Stroke, Cerebrovascular, [http://www.who.int/topics/cerebrovascular\\_accident/en](http://www.who.int/topics/cerebrovascular_accident/en)

# Exploratory Gait Analysis using Wearable Technology

Richard Davies<sup>1</sup>[0000-0002-2663-3979], Paul McCullagh<sup>1</sup>[0000-0002-9060-6262], Huiru Zheng<sup>1</sup>[0000-0001-7648-8709] and Joan Condell<sup>1</sup>[0000-0002-3517-2182]

<sup>1</sup> School of Computing, Ulster University, Jordanstown BT37 0QB, Northern Ireland

**Abstract.** Gait analysis is a rapidly expanding and evolving research area with application to biomechanics and rehabilitation. Current wearable technology that can be used to collect gait information is becoming more accessible in terms of cost and usability compared to lab counterparts which can be expensive and require training to setup and use. This paper provides an exploratory analysis of knee angle versus angular velocity of the lower leg for six healthy participants for at four walking speeds (1.8 km/h, 2.7 km/h, 4.5 km/h, 5.4 km/h) on a treadmill and walking over the ground at 3.9 km, using the Xsens Inertial Measurement Unit. These phase portraits provide a rich data source for qualitative comparison. In future objective gait analysis based on suggested gait parameters can aid with rehabilitation of dysfunctional gait.

**Keywords:** gait analysis, wearable technology, inertial motion capture, accelerometer, Xsens.

## 1 Introduction

Gait Analysis (GA) is an area of research that is continually expanding and evolving across a wide range of domains such as, healthcare, sport science and surveillance. There are a plethora of medical gait applications such as the evaluation of prosthetics, assessment of surgical procedures [1] [2], treatments plans, fall risk in the elderly [3] and assessment of neuropathies [4]. In addition GA has achieved further significance in the monitoring of elite athletes [5] and identification of individuals for forensic biometric purposes [6] [7].

During the past four decades the measurement and assessment of gait has evolved rapidly; tools and technology now provide an objective, quantitative evidence-based approach. Current clinical practice for motor assessment of the lower limb in stroke survivors is based upon a battery of tests, such as the two-minute walking test, timed-up and go, berg balance scale, fugl-meyer assessment, motor assessment scale, rivermead motor assessment of movement, motricity index and stroke rehabilitation assessment of movement. All of the aforementioned motor assessment scales predate the year 1997 and have an average age of 31 years. Although they provide a quantitative score they are based upon human clinical observation and are subject to inter- and intra-rater variability. Additionally, the majority of these assessment approaches are not capable of detecting subtle changes in motor function particularly at the top end of assessment scales as a ceiling effect often occurs [8].

Advances in technology used to measure gait have been instrumental in the evolution of GA. Biomechanical movement of the human body is complex and therefore effective GA requires information such as kinematics, ground reaction forces and influence of muscle activity. Motion Capture (MC) strives to measure kinematic data in an accurate, valid and unobtrusive manner. There are two competing MC technologies: optical capture and the use of force plates to measure plantar pressure. Each offers advantages and disadvantages depending on the context of the application being considered [9] [10].

### **1.1 Optical motion capture and force plates**

Optical Motion Capture (OMC) systems use cameras based upon active or passive markers to accurately detect the position of body worn markers within a 3-dimensional space. There are a number of commercial OMC systems available such as Vicon, Qualysis and Codamotion. These systems tend to be very accurate at sensing marker position to within the sub-millimeter range, however this accuracy heavily relies on the ability of a researcher to place markers accurately and repeatably. Various protocols exist to help locate joint centers but these differing conventions can produce a varied set of results [11]. Due to the complexity involved in camera setup and the configuration of software, OMC systems require a considerable amount of setup time and a need for specialised training. Hence OMC systems are more suited to static deployment in a dedicated gait laboratory, thus impacting upon the information derived as it may not be representative of gait in a real-world context [12]. OMC systems are expensive and occupy a static laboratory space which can be restrictive for particular applications. However, they do offer unparalleled accuracy when configured by a trained Biomechanist and serve as a gold standard or reference point for other less accurate systems.

Traditionally force plates were designed to record single steps with high accuracy and resolution. Pressure sensing technology has evolved through the incorporation of this technology into instrumented walkways such as the GAITRite mat facilitating GA of a sequence of steps [9] [10].

### **1.2 Wearable technologies and inertial motion capture**

Over the last 5 years, there have been significant advances in technologies for inertial motion capture (IMC) systems; in particular insole pressure sensor recording and the measurement and wireless transmission of the electromyogram (EMG). Insole pressure sensing technology has benefited from [13] advancements in microelectronics, wireless charging, energy harvesting, smaller batteries and low power wireless communication. These advancements have made insole technology more pervasive, embedding all of the technology within the insole e.g. Moticon [14]. These developments have paved the way for wearable technologies to replace gait laboratory equipment in the measurement of human kinematics, ground reaction forces and muscle activity. These wearable technologies offer a lower cost, portable, versatile, real-time and highly usable system to provide rich gait information in free-living environments for clinical GA [15], [16].

Inertial Motion Capture (IMC) systems offer benefits over OMC systems due to their portability, wearability and decreasing costs. IMC sensor units provide the

opportunity for more practical, untethered data capture free from the constraints of an indoor observation area. Deploying outside a gait laboratory environment can facilitate diverse spatial settings such as stairs, open space, more natural terrain or other indoor areas. The setup time is shorter repeatability for measurement of joint angles during walking [17] is better than OMC systems for both in-day and between-day recording sessions.

The fundamental assumption of an underlying rigid body such as the human skeleton can be violated by the movement of overlying soft tissue known as skin artefact. This is particularly evident when there is an excess of soft tissue or during highly dynamic movements [18]. Skin artefact is a common issue with IMC and OMC, however, it may be more pronounced with IMC systems as their sensor units are of greater physical size and mass leading to greater displacement. Skin artefact is an important factor in the context of clinical research as a higher than average body-fat index is more common in people suffering from stroke [17]. Ground reaction forces as measured by force plates provide an alternative or complementary means to perform GA.

## 2 Methodology

In this study we conducted exploratory GA on healthy participants ( $n=6$ ) using data gathered from a wearable IMC system. The GA focuses on a subset of parameters to assess feasibility, validity and diversity of gait amongst a healthy cohort. Baseline data were collected to establish a normal set of gait parameters. Factors that may contribute to variations in the data relate to but are not limited to gender, age, height and weight. Walking therapies are part of the rehabilitation pathway as defined by the NICE guidelines [19]; therefore the focus of this study is on walking on a treadmill and walking over ground while incorporating a turn.

In a data driven approach the quality and volume of the data being collected plays a vital role in the capability of a computational model to provide accurate, objective and repeatable assessment of gait. Therefore, it is important to control the recording of activities by applying a consistent clinical protocol. Although the number of participants is low ( $n=6$ ) the size of the dataset can still be adequate to provide sufficient quality of data as the number of steps can reach 2,400. In addition each step can provide further information angle, velocity for 7 locations on the lower body.

Participants wore an IMC system and a pair of smart insoles to collect kinematic and ground reaction forces during walking activities. Participants were required to complete a short anonymous questionnaire that provided information on their age, gender and weight.

**Inertial Motion Capture Configuration.** The IMC system was configured to investigate lower limb which involved donning 7 Inertial Measurement Units (IMUs) using a velcro based strapping system as shown in Fig. 1. The IMUs were attached to the pelvis (sacrum), left/right upper leg (thigh), left/right lower leg (shank) and left/right foot. A number of anatomical measurements (body height, shoulder width, arm span, hip height, hip width, knee height, ankle height, foot size, shoe sole height) were taken from

the participant to help complete the calibration. The IMC was re-calibrated before each recording session.



**Fig. 1.** Xsens inertial capture system showing sensor positions for foot, lower and upper limb.

**Participants.** A group of 6 healthy adults participated in the study. A summary of demographic information is provided in Table 1.

**Table 1.** Demographic information of participants

Participant ID	Gender	Age	Height (cm)	Weight (Kg)
1	Male	61	179	68.7
2	Female	52	158	N/A
3	Male	26	186	N/A
4	Male	71	172	73.0
5	Female	71	164	64.0
6	Male	N/A	177	N/A

Initial feasibility testing was conducted to evaluate at an observational level that the wearable systems were fit for purpose in terms of robustness, reliability, usability, comfort, repeatability and set-up time. Walking activities have been designed to include walking at differing speeds, walking on a treadmill versus over the ground. The study also included turns as this is an important constituent of GA.

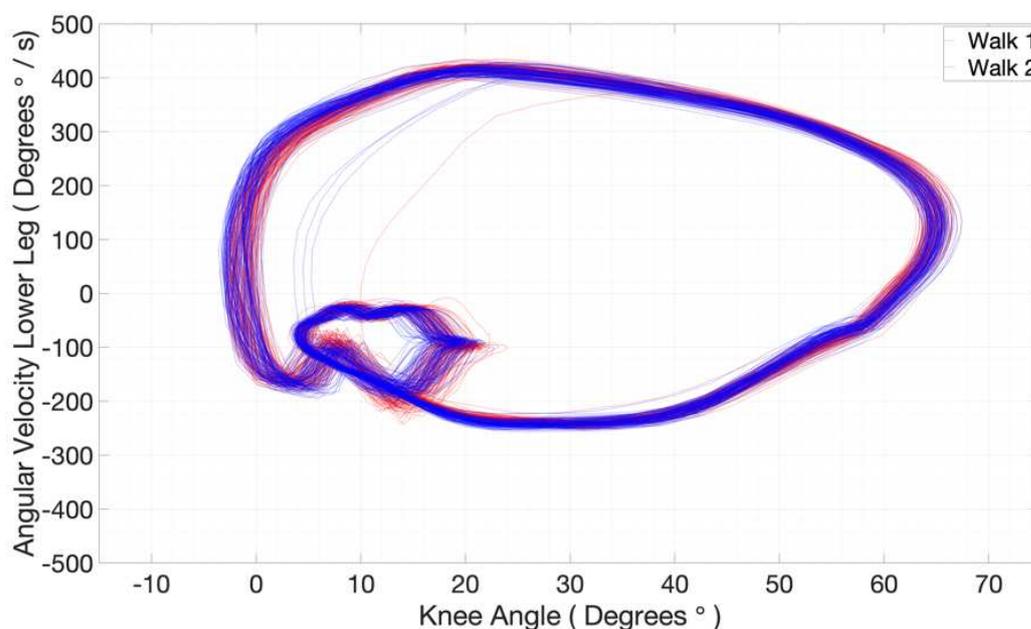
Given that the potential use case scenario for this research will be typically elderly post-stroke survivors some slower walking speeds were included. Participants performed two walking activities. The first involved walking indoors over the ground on a flat smooth surface within a gait laboratory environment for a distance of 10m and turning. This activity was repeated for 2.5 mins at a comfortable walking speed, self-selected by the participant. The second activity required participants to walk for 2.5 minutes at 4 different speeds on a treadmill with zero incline for a total walking time of 10 minutes. The four speeds using the treadmill were: very slow (0.5m/s), slow (0.75m/s), medium (a comfortable speed self-selected by the participant, either 1m/s or

1.25m/s) and fast (1.5m/s). The fastest walking speed test is similar to the 2-minute walking test which is a clinical assessment.

### 3 Results

The Xsens system has already been well validated against OMC systems and has been reported to have a coefficient of multiple correlation  $> 0.96$  for all joints during flexion/extension for level walking activities [20]. However, configuration, calibration and positioning sensors can have varying effects on the quality of the data collected. Repeatability was explored for in-day testing of the Xsens without doffing/donning the sensors.

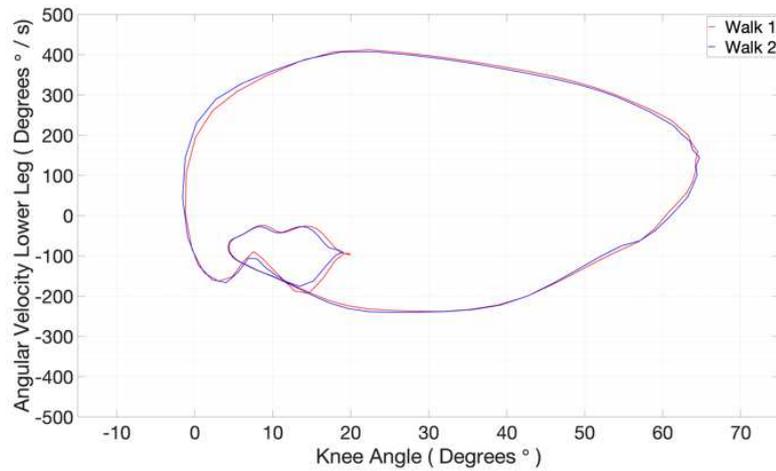
The two phase portraits shown in Fig.2 are highly correlated and present knee angle against angular velocity of the lower limb for participant 1. These phase portraits show the dynamic nature of the knee during walking on a treadmill at a 5.4 m/s for 2 minutes. A single gait cycle is represented by one phase which can be seen as a closed loop. The phases are plotted on top of each other as each gait cycle is repeated, it shows high levels of correlation but with some dynamic and chaotic variations. The variation between gait cycles in the first test can be seen in red while the blue lines show the variation of gait cycles in the second test. Since both tests were recorded within 30 minutes and under the same conditions the variation which is expected to be minimal between walks can be observed by comparing red and blue lines.



**Fig. 2.** A highly correlated phase portrait of knee angle versus angular velocity of the lower right leg for Participant 1 during two separate tests while walking on treadmill at 5.4 m/s for 2 minutes.

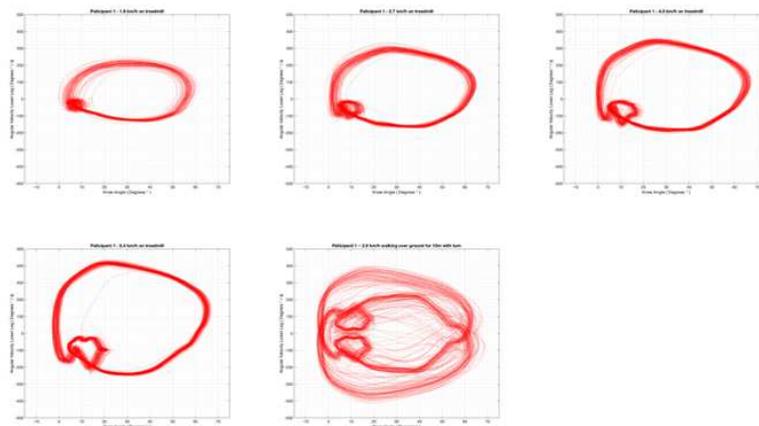
To provide a statistical measure of correlation an average gait cycle was computed for both walks and a correlation coefficient calculated by comparing both average gait

cycles. The average phase portraits can be seen in **Fig. 3**, these are highly correlated as expected ( $r=0.9993$ ), this is the Pearson correlation coefficient as reported by MATLABs 2-dimensional correlation function.



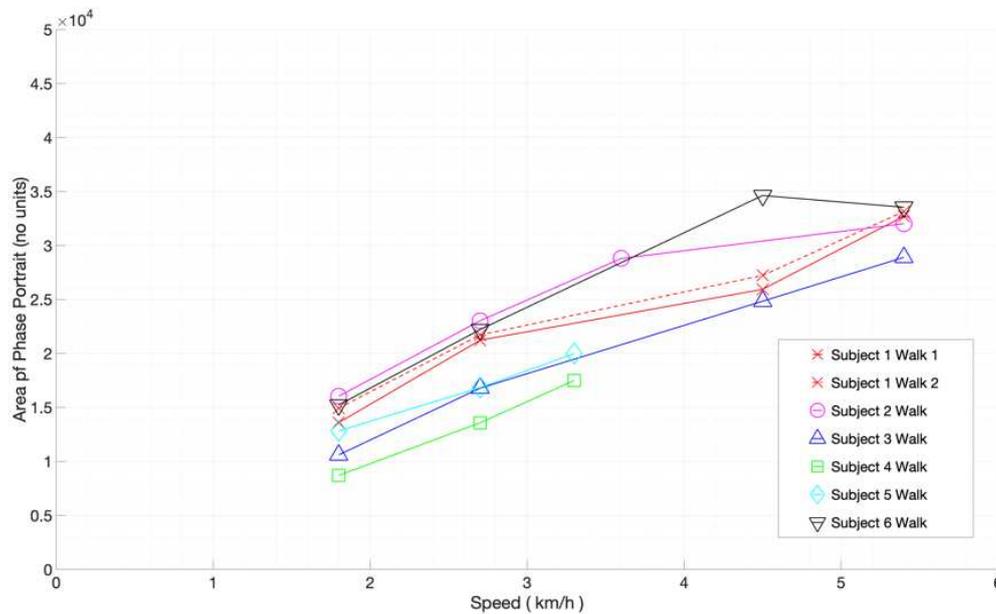
**Fig. 3.** A highly correlated ( $r=0.9993$ ) phase portrait, it represents an average gait cycle from two walks by participant 1.

The next stage of GA was to compare all of the walking activities for participants across a number of different walking speeds; this involved 2 minute treadmill walks at 1.8 km/h, 2.7 km/h, 3.6 km/h or 4.5 km/h and 5.4 km/h and a final 2 minute walk over the ground for 10 metres with a turn. The phase portraits of these walking activities can be seen in **Fig. 4** for participant 1. Initial observations show correlation between walking speed and the area enclosed within the phase portraits. A greater range of motion and increased angular velocity should result in an increased area within the curves. Additionally, as the speed increases the variability in the phase portraits reduces to produce a more rhythmic and stable gait cycle, this is particularly for true for walking activities on the treadmill. It seems that this effect is caused by a combination of the treadmill and higher walking speeds.



**Fig. 4.** Phase portraits of knee angle versus angular velocity of the lower right leg for Participant 1 at four walking speeds (1.8 km/h, 2.7 km/h, 4.5 km/h, 5.4 km/h) on a treadmill walking over the ground at 3.9 km/h.

To quantify the relationship between the area of the phase portraits and speed an average phase portrait was computed for each walk. These were then used as a basis to calculate an average area of each phase portrait and to plot these against speed to quantify any relationship and how it may change across the cohort. Observation of **Fig. 5** shows a common pattern of increased area equating to increased speed. Participants 4 and 5 were both in their early seventies and yet it is interesting that they were able to maintain walking speeds of 1.8 km/h, 2.7km/h and 3.3km/h with a reduced area, this may imply a greater sense of control by reducing the stride length.



**Fig. 5.** Walking activity for six participants showing average phase plot area against speed.

Comparing treadmill walking activities against over the ground walking in **Fig. 4** shows that the variation in knee angle is more apparent while walking over ground. As the walking over the ground activity included turning 180 degrees every 10 metres there are a significant amount of turns ( $n \approx 16$ ) within a 2 minute period. Therefore it makes it more difficult to attribute the increased knee angle variation as a direct result of walking over the ground. The second noticeable feature of the phase portraits for walking over the ground is that there is a mirroring effect which results in two prominent distinct phases. These are a direct result of walking in two opposite directions and may be combined into a single phase if it is possible to adjust the data to accommodate walking direction. This would be a useful analysis feature as it would allow all walking activities to be analyzed and compared irrespective of walking direction.

## 4 Future Work

This paper provided a demonstration that repeatable and interpretable gait analysis is possible using wearable IMC technologies. Phase cycles and repeatability were assessed by observation and quantitative measurement. Further work will be conducted

with more healthy participants and incorporate between-day and inter-rater reliability. There is a significant opportunity to quantify GA through the measurement of spatial, temporal, spatiotemporal and other phasic parameters, as listed in Table 2. Additional features that can be used in assessing the repeatability of normal gait can be derived from the normative dataset. Using these measurements and features will facilitate gait modelling for a healthy population. A gait model for a healthy population can be used to provide a reference to any new gait information that should be compared from a disease specific cohort such as stroke survivors.

There are a number of use cases where gait analysis can be beneficial for both the clinician and patient such as Parkinson's disease, cerebral palsy, lower-limb osteoarthritis, post-stroke and diabetic neuropathy. The authors propose to develop a computational gait model based on intelligent data analysis using non-linear techniques to provide an objective and quantifiable assessment of pathological gait for post-stroke that is accurate, robust and repeatable.

**Table 2.** List of gait features for future work.

<b>Spatial</b>	<b>Temporal</b>	<b>Spatiotemporal</b>	<b>Phasic</b>
Step Length (cm)	Cadence (steps/min)	Gait Speed (m/s)	Stance Time (%GC)
Stride Length (cm)	Step Time (s)	Stride Speed (m/s)	Swing Time (%GC)
Step Width (cm)	Stride Time (s)	Stride variability	SST (%GC)
Step Height (cm)	Stance Time (s)	Smoothness	DST (%GC)
Knee Angle (°)	Swing Time (s)	Centre of Pressure	
Hip Angle (°)	Single Support Time (s)		
Ankle Angle (°)	Double Support Time (s)		

## 5 Conclusion

Gait analysis of the normal population and of different pathologies is an area of research that is expanding rapidly. There are a number of competing technologies that can provide gait information, two such competing sets are research gait lab technology and wearable technology. The former tends to be more expensive, less flexible and with longer setup times often requiring specialised training. With recent advances, wearable technology can offer a cheaper, more accessible, less restrictive and easier to use option without comprising on the accuracy or quality of the information. This is particularly true of recent advances of IMC systems.

The Xsens IMC system captured kinematic data from walking. Due to the exploratory nature of this study only the dynamic nature of knee angle during walking was considered; the gait variation across a number of walking speeds on a treadmill and walking over the ground and gait variation across the population were assessed. Future

research aims to build a computational model that can be used to assess a user's gait during ambulation. A large set of features will be generated to serve as input to the model and as such can be configured in multiple ways via feature selection to ascertain the optimal model and as a result what are the optimal features, technologies and sensors. There is a significant body of research to suggest that spatial temporal gait parameters provide such a feature set [21], [22], [23].

## 6 Acknowledgements

The authors would like to acknowledge the School of Computing at Ulster University and SENDoc (www. Sendocnpa.com) project in the ongoing support of this doctoral research.

## References

1. A. Heikkilä, N. Sevander-Kreus, A. Häkkinen, M. Vuorenmaa, P. Salo, P. Konsta and J. Ylinen, "Effect of total knee replacement surgery and postoperative 12 month home exercise program on gait parameters," *Gait & Posture*, vol. 53, pp. 92-97, 2017.
2. P. Ornettia, J.-F. Maillefert, D. Laroche, C. Morisset, M. Dougados and L. Gossec, "Gait analysis as a quantifiable outcome measure in hip or knee osteoarthritis: A systematic review," *Joint Bone Spine*, vol. 77, pp. 421-425, 2010.
3. M. Zagoa, C. Sforza, D. R. Bonardi, E. E. Guffanti and M. Galli, "Gait analysis in patients with chronic obstructive pulmonary disease: a systematic review," *Gait & Posture*, vol. 61, pp. 408-415, 2018.
4. M. E. Fernando, R. G. Crowther, P. A. Lazzarini, K. S. Sangla, P. Buttner and J. Golledge, "Gait parameters of people with diabetes-related neuropathic plantar foot ulcers," *Clinical Biomechanics*, vol. 37, pp. 98-107, 2016.
5. B. Hanley, C. B. Tucker and A. Bissas, "Differences between motion capture and video analysis systems in calculating knee angles in elite-standard race walking," *Journal of Sport Science*, vol. 36, no. 11, pp. 1250-1255, 2018.
6. I. Bouchrika, M. Goffredo, J. Carter and N. Mark, "On Using Gait in Forensic Biometrics," *Journal of Forensic Sciences*, vol. 56, no. 4, pp. 882-889, 2011.
7. Y. He, J. Zhang, H. Shan and L. Wang, "Multi-Task GANs for View-Specific Feature Learning in Gait Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 102-113, 2019.
8. D. J. Gladstone, C. J. Danells and S. E. Black, "The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties," *Neurorehabilitation and Neural Repair*, vol. 16, no. 3, pp. 232-240, 2002.
9. S. S. Kuys, S. G. Brauer and L. Ada, "Test-retest reliability of the GAITRite system in people with stroke undergoing rehabilitation," *Disability and Rehabilitation*, vol. 33, no. 19-20, pp. 1848-1853, 2020.

10. A. Scully and D. Tan, "Within-Assessment Variability of the GAITRite System in People With Stroke," *Physical Medicine & Rehabilitation*, vol. 98, no. 6, pp. 446-449, 2019.
11. A. Ferrari, M. G. Benedetti, E. Pavan, C. Frigo, D. Bettinelli, M. Rabuffetti, P. Crenna and A. Leardini, "Quantitative comparison of five current protocols in gait analysis," *Gait Posture*, vol. 28, no. 2, pp. 207-216, 2008.
12. A. Ali, K. Sundaraj, B. Ahmad, N. Ahamed and I. Anamul, "Gait disorder rehabilitation using vision and non-vision based sensors: A systematic review," *Bosnian journal of basic medical sciences*, vol. 12, no. 3, pp. 193-202, 2012.
13. T. P. Kakarla, K. A. Varma, P. S. P. Jayaraj Joseph and M. Sivaprakasam, "Accuracy Enhancement of Total Force by Capacitive Insoles," in *IEEE International Symposium on Medical Measurements and Applications*, Istanbul, 2019.
14. M. S. Oerbekke, M. J. Stukstette, K. Schütte, R. A. de Bie, M. F. Pisters and B. Vanwanseele, "Concurrent validity and reliability of wireless instrumented insoles measuring postural balance and temporal gait parameters," *Gait & Posture*, vol. 51, pp. 116-124, 2017.
15. J. Lee, S. Y. Shin, G. Ghorpade, T. Akbas and J. Sulzer, "Sensitivity comparison of inertial to optical motion capture during gait: implications for tracking recovery," in *IEEE 16th Internal Conference on Rehabilitation Robotics*, Toronto, 2019.
16. C. Cao, M. I. Rodríguez-Ferradas, A. Cazón-Martin and P. Morer, "Development of a Low-Cost Wearable Prevention System for Musculoskeletal Disorders Using Inertial Measurement Unit Systems," in *Advances on Mechanics, Design Engineering and Manufacturing II. Lecture Notes in Mechanical Engineering*, Cham, Springer, 2019.
17. T. Cloete and C. Scheffer, "Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems," in *Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems*, Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems, 2010.
18. A. Cappozzo, F. Catani, A. Leardini, M. G. Benedetti and D. U. Croce, "Position and orientation in space of bones during movement: experimental artefacts," *Clinical Biomechanics*, vol. 11, no. 2, pp. 90-100, 1996.
19. "National Institute for Health and Care Excellence Stroke rehabilitation: long-term rehabilitation after stroke. Clinical Guideline 162," NICE, 2013. [Online]. Available: [www.nice.org.uk/guidance/cg162](http://www.nice.org.uk/guidance/cg162). [Accessed 31 08 2020].
20. J.-T. Zhang, A. C. Novak, B. Brouwer and Q. Li, "Concurrent validation of Xsens {MVN} measurement of lower limb joint angular kinematics," *Physiological Measurement*, vol. 34, no. 8, pp. 63-69, 2013.
21. J. H. Hollman, M. K. Watkins, A. C. Imhoff, C. E. Braun, K. A. Akervik and D. K. Ness, "A comparison of variability in spatiotemporal gait parameters between treadmill and overground walking conditions," *Gait & Posture*, vol. 43, pp. 204-209, 206.
22. O. Beauchet, G. Allali, H. Sekhon, J. Verghese, S. Guilain, J.-P. Steinmetz, R. W. Kressig, J. M. Barden, T. Szturm, C. P. Launay, Grenier, L. Bherer, T. Liu-Ambrose, V. L. Chester and M. L. Callisaya, "Guidelines for Assessment of Gait and Reference Values for Spatiotemporal Gait Parameters in Older Adults: The Biomathics and Canadian Gait Consortiums Initiative," *Frontiers in Human Neuroscience*, vol. 11, p. 353, 2017.

23. R. Briggs, D. Carey, P. Claffey, T. McNicholas, O. Donoghue and S. P. Kennelly, "Do Differences in Spatiotemporal Gait Parameters Predict the Risk of Developing Depression in Later Life?," *Journal of the American Geriatrics Society*, vol. 67, no. 5, pp. 1050-1056, 2019.



## Chapter 5

---

# Business and Society



# Requirements for payment systems in e-commerce from the retailers' perspective

Lisa Zuyeva<sup>1</sup>, Ingo Stengel<sup>2</sup> and Stefanie Regier<sup>3</sup>

<sup>1</sup> Deutsche Bank AG, 60325 Frankfurt am Main, Germany

`lisa.zuyeva@web.de`

<sup>2</sup> Karlsruhe University of Applied Sciences, 76133 Karlsruhe, Germany

`ingo.stengel@hs-karlsruhe.de`

<sup>3</sup> Karlsruhe University of Applied Sciences, 76133 Karlsruhe, Germany

`stefanie.regier@hs-karlsruhe.de`

**Abstract.** The selection and use of payment methods is an important success factor for online retailers, as many customers have already cancelled a purchase because their preferred payment method was not offered in the online shop. Online retailers now have a wide range of different payment methods available to them, which they can offer their customers for payment. In order to choose the right payment method, various requirements for e-commerce payment systems must first be determined. In addition, current EU regulations result in legal requirements that payment systems must meet, so that this also affects the requirements of the retailers. The objective of the paper is to identify, describe and evaluate the requirements for payment systems in e-commerce from the retailers' perspective. This should provide the basis for the development of a model for the evaluation of electronic payment procedures. The evaluation of the identified requirements was based on an empirical study with expert interviews. It was determined that high distribution/acceptance by customers and user-friendliness are the two most important requirements for payment systems from the retailer's point of view, which contribute to sales development. A high level of payment security and low total costs are also important requirements, but were rated as less relevant by the experts surveyed. In addition, the legal regulations of the European General Data Protection Regulation (GDPR) and Payment Service Directive 2 (PSD2) also resulted in legal requirements that payment systems must meet.

**Keywords:** payment systems, e-commerce

## 1 Introduction

In today's digital and globalised world, e-business is playing an increasingly important role. In addition to stationary sales, more and more retailers are offering their goods and services online in an e-shop or are concentrating entirely on Internet sales. While

in 2009, sales in the online and mail-order business accounted for around 7.2% of total retail sales in Germany, sales in 2017 were already reaching 13.2% [1]. For consumers, shopping on the Internet is already part of everyday life. More than half of all Germans today order their goods and services via the Internet, with around 50 million online buyers [2]. The increasing number of online users and thus potential online buyers is driving the steady growth of the German e-commerce market. A consumer study conducted by the Federal E-Commerce and Mail Order Association (BEHV) found that in 2018 a gross turnover of 65.1 billion euros was achieved with e-commerce goods throughout Germany [3]. Payment is an essential part of the ordering process. The selection and use of payment methods is an important success factor for online retailers, as 63% of consumers have not made a purchase at some point because the payment methods offered were not sufficient [4]. A challenge in online trade arises from the fact that ordering, delivery and payment are decoupled in time and space, whereas in stationary trade the goods are paid for directly [5]. Online retailers today have a large number of different payment methods at their disposal, so they are faced with the question of which payment methods they should offer their customers for payment processing [7]. In order to choose the right payment method, various requirements for payment systems in e-commerce must first be identified. This work focuses primarily on the retailer perspective, although customer requirements also influence this. A further reason is that the selection of payment methods that are not accepted by customers generally do not bring the desired success for retailers [8]. In addition, current legal EU regulations also result in legal requirements that payment systems must fulfil, thus this also affects the requirements of retailers. The objective of the paper is to identify, describe and evaluate the requirements for payment systems in e-commerce from the retailers' point of view. This provides a basis for the development of a model for the evaluation of electronic payment procedures.

## 2 Related Research

Since every payment represents a transaction, the basic technical requirements of a transaction must be reviewed for each new electronic payment procedure [7]. The requirements for a transaction include Atomicity, Consistency, Independence, and Durability (ACID). These four ACID properties together form the requirements for a transaction system known in computer science [7, 9]. By fulfilling the ACID properties, the successful execution of payment transactions can be guaranteed. Since the transactions of payment systems in e-commerce refer to large distances in the anonymous digital space, a much higher level of security must be provided than, for example, in the payment process in stationary retail trade [10]. In particular, the electronic payment procedures must ensure adequate protection against attacks, misuse or manipulation of data and financial transactions on the Internet [9]. In e-business, the (perceived) security is

also an absolute knockout criterion, because customers only accept offers from operators that can credibly exclude the misuse and manipulation of their data [10]. Protection against misuse and manipulation is an important criterion, because in addition to the direct damage, trust in electronic payment systems can be lost [11]. Data and transaction security in electronic payment transactions is perceived as given and is measured when a payment system meets the criteria of availability, confidentiality, integrity, authenticity and authorization [9, 10, 12-14]. In addition to the security requirements described above, electronic payment systems must be flexible in its use and be widely used and accepted by both customers and retailers [13, 15]. The flexibility of a payment system must be given from a technical, economic and geographical point of view [15]. The acceptance or willingness to accept a new payment procedure must be guaranteed by the players involved, so that misinvestments or loss of revenue can be prevented [9]. Acceptance is also related to the degree of penetration, which describes the density of participants in the payment system [9]. The interaction between the actors is a necessary condition for the successful dissemination of an Electronic Payment System (EPS) [16]. The critical mass of customers and retailers is crucial for the success of a payment system, because if this is not achieved, there will be too few users and the EPS will not be able to spread further in the market [16]. This means that acceptance by users increases with the number of participants [9]. In fact, after reaching a certain critical mass, the number of users grows exponentially until saturation is reached [16]. Furthermore, the current state of research shows that the acceptance of an EPS is an important and frequently researched area. Accordingly, acceptance and distribution is an important criterion for the selection of EPS. Since there must be enough users on both sides of the market for a payment method to be offered as a payment option in an online shop, the needs of both retailers and consumers must be considered. For a retailer, it is only worthwhile entering a particular EPS if there are currently enough customers using it or at least in the near future. Whereas customers choose the systems that are accepted in most online shops [7, 17].

From the customer's point of view, there are a number of requirement criteria that are decisive in determining why they choose a particular payment method for their online purchase. The results of the study conducted by the E-Commerce-Center Handel at the „Institut für Handelsforschung“ showed that consumers demand in the first place a high level of security, user-friendliness, distribution and cost of a payment procedure [18].

From the retailer's point of view, there are also some requirement criteria that are decisive for the selection of the payment methods they want to offer in their online shop. The aim of retailers is to secure and increase their sales [8]. The current situation in the e-commerce market is positive, since it continues to show double-digit growth and the e-commerce sales will continue to increase in the future [3, 19]. Retailers should be able to access and collect these revenues flexibly, easily and securely [19]. Payment systems in this context should offer a high level of security, be flexible with regard to

the amount of the sum involved, be suitable for international use and be platform or multi-channel capable. In addition, the wide distribution and acceptance to customers was also mentioned. This is influenced by the familiarity of payment systems and the payment habits of the customers, as well as by user-friendliness and simple handling. A high level of payment security also plays an important role in generating and securing sales, i.e. online retailers demand protection against non-payment and want to minimize the associated risk to a calculable level [8, 19]. The reduction of the total costs of a payment procedure and close customer relationships are also sales factors and influence a retailer's profit [8, 19, 20].

### 3 Requirements for Payment Systems from Retailers' Perspective

As already mentioned, part of the paper's objective is to identify the requirements for payment systems in e-commerce from the retailer's perspective. These requirements were determined on the basis of the literature study. In connection with the goal of a high turnover and the use of a payment method, requirements from the retailer's point of view can be divided into the following three basic property categories: revenue generation, revenue protection, costs [8, 21]. In Table 1, the essential requirements from the retailer's point of view are assigned to these property categories. The assignment of the requirements to the three categories is based on the study by van Baal & Hinrichs [8]. In addition, legal requirements are included as a fourth category.

<b>Development of Sales Revenue</b>	
Usability/Ease of Use	High Distribution/Acceptance by Customers
International Applicability	Platform Independence
Reputation & Awareness of a Payment System	
<b>Assurance of Sales Revenue</b>	
High Payment/Reliability/Low Risk of Non-Payment	
<b>Costs</b>	
Cost-effectiveness/Low Total Costs	Simple Integration Capability/Low Integration Effort
Early Date of Payment/Rapid Receipt of Payment	Reduction of Internal, Manual Processes
<b>Statutory Provisions</b>	
Data Privacy/Protection of Personal Data	Data Security/Protection against Misuse

*Table 1: List of requirements for payment systems in e-commerce from the retailers' perspective*

### **a) Development of Sales Revenue:**

These requirements are essential for the number of customers reached by a payment system and for the number and amount of possible transactions. For retailers, it is a necessary condition that a sufficient number of potential customers can pay in the online shop and are not excluded from the transaction because, for example, there is no "suitable" payment method available. In this case, the customer would then most likely abandon the checkout process and order his or her goods in another online shop that is only a mouse click away and offers his or her preferred payment method. The requirements assigned to the category of revenue generation have a strong customer reference. The user-friendliness or simple handling of a payment method also includes in particular the features of easy registration for new customers, low usage effort, fast and convenient handling as well as transparency and traceability. For online retailers, this requirement is relevant in order to enable a smooth order or checkout process for customers, so that they can complete the purchase and do not cancel it due to the cumbersome use of a payment system. Another important requirement for retailers to generate sales is a high level of distribution and acceptance among customers. Only if the retailer offers the well-known and popular payment methods of the customers in his online shop, customers will buy goods there and the retailer will achieve a high conversion rate. The reputation and familiarity of a payment system is closely linked to customer acceptance, which is also included here as a requirement of the retailers. Retailers are more willing to offer a well-known payment system, since such a system is usually widely used. In addition, a good image of the payment provider leads to higher customer acceptance. Furthermore, the flexible applicability of a payment system is also a relevant characteristic in terms of revenue generation. This includes the requirements for international applicability and platform independence, as this enables online retailers to reach more potential customers and target groups for their online shop.

### **b) Assurance of Sales Revenue:**

This includes features that directly or indirectly influence the security of the payment receipt for the retailer. The actual receipt of the customer's financial consideration can be prevented by system errors or fraudulent intentions.

In the category of sales revenue assurance, the requirement for high level of payment security is very important. For online retailers, it is crucial for their business and existence that the payment amounts requested are received by them. A high level of payment security of a payment procedure includes, among other things, the reduction of payment defaults and the non-repudiation of the payment. With regard to the reduction of payment defaults, payment systems must be examined to determine whether certain risk checks and security measures are necessary on the retailer side. This represents an additional cost factor. Non-repudiation of payment refers in particular to attempts and

cases of fraud on the customer side and cases of abuse by third parties. Here the question of liability plays an important role.

### **c) Costs:**

These include one-off and ongoing costs incurred by the trader, including opposition costs. Retailers want to ensure that the payment process runs smoothly and cost-effectively.

Various costs are incurred by the retailer when using a particular payment method. The lowest possible total costs of a payment method are decisive in determining whether a retailer will include it in its online shop. The total costs include one-time costs such as setup and integration costs, direct costs such as transaction-independent and transaction-dependent costs, costs of risk and receivables management, opportunity costs due to delayed incoming payments, costs due to service disruptions and costs due to returns [19, 22, 23]. Regarding integration costs, retailers have to consider the integration effort, i.e. whether the payment procedure can be easily integrated into the shop system and into other internal systems such as the accounting system. For example, retailers can integrate some payment methods into their shop system using plug-ins without additional programming effort. Payment methods can also be connected directly via PSP using interfaces. The reduction of internal, manual processes also plays a role in the cost criterion, especially with regard to billing and returns. For retailers, the degree of automation of the payment system is important, e.g. some invoices must be assigned manually, which will cause additional work and costs. In addition, the cost category also includes the requirement for a fast receipt of payment, which influences payment security, but is assigned to the cost category by van Baal & Hinrichs [8]. The background to this is that not negligible costs are incurred, for example, for the storage of the ordered goods until dispatch. Accordingly, the earlier or faster a payment is received by the retailer, the earlier he can ship the goods.

### **d) Legal requirements:**

Current legal regulations result in legal requirements that payment systems must fulfil. For example, EU- GDPR, which came into force in 2018, and the PSD2 have led to far-reaching changes in e-commerce and also in the payment sector within the EU.

This includes the requirements for data protection and protection of personal data and for data security and protection against misuse. Payment service providers as well as providers and online retailers must implement these two requirements in a legally binding manner in accordance with GDPR and PSD2 (implemented into German law in the ZAG). Among other things, secure encryption of data during transmission, entry of data in a protected environment and authentication play an important role. In addition, the protection of personal customer data also includes their anonymisation or pseu-

donymisation, which is also related to data economy. The requirement for data minimization for simplification is covered by the requirement for data protection. For the examination of a payment procedure, data protection also takes into account which and how much data the procedure or the provider needs and collects, and to how many and to which third parties these data are passed on. In this context, it is also relevant whether the payment provider is at least an European company, as the understanding of data protection is present and is regulated more strictly by law (GDPR). This allows retailers to collect and process data for a specific purpose. The requirement for data security describes how well a payment procedure is protected against abuse and fraud.

## 4 Methodology and Evaluation

The identified catalogue of requirements from the retailer's perspective forms the initial basis for the evaluation of different payment systems in e-commerce. To be able to compare and evaluate payment procedures, it is necessary to identify which requirements are actually relevant from the retailer's point of view and how important they are. The evaluation of the requirements was carried out on the base of a qualitative investigation with the help of expert interviews. Therefore a semi-structured guideline interview was selected as the method of collection, since requirements had already been identified in the testing phase of the interviews. This was also suitable for evaluating the requirements, since the experts were asked to select the degree of importance per requirement. In addition, the experts were able to make additional remarks and thus also specify requirements. During the course of the interview, it was possible to ask questions or additional questions based on the respective interview partner and his expertise. A total of four interviews with experts in the fields of e-banking, e-payment and e-commerce were conducted, recorded, transcribed and evaluated within a period of five weeks.

The evaluation of the interviews was based on the qualitative content analysis according to Mayring [24]. In the first step of the qualitative content analysis, the central analysis units were determined. Here, the categories of the guide were defined as analysis units. In the second step, the main text passages of these five categories were paraphrased. Decorative, repetitive, clarifying text components were deleted, e.g. "hm". The third step involved the generalisation of the experts' paraphrased statements. Since the experts have answered the questions precisely and clearly, the level of abstraction here is low. In the fourth step, generalised paraphrases with the same meaning were deleted, which is the first reduction phase. The second reduction phase took place in the fifth step, in which several paraphrases, which refer to each other and were often distributed through the transcribed material, were bundled and summarised and reproduced by a new statement. Accordingly, the already reduced contents of the categories were reduced once again by a higher level of abstraction. Step six contains the compi-

lation of the new central statements of the corresponding categories. Finally, in the seventh step, the summarised category contents were compared again with the source material in order to correct possible errors [24].

## 5 Results

In the course of the interviews, the experts were asked how important they consider the requirements listed in Table 1 to be from the dealer's point of view and which of the requirements they consider to be the most important. The experts were able to choose between four options: "very important", "important", "less important" and "unimportant". An even number of options was chosen so that the experts could not give a neutral assessment, so that they had to decide between important and unimportant and thus a tendency could be identified. All experts identified the requirements for a high level of dissemination and acceptance among customers as very important, whereby for three of the four experts this is the most important requirement for online retailers. In their opinion, the customers ultimately decide which payment methods the online retailer must offer in order to sell his goods successfully. According to expert C, the cost of the process is less important than customer acceptance, because a turnover with a poor margin is still better than no turnover at all. This statement refers to the case when the customer abandons the purchase during the checkout process because his preferred method of payment is not offered. As a result, Online retailers are forced to offer expensive payment methods if they are popular and widely used by customers. Expert B, on the other hand, considers user-friendliness and ease of use to be the most important requirements. After all, an online retailer wants to sell his goods and the threshold for this should be as low as possible for the customer, which is why a payment system should be user-friendly and easy to use. The retailer benefits if the customer can pay for the ordered goods easily and quickly in an unobstructed payment flow. Overall, all four experts rated user-friendliness and simple handling as a very important requirement.

The experts also agree on the requirement for international applicability. They rate this requirement as less important. Many online shops focus on only on regional customers. There are only a few retailers who sell cross-border, so that for most online retailers the international applicability of a payment procedures plays a secondary role. In contrast, payment habits and the payment methods offered are different in every country. There are payment methods that are widespread regionally but not relevant internationally or in other countries, such as the popular purchase on account in Germany. Thus, international applicability is not considered as a requirement criterion in the evaluation of the selected payment systems.

A simple integration capability or a low integration effort is considered by the experts to be a rather important requirement. Only for expert B, this requirement is less

important. For all online shops, it is an integration effort when a seldom used payment procedure is introduced.

In addition, some payment methods can be switched on or off relatively quickly if a plug-in is used. However, expert A believes that the trend with plug-ins is declining because many retailers want to have a customised website with their desired design to provide customers with a continuous "flow". If a retailer wants something special or unusual in his online shop, he would have to change a lot of things when using a plug-in, which may be more expensive than implementing it himself. These statements make it clear that the integration effort depends very much on the individual wishes and requirements of the online retailer. For this reason, the ease of integration or the low integration effort is not defined as a separate requirement criterion for the evaluation of the selected payment systems, but is summarised in the requirement for low total costs and taken into account in the evaluation.

An early payment date or a fast receipt of payment is considered to be a rather important requirement. According to expert B, the relevance depends primarily on the online retailers, whether they need the money in their bank account faster or not. An early and fast receipt of payment reduces the risk of non-payment on the one hand, and on the other hand it influences the total costs, since, for example, delayed receipt of payment results in storage costs for payment in advance or costs for collection measures for a purchase on account. As a consequence, the early payment date or the fast receipt of payment is not adopted as a separate request criterion for the evaluation of the selected payment alternatives, but is summarised in the requirements for high payment security and low total costs and taken into account in the evaluation.

The reduction of internal, manual processes represents a rather important requirement for the experts. Two experts agree that today's payment processes are automated and that the manual processes are manageable. After entering customer data and completing the order in the shopping cart, the payment process usually automated, including, for example, the address and creditworthiness checks of the customer when purchasing by invoice or instalment. In the meantime, a great deal can also be automated by means of image recognition procedures or similar for invoice purchases. Expert B is agrees that although the online retailer's internal processes can be implemented and implemented in a leaner manner, the time savings or degree of automation that can be achieved is insignificant from the retailer's point of view. Based on the statements of the experts, it is assumed within the scope of this study that the internal processes of online retailers are largely automated with regard to payment processing and that any process optimisations will not result in any significant changes. Accordingly, the reduction of internal, manual processes will not be included in the final catalogue as a requirement criterion.

Thus, of the ten requirements listed in the guide, only five are included as criteria for evaluating the selected payment systems, with two legal requirements being added.

These were not listed in the guide for the experts' assessment and evaluation, as they have to be fulfilled by the selected payment systems anyway.

In summary, the following seven requirement criteria are:

- usability/ease of use
- high distribution/acceptance by customers
- platform independence
- high payment reliability/low risk of non-payment
- cost-effectiveness/low total costs
- data privacy/protection of personal data and
- data security/protection against misuse.

## 6 Conclusion and Critical Discussion

In the course of this work, various requirements for payment systems were identified from the retailers' point of view. Although this work focuses on the retailer perspective, the requirements of customers were also considered, since they ultimately decide which payment method they want to use. In addition, current legal regulations have resulted in legal requirements that payment systems must meet. In particular, GDPR and PSD2 have led to far-reaching changes in payment processing in online retailing. The empirical study showed that the two most important requirements for payment systems from the retailer's point of view are the high level of dissemination and acceptance among customers and user-friendliness. According to the study, online retailers should focus on the distribution and acceptance by customers as well as on user-friendliness and ease of use when choosing payment methods. This enables them to further develop their sales and increase their conversion rate. High payment security and low overall costs are also important requirements, but were rated as less relevant by the experts surveyed. Platform independence and international applicability were also identified as requirements of retailers for sales development. However, these two requirements are less relevant and play a secondary role, especially with regard to international applicability.

The two most important requirements show that online retailers are, to a certain extent, externally controlled by the customers and that it is therefore advisable to orientate oneself according to the customers' wishes, even if this means that the retailer has to offer the more expensive payment methods. The general rule here is that the more payment methods an online retailer offers, the more sales it generates and the higher the probability that customers will find their preferred payment method. As far as the costs and effort involved are justifiable, online retailers should therefore offer as many payment methods as possible that are known and accepted by customers. Since the payment industry is in a constant state of flux due to technical developments, legal regulations and changing consumer behaviour, the topic of payment systems in e-commerce will continue to be an interesting and varied field of research in the future.

## References

1. Bundesverband E-Commerce und Versandhandel e.V. 2018. E-Commerce – der neue Nahversorger? [https://cloud.bevh.org/index.php/s/bVmooV05I64DkQD/download?path=%2F&files=180122%20bev\\_h\\_Praesentation%20E-Com-merce%20der%20neue%20Nahversorger%3F.pdf](https://cloud.bevh.org/index.php/s/bVmooV05I64DkQD/download?path=%2F&files=180122%20bev_h_Praesentation%20E-Com-merce%20der%20neue%20Nahversorger%3F.pdf). Accessed 18. Mar 2019.
2. Statistisches Bundesamt. 2018. Private Haushalte in der Informationsgesellschaft – Nutzung von Informations- und Kommunikationstechnologien. [https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400187004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400187004.pdf?__blob=publicationFile). Accessed 18. Mar 2019.
3. Bundesverband E-Commerce und Versandhandel e.V. 2018. Interaktiver Handel in Deutschland. Ergebnisse 2018. [https://www.bevh.org/fileadmin/content/05\\_presse/Auszuege\\_Studien\\_Interaktiver\\_Handel/Inhaltsverzeichnis\\_fu\\_r\\_bevh\\_Gesamtbericht\\_Interaktiver\\_Handel\\_in\\_Deutschland\\_2018.pdf](https://www.bevh.org/fileadmin/content/05_presse/Auszuege_Studien_Interaktiver_Handel/Inhaltsverzeichnis_fu_r_bevh_Gesamtbericht_Interaktiver_Handel_in_Deutschland_2018.pdf). Accessed 18. Mar 2019.
4. Bundesverband Digitale Wirtschaft. 2019. Bevölkerungsumfrage zum Thema Zahlungsmethoden. [https://www.bvdw.org/fileadmin/user\\_upload/190826\\_Studie\\_PSD2.pdf](https://www.bvdw.org/fileadmin/user_upload/190826_Studie_PSD2.pdf). Accessed 20. Sep 2019.
5. Böhle, K. and Riehm, U. 1998. Blümenträume – Über Zahlungssysteminnovationen und Internet-Handel in Deutschland. Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), Karlsruhe. <http://www.itas.kit.edu/pub/v/1998/bori98a.pdf>. Accessed 19. Mar 2019.
6. Hudetz, K. and Brückes, S. 2019. Zukünftige Payment-Lösungen im digitalen Zeitalter – Bestandsaufnahme und aktuelle Trends. In *Handel mit Mehrwert. Digitaler Wandel in Märkten, Geschäftsmodellen und Geschäftssystemen*, G. Heinemann, H. M. Gehrckens, T. Täuber and Accenture GmbH, Eds. Springer Gabler, Wiesbaden, 423–440.
7. Henkel, J. 2001. Anforderungen an Zahlungsverfahren. In *E-Commerce und E-Payment. Rahmenbedingungen, Infrastruktur, Perspektiven*, R. Teichmann, M. Nonnenmacher and J. Henkel, Eds. Gabler Verlag, Wiesbaden, 103–121.
8. Van Baal, S. and Hinrichs, J.-W. 2006. Internet-Zahlungssysteme aus Händlersicht. Bedeutung, Bewertung, Eigenschaften. In *Handbuch E-Money, E-Payment & M-Payment*, T. Lammer, Ed. Physica-Verlag, Heidelberg, 293–305.
9. Kollmann, T. 2019. *E-Business. Grundlagen elektronischer Geschäftsprozesse in der digitalen Wirtschaft*. Springer Gabler, Wiesbaden.
10. Heng, S. 2006. E-Payment-Systeme. Treiber einer notwendigen Evolution der Zahlungssysteme. In *Handbuch E-Money, E-Payment & M-Payment*, T. Lammer, Ed. Physica-Verlag, Heidelberg, 419–428.
11. Herzig, R. 2016. Digital Payments aus Sicht eines Handelsunternehmens. In *Digital Payments - Revolution im Zahlungsverkehr*, M. W. Mosen, J. Moormann and D. Schmidt, Eds. Frankfurt School Verlag, Frankfurt am Main, 179–196.
12. Himmelspach, A., Runge, A., Schubert, P., and Zimmermann, H.-D. 1996. Anforderungen an elektronische Zahlungssysteme. Universität St. Gallen, St. Gallen. [https://irf.fhnw.ch/bitstream/handle/11654/9353/\\_ppt\\_content\\_pub\\_anforderungen-an-elektronische-zahlungssysteme\\_Ab\\_bm51.pdf?sequence=1](https://irf.fhnw.ch/bitstream/handle/11654/9353/_ppt_content_pub_anforderungen-an-elektronische-zahlungssysteme_Ab_bm51.pdf?sequence=1). Accessed 06. Mar 2019.
13. Dannenberg, M. and Ulrich, A. 2004. *E-Payment und E-Billing. Elektronische Bezahlungssysteme für Mobilfunk und Internet*. Gabler Verlag, Wiesbaden.

14. Ul, B., F., R., Mehraj, A., Ahmad, A., and Assad, S. 2017. A Compendious Study of Online Payment Systems: Past Developments, Present Impact, and Future Considerations. 256–271.
15. Zapkau, F. and Schwickert, A. C. 2006. E-Payment-Systeme – Funktionsweisen, Marktüberblick, Bewertung. Arbeitspapiere Wirtschaftsinformatik, Professur BWL – Wirtschaftsinformatik, Justus-Liebig-Universität Gießen, 4.
16. Oh, S. 2006. A Stakeholder Perspective on Successful Electronic Payment Systems Diffusion. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06). IEEE, Los Alamitos, CA, USA, 186b-186b. DOI=10.1109/HICSS.2006.31.
17. Wichmann, T. 2002. Internet-Zahlungssysteme aus Sicht von Online-Händler. In Handbuch ePayment. Zahlungsverkehr im Internet: Systeme, Trends, Perspektiven, K.-H. Ketterer and K. Stroborn, Eds. Deutscher Wirtschaftsdienst, Köln, 119–133.
18. E-Commerce-Center Handel. 2013. Der Internet-Zahlungsverkehr aus Sicht der Verbraucher in D-A-CH. Ergebnisse der Umfrage IZV11. Eine Zusammenfassung der Studie des ECC über den Online-Payment-Markt in Deutschland, Österreich und der Schweiz, Köln. [https://www.ifhshop.de/media/pdf/6e/83/7f/Der-Internet-Zahlungsverkehr-aus-Sicht-der-Verbraucher-in-D-A-CH\\_2013\\_Summary.pdf](https://www.ifhshop.de/media/pdf/6e/83/7f/Der-Internet-Zahlungsverkehr-aus-Sicht-der-Verbraucher-in-D-A-CH_2013_Summary.pdf). Accessed 15. Aug 2019.
19. Stahl, E., Wittmann, G., Krabichler, T., and Breitschaft, M. 2015. E-Commerce-Leitfaden. Noch erfolgreicher im elektronischen Handel. <https://www.ecommerce-leitfaden.de/studien/item/e-commerce-leitfaden-3-auflage>. Accessed 04. Mar 2019.
20. Paul, M. and Stahl, E. 2017. Optimierung des Checkouts – Mission possible! So steigern Sie Ihre Konversionsraten im Bestellprozess, Regensburg.
21. Hinrichs, J.-W., Stroborn, K., and van Baal, S. 2004. (Mobiles) Bezahlen aus der Sicht des Online-Händlers. Status Quo und Perspektiven. In Mobile Economy -- Transaktionen, Prozesse, Anwendungen und Dienste, Proceedings zum 4. Workshop Mobile Commerce, Universität Augsburg 2.-3. Februar 2004, 63–78.
22. Penzel, H.-G. and Seidenschwarz, H. 2016. Kosten von Zahlungsverfahren im E-Commerce aus Händlersicht. In Digital Payments - Revolution im Zahlungsverkehr, M. W. Mosen, J. Moormann and D. Schmidt, Eds. Frankfurt School Verlag, Frankfurt am Main, 197–214.
23. Deichner, N., Seidenschwarz, H., and Stahl, E. 2019. Gesamtkosten von Zahlungsverfahren im deutschen E-Commerce 2019. Eine empirische Erhebung unter Online-Händlern, Regensburg.
24. Mayring, P. 2015. Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz, Weinheim.

# An Innovative IoT Based Financing Model for SMEs

Marcel Kehler<sup>1</sup>, Stefanie Regier<sup>1</sup> and Ingo Stengel<sup>1</sup>

<sup>1</sup> University of Applied Sciences Karlsruhe, Germany  
marcel.kehrer91@gmail.com

**Abstract.** Industry 4.0 and the developed technology of the Internet of Things are changing the global economy rapidly. Companies have completely new opportunities to sell their products and services. This creates new business models, such as pay-per-use, in which billing is based on the data collected. In this paper, a way is being worked out how banks can use this data for themselves in order to offer their business customers new financing products. For this purpose, the requirements of companies wishing to offer pay-per-use were addressed qualitatively. A financing model was then developed on the basis of this, which in turn was qualitatively assessed by banking experts. In a final online survey, the financing model was assessed quantitatively by the target group. In addition, the risks associated with the changes are discussed and proposed solutions for eliminating them are presented.

**Keywords:** finance 4.0, IoT, banking, financial innovation, SME.

## 1 Introduction

Industry 4.0 is currently considered the fourth industrial revolution worldwide. Following the invention of the steam engine, the discovery of electricity and automation through electronics, and the invention of the personal computer, "smart" objects and the associated technology of Internet of Things (IoT) are now leading to a further paradigm shift [16].

However, this revolution is associated with enormous investment costs [3]. Shorter innovation cycles and constant cost pressure are putting increasing pressure on small and medium-sized enterprises (SMEs). But it is not only in industry that major changes are currently taking place. In the banking sector, the low interest rate period, new regulations and digitalisation mean that those involved will have to rethink and change their processes in order to remain competitive [4]. Industry 4.0 offers a great opportunity for banks in this context [14]: with the end-to-end networking of devices and the resulting data, new business models can be developed that are optimally tailored to the personal needs of customers [14]. In the following, a financing model will be presented that uses IoT to flexibly adjust the liquidity burden of investment loans to the respective situation of the company in order to take account of economic and seasonal fluctuations.

## 2 Conceptual basis and relevant literature

Kevin Ashton first described the Internet of Things (IoT) in 1998 as the ability to connect people and objects at any time, in any place, with anything and everything, and ideally through any network or service [1]. IoT should represent all devices and persons in the virtual world and connect them with each other with the help of a virtual footprint [9]. The development of IoT is divided into five phases: the networking of two computers, the introduction of the Internet, the development of smartphones and the associated mobile Internet, social networks and finally the networking of objects from daily life, such as cars, cash machines, lamps or refrigerators [14],[18].

In the banking sector, too, this concept is becoming increasingly important in order to offer bank customers benefit. According to Drinkwater, this includes not only the use of wearables and Bluetooth beacons but also the use of intelligent cars as mobile ATMs, block chain technology and the development of chatbots [5]. Wang complements these with the possibility of real-time analysis of the devices and the use of capacity utilization and idle time for better pricing of leasing offers [7]. In their study, Ramalingam and Venkatesan are working on an ATM 2.0 that enables the use of IoT to connect to smartphones. In this way, they make it possible to withdraw money without a bankcard [20].

When using intelligent objects, large amounts of data are generated that have to be stored, processed and analysed. This enormous data is summarised under the term Big Data. The use of large amounts of data also offers great advantages in banking. The application scenarios described in the literature can be divided into the following categories: security, risk management, customer relationship management (CRM) and other uses. The studies [10], [11], [19] and [21] provide a very good overview of Big Data in the banking sector. In the security category, Wongchinsri and Kuratach's [23] meta-study provides an overview of the current technologies in this area, especially with regard to error detection and new customer selection. They analyse 41 articles dealing with this topic area. The articles mentioned there are very well suited to gain a deeper understanding of the topic. Kharote [15] uses Big Data to detect money laundering. To make this possible, he has developed a framework that analyses transaction data using a new algorithm. Through this, he tries to detect anomalies and thus identify possible money laundering. In the CRM category, the researchers try to make statements and predictions about the banks' customers. A wide variety of approaches is used, each with different objectives. In their 2014 study [8], Fuschi and Tvaronaviciene make recommendations on how to deal with Big Data so that service quality can be guaranteed. They recommend that quality guidelines and the way data is collected to be constantly redefined. These should be adapted to the constantly changing requirements. Others try to analyse their customers with large amounts of data. In their study, Srivastava and Gopalkrishnan [22] use the example of an Indian bank to show how big data can be used to analyse customers and understand their behaviour. They focus on the output patterns of their customers, the sales channels used, customer segmentation and profiling, product cross-selling, mood and feedback analysis, and security and error detection. The data provided by the bank is processed and analysed using descriptive statis-

tics. In this way, the authors clearly show how Big Data can be used to analyse customers and thus generate significant benefit for banks. Other researchers try to use Big Data to analyse their customers' supply chains in order to identify potential customers and gain advantages over the competition. In their study, Hung and colleagues [25] were able to clearly show that the response and completion rates of potential customers previously analysed by Big Data were significantly higher. Risk management attempts to make predictions or assessments of the risk associated with the loan. The data is often included in the scoring of the customer. This can be done in various ways. Yadav and Thakur [24] have used past customer usage data to assess the risk of each customer. Calis et al. [26] use clustering and classification to assess the risk of customers. In his 2018 study [17], Odinet identified several frameworks that use large amounts of data to identify unfair credit to customers. Hurley and Adebayo [12] take a different approach. In their investigations, they uncover Big Data's problems in credit scoring and the gaps in legislation.

A review of the relevant literature on IoT and banking shows, that there are no studies and concepts that use the resulting data volumes for a financing model. Thus, provide the customer with innovative financing products. Therefore, this subject will be addressed in this paper. The focus is set mainly on small and medium-sized companies that offer their customers the pay-per-use business model or plan to offer it in the future.

### **3 Research methodology and study design**

The following results were compiled using the mixed-method approach. This approach combines elements of qualitative and quantitative research and thus enables a deeper understanding of the interrelationships [13].

In qualitative research, verbal, visual and audio-visual data are collected, structured, and analysed in the course of the project. This approach attempts to understand the behaviour of the participants and then to transfer it inductively to the general public [2]. In this work, guideline-based interviews with decision-makers from small and medium-sized enterprises were conducted. These interviews aimed to gain new insights into the current situation and the challenges facing companies. In addition, requirements for banks and their products are defined. Based on this, a financing model is designed. This was analysed and improved in further interviews with banking experts. The focus here is primarily on the requirements for banks, to assess the feasibility and to identify problems that are likely to arise. The aim of the second qualitative study was to examine the feasibility from the banks' point of view. A total of eight interviews with bank representatives and SME decision-makers were conducted in March and April 2020.

The modified financing model was then presented in the form of a standardised survey to SME decision-makers, who were asked to evaluate the concept. The aim of the quantitative survey in April and May 2020 was to obtain an initial assessment of the acceptance of the developed financing model in the market and the willingness of companies to pass on customer data. With almost 30 decision-makers from SMEs having completed the survey, the sample is not representative due to the number alone, but it

does provide initial indications of the acceptance and possible weaknesses of the concept. The new financing model as well as challenges and evaluation have been developed and optimised directly from the interviews and the survey and will be presented in the following sections.

## 4 The IoT financing model

The target group for the developed financing model are SMEs that want to offer their customers the pay-per-use business model or already do so. With pay-per-use, the customer does not buy the product, but only pays for the use of the product. Since this model is designed to generate more money in the medium to long term, the SME must first make advance payments. They must make their products available without directly receiving the full sales price. This means an enormous capital requirement, especially when scaling up. After a successful launch, the recurring revenues will finance further development and production of new products.

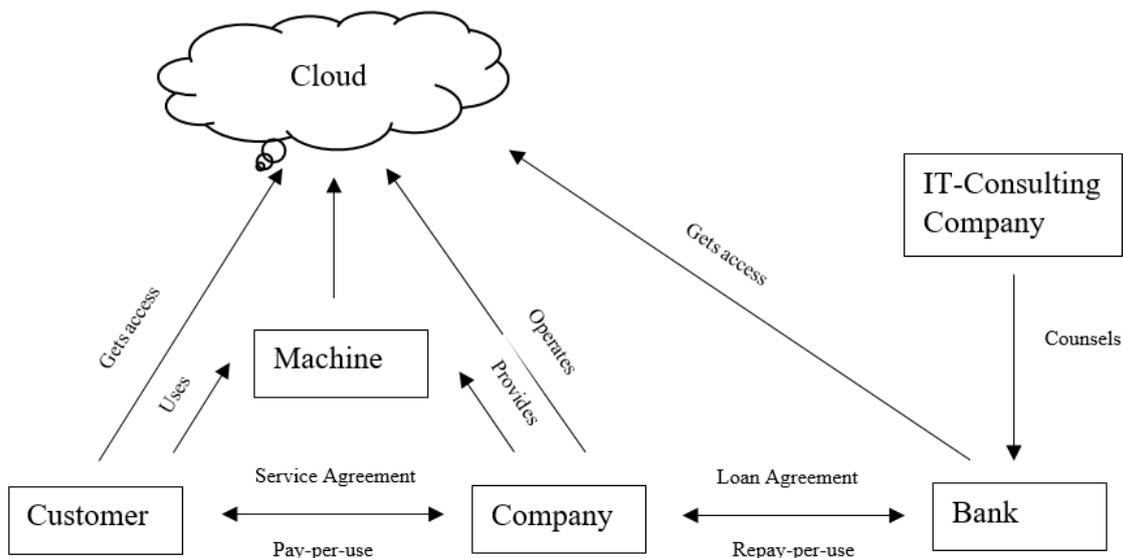


Fig. 1. Overview of the different roles

The financing model is based on the pay-per-use business model and therefore uses the same calculation basis as used previously mentioned. With the financing model, the repayment amount is calculated flexibly on the basis of the end customer's user data. In order to ensure that planning security is still required, a maximum financing period is agreed during which the loan must be paid off, if the financed product is not used. However, if the product is used, an additional, previously agreed amount is repaid for each unit. Thus, the more intensively the financed product is used by the end customer, the higher the manufacturer's repayment to his bank and the shorter the term.

The roles involved in the financing model are shown below and the respective relationships are clarified. The interaction of these roles is shown in Figure 1.

The SME plans to offer its products in the future with the pay-per-use business model. This products could be for example machines (Machine-as-a-Service), lamps (Light-as-a-Service) or cars (Mobility-as-a-Service). To this end, it will make its products available for use by its end customers and bill them on the basis of usage. In order to be able to offer the model across companies, they need a cloud in which they will consolidate, evaluate and automatically calculate their customers' usage data. This will be operated as part of his billing. By selling via service contracts, it will generate recurring income, but this only covers the costs incurred on a medium to long term. When scaling, this results in a significantly higher capital requirement for the products, which has to be covered with the help of a loan.

The end customer wants to avoid the risk of a capital-intensive purchase and still use the manufacturer's products. In addition, they want to have greater planning security and avoid unforeseen expenses for repairs or new investments, for example. For these reasons, they are looking for new business models with more planning security and less risk of default. He uses the new business model from the product manufacturer and uses his product without being the owner. Billing is based on the intensity of use.

The intelligent product is provided by the manufacturer and used by the end customer. It records a wide variety of data via sensors built into it, which is then transferred to the cloud operated by the manufacturer. From there, all participants can access data and analyse, evaluate and use it for billing.

The bank provides the manufacturer with the required capital and at the same time, it offers him an innovative, flexible financing model that will adjust the repayment rate to the end customer's usage. To do this, the bank is given access to the data stored in the cloud.

## 5 Challenges

When introducing pay-per-use or using the financing model, there are four challenges, which are explained in more detail below:

- Early repayment of the loan,
- Minimum redemption rate in the absence of turnover,
- Cluster risks,
- Change in the manufacturer's balance sheet.

Thanks to the design of the financing model, the early repayment of loans is no longer an exception, but has been deliberately brought about. However, this entails an increased risk for banks. They have to borrow money from other credit institutions, such as the European Central Bank (ECB), in order to grant loans to their customers. For private customers, this will generate costs. A good indicator of the costs incurred is the swap rate, which describes what fixed interest rate banks are prepared to pay for certain terms [6]. Due to a high utilization of the machine, the loans will most likely be repaid before the current financing term. As a result, the costs will continue to exist,

but on the other hand no income will be generated. If this is not taken into account in the calculation, it is possible that banks will suffer losses. In order to counteract this, banks must consciously plan and calculate risks. Early repayment interest rates, which are based on the swap rate, are suitable for this purpose. In this way, banks will protect themselves against a loss and the borrowers still will save parts of their costs (from a holistic perspective).

A further challenge is the dependence on customer benefit. In extreme cases, this can lead to the end customer not using the machine and thus not generating any income for the manufacturer. However, the manufacturer has to service the minimum redemption amount toward the bank. This situation can be life threatening for the manufacturer. In order to counteract this, the manufacturer should always agree a minimum purchase quantity or a monthly rental price with his end customer with which he can certainly cover the redemption payments. Regardless of this, an end customer can of course also become insolvent. However, this risk also exists with other types of credit. Moreover, the manufacturer has spread his risk over several customers and can make the product available to another customer as quickly as possible.

Another risk that was identified during the expert interviews is the cluster risk at banks. In the case of capital-intensive products, before the introduction of pay-per-use, the end customers were obliged to finance the products. As a result, the necessary amount of loan was spread over many companies and the bank was diversifying the existing risks. Furthermore, it is very unlikely that all end customers were at the same bank. As a result, the amount of credit was not only distributed among different customers, but also among different banks. Pay-per-use increases the risks for banks, since the manufacturers alone now need full financing. Although, it can be assumed that the manufacturing costs are significantly lower than the current sales price, the financing requirements of the manufacturer are significantly higher than those of individual end customers. In addition, there is no need for the manufacturer to distribute the loan among several banks, since the manufacturer normally wants to take out a loan with his house bank. In conjunction with the regulatory regime to which all banks are subject, this can mean that the loan cannot be guaranteed by just one bank. In addition, this leads to an increase in the risk and as a consequence also to an increase of the costs of the credit. In this case, the bank's action possibilities are limited due to the regulation controlled by the banking supervisory authorities. This means that to jointly cover the forthcoming large capital requirements, banks can only look for partner banks in advance.

The final challenge for the financing model is the change in the manufacturer's balance sheet and the resulting lower rating. Assessing the manufacturer's creditworthiness, calculating the risk and the associated premium is always a particular challenge with a new business model. According to the Basel 1, 2 and 3 resolutions, banks are required by the Banking Supervisory Authorities to carry out a rating for each borrower, in which they assess the borrower's creditworthiness. The worse the rating, the greater the risk and the higher the interest on the loan. In the calculation, a worst-case consideration must always be carried out and thus both qualitative and quantitative indicators are used. For example, the industry and the competitive position of the company are used as qualitative indicators. The annual financial statements and the balance sheet contained therein are used as quantitative characteristics for the rating. From these, key

figures are calculated that are used for the rating. A very important key figure in the rating is the equity ratio, which expresses the proportion of equity to total capital. The lower the equity ratio of a company, the worse is its rating. The introduction of the pay-per-use billing model has far-reaching effects on the provider's balance sheet and thus also on the key figures calculated for the rating. With pay-per-use billing, the products are no longer sold, but made available and leased using a service contract. As a result, the products manufactured are listed on the assets side of the balance sheet. These are typically, in the best case, financed by debt using the financing model developed here. This external financing is noticeable on the liabilities side of the balance sheet. As a result, the balance sheet total rises significantly and the equity ratio, which is important for the rating, deteriorates significantly. In turn, this will lead to higher risk interest rates, which will make the model unprofitable for manufacturers. This problem can be solved in two ways. The first option is to develop a new rating system that takes into account the exact conditions and might include data generated by the product. The other possibility is to spin off the company into a subsidiary. In this way, the manufacturer can continue to sell its products in the parent company as before and implement the pay-per-use business via a subsidiary.

This can then be rated by the bank in the role of a leasing company. Experience shows that these have a similar balance sheet structure.

## 6 Evaluation and Conclusion

The evaluation of the financing model by SME decision-makers provides initial indications of the acceptance, possible problems and opportunities of the model on the market.

Already 50 % of the companies surveyed are already recording data, and are using it for a variety of reasons. 10 % of the SME's already use the data for billing their services. Another 10 % use the data within the scope of service and maintenance contracts. Surprisingly, 25% of those surveyed record data but do not yet use it or only use it in individual cases. The most important finding is the possible potential of the financing model. 3 out of 4 companies consider recording the usage behaviour of their customers to be useful and are therefore also considered as a target group for the financing model.

Overall, it can be said that the IoT financing model has met with great interest among the test persons. On a scale of 5, ranging from very unattractive to very attractive, the test persons had to assess the attractiveness of the IoT financing model presented. 84 percent of the respondents rated the model as very attractive or attractive. Only 16 percent opted for the neutral middle. The survey participants saw the advantages of the idea of calculating the repayment participation on the basis of IoT data primarily in the flexible liquidity burden.

However, the survey also shows that many companies are still uncertain whether they are allowed to share the data with third parties from a data protection perspective and whether they need the consent of their customers to do so. For this reason, it is extremely important to sensitise customers to the topic of data protection from the start

and to explain to them in a comprehensible manner when the DSGVO takes effect, which regulations must be complied with and how this is implemented in the present financing model.

In addition, the lack of skilled workers is a major problem. The technical professions, which are also indispensable for the introduction of pay-per-use, are particularly affected. Companies have to think of suitable concepts for attracting talented and well-trained people and retaining them in the long term. Banks have to think about how they want to counter the conflict of objectives between the shortage of skilled workers and cost reduction in the long term. This is the only way to implement and change towards usage-based billing models.

In addition to the problems already mentioned, both the expert interviews and the online survey revealed that banking regulations are an enormous challenge. Here, it is of utmost importance for the success of the financing model that a new regulations are found that enables banks to provide a suitable rating for companies that act on the market as both product and service providers. Only in this way companies can take this innovative path without being put at an enormous disadvantage in future lending.

In summary, it can be said that the model presented here is suitable for companies who, in addition to selling their products, also want to position themselves as service providers in the market. This business model is less suitable for companies that want to act entirely as a service provider in the future, as banking regulations will push up risk costs too much. This is mainly due to the fact that there is no suitable rating model for the pay-per-use business model. This is precisely where further research will start.

## References

1. Ashton, K. (2009): That 'Internet of Things' Thing. In: *RFID Journal*. <https://www.rfidjournal.com/articles/pdf?4986>, last accessed 19.12.2019.
2. Berekoven, L.; Eckert, W.; Ellenrieder, P., A. (2009): *Marktforschung Methodische Grundlagen und praktische Anwendung*, Wiesbaden: Gabler.
3. Bitkom Research (2019): *Industrie 4.0: Anwendungshemmnisse in deutschen Industrieunternehmen 2019* | Statista. <https://de.statista.com/statistik/daten/studie/990517/umfrage/umfrage-zu-hemmnissen-bei-industrie-40-anwendungen-in-industrieunternehmen/>, last accessed 03.12.2019.
4. Böhnke, W.; Rolfes, B. (Hg.) (2018): *Neue Erlösquellen oder Konsolidierung? – Geschäftsmodelle der Banken und Sparkassen auf dem Prüfstand*. Beiträge des Duisburger Banken-Symposiums. Wiesbaden: Springer Gabler.
5. Drinkwater, D. (2017): *Eight examples of how IoT is improving retail banking* | Internet of Business. <https://internetofbusiness.com/eight-examples-iot-retail-banking/>, last accessed 31.01.2020.
6. Durchblicher(2020):*Swapsatz*. <https://durchblicher.at/immokredit/ratgeber/swapsatz>, last accessed 29.05.2020.
7. Wang, E. (2017): *IOT in Banking - Enabling Bank's Digital Future*. Hg. v. Edgeverve. <https://www.edgeverve.com/finacle/finacleconnect/trulydigital-2017/perspective/iot-banking-enabling-banks-digital-future/>, last accessed 31.01.2020.
8. Fuschi, D. L.; Tvaronavičienė, M. (2014): *Subsustainable Development, Big Data and supervisory control: Service quality in Banking Sector*. In: *JSSI* 3 (3), pp. 5–14.

9. Govinda, K.; Saravanaguru, R. A.K. (2016): Review on IOT technologies. In: *International Journal of Applied Engineering Research* 11 (4), pp. 2848–2853.
10. Hassani, H.; Huang, X.; Silva, E. (2018a): Banking with block chained big data. In: *Journal of Management Analytics* 5 (4), pp. 256–275.
11. Hassani, H.; Huang, X.; Silva, E. (2018b): Digitalisation and Big Data Mining in Banking. In: *BDCC* 2 (3), p. 18.
12. Hurley, M.; Adebayo, J. (2017): Credit scoring in the era of big data. In: *Yale Journal of Law and Technology* 18 (1), p. 5.
13. Johnson, R. B.; Onwuegbuzie, A. J.; Turner, L. A. (2007): Toward a Definition of Mixed Methods Research. In: *Journal of Mixed Methods Research* 1 (2), pp. 112–133.
14. Khanboubi, F.; Boulmakoul, A.; Tabaa, M. (2019): Impact of digital trends using IoT on banking processes. In: *Procedia Computer Science* 151, pp. 77–84.
15. Kharote, M. (2014): Data Mining Model for Money Laundering Detection in Financial Domain. In: *International Journal of Computer Applications* (Volume 85), pp. 61–64. <https://research.ijcaonline.org/volume85/number16/pxc3893337.pdf>, [accessed 22.06.2020].
16. Lasi, H.; Fettke, P.; Kemper, H.-G.; Feld, T.; Hoffmann, M. (2014): Industrie 4.0. In: *Wirtschaftsinf* 56 (4), pp. 261–264.
17. Odinet, C. K. (2018): Consumer Bitcredit and Fintech Lending. In: *International Political Economy: Investment & Finance eJournal*.
18. Perera, C.; Liu, H.; Jayawardena, S.; Chen, M. (2014): A Survey on Internet of Things From Industrial Market Perspective. In: *IEEE Access* 2, pp. 1660–1679.
19. Prabhu, C.S.R.; Chivukula, A. S.; Mogadala, A.; Ghosh, R.; Livingston, L. J. M. (2019): Big Data Analytics for Financial Services and Banking. In: C. S. R. Parbhu, Aneesh Sreevalabh Chivukula und Aditya Mogadala (Hg.): *Big data analytics. Systems, algorithms, applications*. Singapore: Springer Singapore, pp. 249–256.
20. Ramalingam, H.; Venkatesan, V.P. (2019): Conceptual analysis of Internet of Things use cases in Banking domain. In: *Proceedings of The TENCON 2019. Technology, Knowledge, and Society : 17-20 October 2019, Grand Hyatt Kochi Bolgatti, Kerala, India. TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. Kochi, India, 10/17/2019 - 10/20/2019. Piscataway, NJ: IEEE, pp. 2034–2039.
21. Sharma, S.; Sharma, T.; Kotak, B.; Hasotkar, A. (2019): Big Data Analysis in Banking Sector. In: *International Journal of New Technology and Research* 5 (10).
22. Srivastava, U.; Gopalkrishnan, S. (2015): Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. In: *Procedia Computer Science* 50, pp. 643–652.
23. Wongchinsri, P.; Kuratach, W. (2016): A survey - data mining frameworks in credit card processing. In: *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1–6.
24. Yadav, S.; Thakur, S. (2017): Bank loan analysis using customer usage data: A big data approach using Hadoop. In: Balvinder Shukla (Hg.): *2nd International Conference on Telecommunication and Networks - TEL-NET 2017*. 10th-11th August 2017, Amity University Uttar Pradesh, Noida, India. 2017 2nd International Conference on Telecommunication and Networks (TEL-NET). Noida, India, 8/10/2017 - 8/11/2017. *International Conference on Telecommunication and Networks; Amity University; Institute of Electrical and Electronics Engineers; TEL-NET*. Piscataway, NJ: IEEE, pp. 1–8.
25. Hung, Jui-Long; He, Wu; Shen, Jiancheng (2020): Big data analytics for supply chain relationship in banking. In: *Industrial Marketing Management* 86, pp. 144–153.

26. Calis, Asli; Boyaci, Ahmet; Baynal, Kasim (2015): Data mining application in banking sector with clustering and classification methods. In: International Conference on Industrial Engineering and Operations Management (IEOM), 2015. 3 - 5 March 2015, Hyatt Regency, Dubai, United Arab Emirates. 2015 International Conference on Industrial Engineering and Operations Management (IEOM). Dubai, 3/3/2015 - 3/5/2015. Industrial Engineering & Operations Management Society; Institute of Electrical and Electronics Engineers; International Conference on Industrial Engineering and Operations Management; IEOM. Piscataway, NJ: IEEE, S. 1–8.

# Operational Management of Data Centers Energy Efficiency by dynamic optimization -Based on a Vector Autoregressive Model- Reinforcement Learning(VAR-RL) approach

Yang Hai<sup>1</sup>[0000-0003-4656-1594], Truong Van Nguyen<sup>2</sup>[0000-0001-9380-5710], Youngseok Choi<sup>3</sup>,  
Mohammed Bahja<sup>4</sup>[0000-0002-2138-1784], and Habin Lee<sup>5</sup>[0000-0003-0071-4874]

<sup>1,2,3,5</sup> Brunel University London, Middlesex UB8 3PH, UK

<sup>4</sup> University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Yang.Hai@Brunel.ac.uk

TruongVan.Nguyen@Brunel.ac.uk

Youngseok.Choi@Brunel.ac.uk

M.Bahja@Bham.ac.uk

Habin.Lee@Brunel.ac.uk

**Abstract.** With the increasing demands of digital computing, Data Centers (DCs) have become a leading scheme for global energy issues. Major efforts that can be observed for DC energy efficacy solutions are focusing on relatively problematic infrastructure designs. Nevertheless, we emphasised the managerial strategies of using the existing facilities to achieve energy efficiency through active intervention. It is believed that there exists a trade-off between the cooling devices and IT devices. Accordingly, the Vector Autoregressive Model- Reinforcement Learning(VAR-RL) approach will be proposed as a combination of traditional multivariate time series modeling technique and the artificial intelligence technique which allows us to predict and adjust the prediction of an error would help to explore the complex dynamic interrelationships between the two types of devices. Moreover, an optimization decision support system will also be conducted subsequently to optimize Power Usage Effectiveness (PUE) by controlling the combination of Air Conditioners (ACs). The proposed VAR-RL approach would not only increase the forecasting accuracy but also would adapt to the environment changes dynamically, this would give a better foundation for the DC energy efficiency optimization. The data we adopted is the real-time data from a DC located in Turkey. Consequently, the novel of this study would save the DC energy consumption tremendously.

**Keywords:** DC, Energy consumption, optimisation.

## 1 Introduction

DCs are energy-intensive industries and they are taking 1-1.5% of global electricity usage every year [1]. There are two main units of DCs that are supplied the energy, IT,

and cooling units. Widely cited studies and conservatively evaluation demonstrated the fact that there are around 40% of energy has been taken by the cooling system in a typical air-cooling DC [2]. IT equipment, along with its respective power supply, creates heat when being in use and will raise the surrounding ambient temperature. In consequence, the equipment is likely to fail if the temperature becomes too high. Previous to 2004, IT companies were too fixated upon the performance of their equipment and adjusted the equipment's environment with only reliability in mind, with little weighting to energy costs. Nowadays the temperature range has been wider to 18 – 27°C (ASHRAE 2016)[3]. However, there are over 90% of DCs still keeping a constant temperature which means that the DCs are over-cooled and energy inefficiency [4].

Since most of the energy consumption lies in cooling, efforts have been made on reducing the cooling energy consumption in the DCs. It has been found that configuration design predominantly affects on energy consumption. As long as there is a design revolution, the application still requires a long time. As a result, traditional air-cooling will still domain the DC cooling system in the next few decades [5]. Therefore, we are seeking DC cooling efficiency solutions from a different aspect. As we mentioned earlier, cooling devices are the biggest energy consumers in the DC however there is no guideline on how to make the optimal usage of them. The common operations in DCs are still following traditional rules by turning off some certain number of ACs to save energy in winter. But to our knowledge, there is no sophisticated analysis so far to guide the optimal use of ACs combination, which gives us ideas to fill up this gap.

According to the relationship of energy consumption units both IT and Air conditioning devices, a lower temperature will increase the energy supplied, inversely, will reduce the energy consumption of IT devices due to an increasing computing efficiency. Therefore, smart operation management on DCs to find out the optimal solution on temperature control to minimize the energy consumption without affecting the performance of IT devices and meeting the service-level agreement has been investigated. Therefore, this has become our motivation for this study. The question can be modeled as a Linear Programming (LP) problem. However, applying the LP method to the problem requires understanding the complex interactions among many variables within the DC. To solve and simplify this issue, we adopt VAR model to identify such complex interactions. It has been taken to account for common features of the industry big data

Changes rapidly in the structures. Changes in server workload, outside environment, device locations or human intervention all can be reasons that lead to structural breaks of the series. Numerous empirical studies put attention on post-event detection, which wildly used for economic or business analysis, while rarely of them are looking at this issue in a real-time. Instead, we expect the model would able to autonomous evaluate itself and correct the mistake once it notices it. In this study, we adopt RL approach for dynamic real-time adjustment of VAR model. It would take the responsibility to detect the structural break and trigger the parameter re-estimated system. With the proposed VAR-RL approach, the subsequent optimization problem can be solved with the consideration of the changes in the environment in real-time.

Our contributions to the literature including the following: (1) Give the DC energy-efficient solution without changing DC configurations. (2) A dynamic simulation sys-

tem based on VAR-RL approach has been made, this will provide an efficient and accurate forecast for the complex environment with the adjustment of structural changes in the data. (3) Real-time optimization will be conducted based on the simulation result and future optimization also can be made by the forecasted data set. (4) This study will also arouse the environmental awareness of energy saving.

## 2 Literatures review

We have searched the empirical studies on DC effective cooling strategy of simulation-based optimization. Among the ten results, eight of them adopted the Computational Fluid Dynamic (CFD) while and the rest of them used Data-driven Models (DDM) model and other configuration design simulation tools for airflow simulation.

Most of them are from a pure configuration design and layout aspect, ie. [6] on sensors placement strategy and [8] on air aisle and racks layouts [7]. There are only [8, 9, 10] among the results that looking at the optimal temperature solutions, however [9] it doesn't take the trade-off relationships between cooling and IT into considerations and [8] is an equation-based simulation that looking at system network control. Also [10] studied the combination of water and airflow in Indirect Adiabatic Cooling (IAC) DC. Numerous studies that use the First principle (FP) in terms of DC objectives largely rely on pre-defined algorithms. However, in practical, there are a variety of unknown relationships that cannot be acquired from physical principles. Data-Driven Models (DDMs) avoid this problem by adopting experimental data to train a system. There is study compares temperature prediction performance of four different types of DDMs including Artificial Neural Networks (ANN), Support Vector Regression (SVR), Gaussian Process Regression (GPR) as well as Proper Orthogonal Decomposition (POD) in a DC, the training data is given by CFD simulation and the result demonstrated that most of them can give a relatively accurate prediction however only ANN could handle multiple output points in one model. Because of the unknown features of the system and multi-dimensional problem need to be solved in one model, so that it requires a large volume of data to feed in the model and moreover, all these types of models are facing similar difficulties which are computational expensive practical cases and relatively time-consuming [11].

We conservatively conclude that our VAR-RL approach would be the first study that further extended Linear Regression (LR) based RL to analyze complex industrial environment, then apply the simulation result to real-time optimization in industrial practical case. Due to the limited resources, we reviewed similar studies that used the similar method for different problems. RL approach has been used for an auto-select different combination of data streams to feed to the parameters-fixed LRs and practical application on typhoon rainfall prediction shows a better performance than traditional LRs [12]. More comprehensively, a Multi-Agent reinforcement learning (P-MARL) on predicting the future environment which allows the agents to adapt to the changes off-line by the combination of ANN and Autoregressive Integrated Moving Average (ARIMA) models, this joint approach also increased the prediction accuracy of the agents [13]. Moreover, efforts have been made on adapting RL to the side of the sensor to reduce

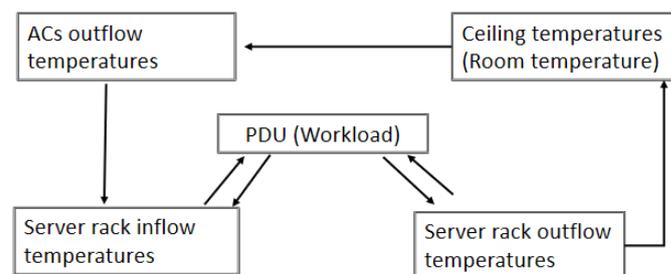
the energy transmission cost in the wireless sensor networks for signal prediction [14]. These approaches provided evidences that with RL adjustment, the prediction accuracy would be largely increased and gives the model more flexibility by self-learning during prediction without any training data which would require large storage and costs for computing.

Our VAR-RL approach would take advantage of these empirical studies, moreover, we will adapt this into DC industrial practice: (1) Our model parameters will be dynamically changed according to the feedback from the learning process that would allow our model to adapt to the environment changes. (2) We will propose a time-series linear regression model that will not only consider the previous one step (MACOV) but the whole period which has an effect on the present. (3) RL tool will be plugin to determine whether the model should be reused or rebuilt. (4) We will also avoid using the technique which has a black-box property such as Artificial Neuron Network (ANN) because it hides the interrelationships into the black-box procedure which limited the interpretative of the model. (4) Also, as we expect to perform a light, fast, and efficient model to adapt to the rapid industrial practice. we will avoid the use of techniques that requires huge size of training data and local storage. (5) We will not change any DC configurations including the sensors, only managerial strategies will be applied to the DC energy-saving practice.

### 3 Methodology.

#### 3.1 VAR model- A fundamental simulator

Based on the field study in TUKSAT DC target IT room, we identified the main factors that participated in the IT room computing environment. We are going to include ceiling sensors temperatures, Server rack inlet and outlet temperatures, air conditioner outflow temperatures as well as PDU values of the servers as our endogenous variables. To our knowledge, IT room objectives are mutually affected by each. As the graph shows below, we can infer that each variable can affect the others in two directions (direct or indirect), shown as a circulation (Fig.1). Statistical test (Granger causality test) results also confirmed the underlined inferences. Therefore, we briefly include all the related variables into one VAR model.



**Fig. 1.** The DC variables and relationships

Here we present how VAR modelling the above dynamic. VAR assumes all the variables to be endogenous and explain those endogenous variables one by one by all their

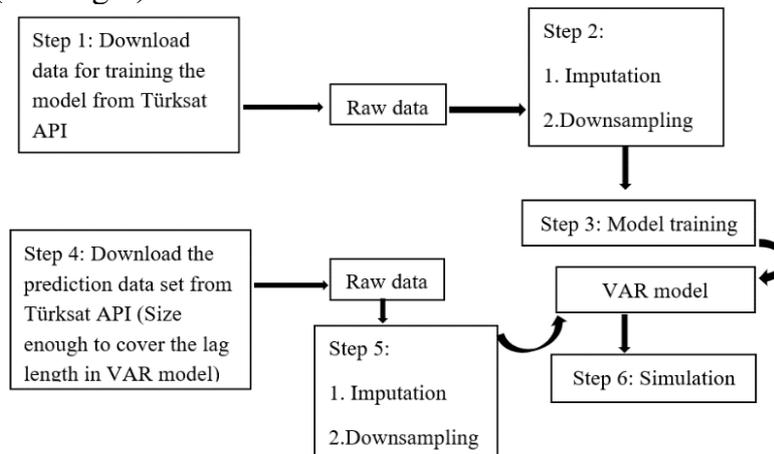
past values. This allows us to use the estimated model to predict the future values of variables. A  $p^{th}$  order  $VAR(p)$  can be represented as:

$$X_t = \sum_{i=1}^k \Pi_i X_{t-i} + C + u_t \quad (1)$$

Assume we have  $N$  variables, then  $X_t$  is the  $N \times 1$  order time series vector,  $C$  is the  $N \times 1$  order constant vector, the  $\Pi_i$  is the  $N \times N$  order parameter matrix,  $u_t$  is the  $N \times 1$  order random error vector. We can extend the above equation to the matrix formula as following (The lag length will be selected by the combination of "AIC", "HQ", "SC", "FPE" criteria):

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \dots \\ x_{N,t} \end{pmatrix} = \begin{pmatrix} a_{1,1} & \dots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \dots & a_{N,N} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ \dots \\ x_{N,t-1} \end{pmatrix} + \begin{pmatrix} b_{1,1} & \dots & b_{1,N} \\ \vdots & \ddots & \vdots \\ b_{N,1} & \dots & b_{N,N} \end{pmatrix} \begin{pmatrix} x_{1,t-2} \\ x_{2,t-2} \\ \dots \\ x_{N,t-2} \end{pmatrix} + \dots + \begin{pmatrix} n_{1,1} & \dots & n_{1,N} \\ \vdots & \ddots & \vdots \\ n_{N,1} & \dots & n_{N,N} \end{pmatrix} \begin{pmatrix} x_{1,t-p} \\ x_{2,t-p} \\ \dots \\ x_{N,t-p} \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \\ \dots \\ q_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_N \end{pmatrix} \quad (2)$$

As VAR model is a dynamic forecasting model. We can use it to simulate the DC environment as well as forecast the future values of each variable. After we get the model parameters by real-time estimation, we will feed the data that cover the lag length and forecast the future value of each variable. To make it clear, here we summarize the procedure to train a VAR model and use it as a simulator to forecast DC environment in a flowchart (See Fig.2).

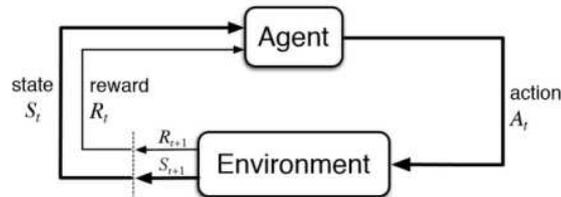


**Fig. 2.** The structure of the simulator

We first downloaded data from the historical Application Programming Interface (API) (Step 1). After data processing (Step 2), we use this data set to train a VAR model (Step 3). Then download historical data again (Step 4), after processing the data for another time (Step 5), we extract the most recent data which cover our lags, then input to the VAR model. The simulator will carry out iterations until reach to the requested forecasting length.

### 3.2 RL -A dynamic environment adaptor

As we mentioned earlier, the data estimated in DC has a rapidly changing feature, therefore our fundamental simulator VAR may not apply for some special cases: ie., suddenly changing load by holidays or online exams, temperatures changes, or other human interventions. Therefore, we need to adjust our model to ensure the prediction accuracy and be able to detect the environment changes and adapt itself to the changing world. Reinforcement learning as an environment adaptor to VAR model will be introduced in this section. The process is shown in the following graph (Fig.3).



**Fig. 3.** The environment adaptor based on RL

With every prediction, we will have an evaluation of the accuracy. And we will give a reward (or punishment) to each prediction. The accumulated reward would be:

$$R^{cum} = \sum_{g=1}^n R_{t+g} \quad (3)$$

Where  $R^{cum}$  is the total cumulated reward values,  $R$  is the reward for each forecast evaluation. With the number of time steps increasing, the difficulty level to predict would be increasing too, to make it fair enough for the judgement, we assign the weight to each reward, and the weight of the reward would be decreasing over the time.

$$R^{cum} = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \gamma \in [0,1) \quad (4)$$

Where,  $\gamma$  is the weight.

Time windows will be plugin at every seasonality changing point. And the system will trigger RL to evaluate the prediction result. In this case, the prediction result from the fundamental simulator VAR will be evaluated by the error rate. A reward will be given to each evaluation. When the accumulated reward value reaches to a certain boundary, the environment changes will be detected. Then the RL adaptor will trigger the alarm then the VAR model parameters and features will be rebuilt.

### 3.3 LP approach- An energy efficiency optimizer

Empirical evidences show that increasing the AC setpoint by a single degree can result in 4-5% energy cost savings; and increasing the setpoint by 10 degrees, which is also a realistic number, can result in savings of over 40% [15]. Although this sounds straightforward and simple, considering the complex nature of DC assets, it is hardly the case. Increasing the AC setpoints blindly can jeopardize the health of servers and other hardware, as existing hot spots may become even hotter and higher hot aisle temperature

may activate server fans and offset efficiency gains. Therefore, a rigorous plan for optimizing the AC temperature setpoint is critical to increasing the energy efficiency of the DC. Particularly, we aim to optimize the energy efficiency in the DC by determining the optimal combination of the supplied temperature of AC units, while taking into consideration the dynamic nature of IT power consumption, as well as satisfying the temperature constraints.

We will use the following notations:

$S = \{S_1, S_2, \dots, S_m\}$  denotes the Server rack number 1 to  $m$

$C = \{C_1, C_2, \dots, C_l\}$  is a set of ACs unit number 1 to  $l$ .

$T_x^{sup}(t)$ : the temperature supply of the  $x^{th}$  AC unit at time  $t$ .

$T_j^{in}(t)$ : the inlet temperature of the  $j^{th}$  server rack at time  $t$ .

$T_j^{out}(t)$ : the outlet temperature of the  $j^{th}$  server rack at time  $t$ .

$T^{room}(t)$ : the room temperature at time  $t$ .

$P_j^{Comp}(t)$ : the computational power (PDU) for the  $j^{th}$  server rack at time  $t$ .

$P_x^{Cool}(t)$ : the computational power for the  $x^{th}$  AC unit at time  $t$ .

$E[W_j(t)]$ : the estimated workload (CPU usage) for the  $j^{th}$  server rack at time  $t$ .

$CoP_x$ : the coefficient of performance.

$CTI_{jx}$  is the thermal correlation index.

#### The IT consumption

. Assumably, the IT power consumption would not only be influenced by the computational workload, but also will be affected by the working temperature because the temperature will affect its working performance. Hence, at any time  $t$ , the total IT computing power of server rack  $S_j$  is the function of power spent on executing IT jobs and the rack inlet temperature.

$$P_j^{Comp}(t) = \alpha_j E[W_j(t)] + \beta_j T_j^{in}(t) \quad (5)$$

Where  $\alpha_j$  and  $\beta_j$  are weight coefficients.

#### The cooling consumption

. Based on [16,17,18], the cooling cost of AC device  $C_x \in C$  can be presented as:

$$P_x^{Cool}(t) = \frac{CTI_{xj} \sum_{j=1}^m P_j^{Comp}(t)}{CoP_x(T_x^{sup}(t))} \quad (6)$$

Where  $CoP_x$  is the performance coefficient, shown as the ratio of the amount of heat the AC device  $C_x$  needs to remove to the energy it needs to consume to perform the removal.  $CoP_x$  indicates the efficiency of the AC device, and is typically a non-linear, increasing function of the supplied cold air temperature,  $T_x^{sup}(t)$ . It means that operating the AC system at a higher temperature is saving energy, as providing colder air requires the AC to work harder and consume more energy to remove heat. Hence, we can minimise  $P_x^{Cool}(t)$  by maximise the allowable supplied cold air temperature,  $T_x^{sup}(t)$  that satisfies the constraint of redline thresholds. The simulation approach will

also be used to get the function of  $CoP_x \left( T_x^{sup}(t) \right)$ .  $CTI_{jx}$  is the thermal correlation index,  $CTI_{jx} = \frac{\Delta T_j^{in}}{\Delta T_x^{sup}}$ , which represents the influence of each AC unit  $C_x$  on inlet temperature of server rack  $S_j$ . As defined in Eq. (6), it quantifies the response of the server  $S_j$ 's inlet temperature  $T_j^{in}$  to a step-change in the supply temperature  $T_x^{sup}$  of  $C_x$ .  $CTI_{jx}$  is a static metric, which is stable with time but based on the physical configuration of the DC. Hence, we use the simulation approach to get the value of  $CTI_{jx}$ . The detailed explanation of this metric can be seen in [17,18].

### Thermal modelling

. According to the law of energy conservation, almost all the computing power consumed by a server is transformed into heat, hence the relationship between the power consumption and inlet/outlet temperature of server rack  $S_j$  can be presented as:

$$T_j^{out}(t) = T_j^{in}(t) + K_j P_j^{Comp}(t) \quad (7)$$

Where  $K_j = pf_j c$  is the thermal-physical term. This can be estimated by our data obtained in DC.

Typically, the server's inlet temperature ( $T_j^{in}$ ) tends to be higher than the AC's supplied air temperature ( $T_x^{sup}$ ) due to the phenomenon so-called heat recirculation where the hot air from the server and the supplied cool air from the AC are mixed then recirculates in the room. Based on the energy conservation as described in Eq. (8) and the assumption of the fixed airflow pattern in the computer room, prior studies (eg. [19]) characterise this phenomenon with a heat distribution matrix  $A = \{a_{jo}\}$  where  $a_{jo}$  is the temperature increase at the inlet of server rack  $M_j$  due to the heat emitted at the outlet of the server rack  $M_o$ . Here we adjust this to matrix  $A = \{a_j\}$  denotes the heat increased at the inlet of server rack  $S_j$  caused by the heat recirculated inside the rack due to computation. Hence, the inlet temperature of a server rack  $S_j$  comes from the combination of the supplied cold air from the AC and hot air recirculated inside the rack. This relationship can be written as:

$$T_j^{in}(t) = \sum_{x=1}^l c_{jx} CTI_{jx} T_x^{sup}(t) + d_j P_j^{Comp}(t) \quad (8)$$

Where  $c_{jx}$  is a binary variable which equals to 1 if the AC unit  $C_x$  is assigned to supply cold air to the rack slot of server rack  $S_j$ , and 0 otherwise. As the DC layout is fixed, the value of  $c_{jx}$  will be given by VAR estimation.

With equations (5) and (8), we can transfer  $T_j^{in}(t)$  to the function of  $T_x^{sup}(t)$ .

$$T_j^{in}(t) = \frac{\sum_{x=1}^l c_{jx} CTI_{jx} T_x^{sup}(t) + \alpha_j E[W_j(t)]}{1 - \beta_j} \quad (9)$$

### Optimization solution

. PUE value is a measurement of the power utilization efficiency of DCs that is adopted internationally. It is the ratio of total power consumed by the DCs to the power consumed by the IT load.

$$PUE = \frac{P^{total}(t)}{P_j^{Comp}(t)} = 1 + \frac{P_x^{Cool}(t)}{P_j^{Comp}(t)} \quad (10)$$

The closer the PUE value is to 1, the higher the greenness of a DC.

$$\text{Let } G^* = \frac{P_x^{Cool}(t)}{P_j^{Comp}(t)} \quad (11)$$

Then to optimize PUE is equal to the question to minimize  $G^*$ .

Let denote  $[t_1, t_2]$  be the interval of interest. Based on the Eq. (5, 6) above, our objective function can be written as:

$$\begin{aligned} \text{Min } G^* &= \frac{P^{Cool}}{P^{Comp}} = \int_{t_1}^{t_2} \left( \frac{\sum_{x=1}^l P_x^{Cool}(t)}{\sum_{j=1}^m P_j^{Comp}(t)} \right) dt = \int_{t_1}^{t_2} \left( \frac{\sum_{x=1}^l \left( \sum_{j=1}^m \left( \frac{CTI_{xj} \sum_{j=1}^m P_j^{Comp}(t)}{CoP_x(T_x^{sup}(t))} \right) \right)}{\sum_{j=1}^m (\alpha_j E[W_j(t)] + \beta_j T_j^{in}(t))} \right) dt = \\ &\int_{t_1}^{t_2} \left( \frac{\sum_{x=1}^l \left( \sum_{j=1}^m \left( \frac{CTI_{xj} (\sum_{i=1}^m (\alpha_j E[W_j(t)] + \beta_j T_j^{in}(t)))}{CoP_x(T_x^{sup}(t))} \right) \right)}{\sum_{j=1}^m (\alpha_j E[W_j(t)] + \beta_j T_j^{in}(t))} \right) dt \end{aligned} \quad (12)$$

Align Eq.(12) with Eq.(9), we can transfer the objective function to the function of the cooling temperature combinations  $T_x^{sup}(t)$ . Hence, to minimise  $G^*$ , we can optimise  $T_x^{sup}(t)$ , which is also the decision variable in this model.

Nevertheless, the adjustment of the supplied cold air temperature is subject to the constraint that the inlet temperatures of all server racks are below the redline temperature threshold specified by the device manufacturers (i.e. typically below 25oC). Hence, based on Eq. (8)(5), the constraint of redline threshold ( $T^{red1}$ ) can be presented as:

$$T_j^{in}(t) = \sum_{x=1}^l c_{jx} CTI_{jx} T_x^{sup}(t) + d_j P_j^{Comp}(t) = \sum_{x=1}^l c_{jx} CTI_{jx} T_x^{sup}(t) + d_j. (\alpha_j E[W_j(t)] + \beta_j T_j^{in}(t)) \leq T^{red1} \quad (13)$$

Also, we will restrict the room temperature to be within the allowance (18~27 °C according to AHREA 2016). As we define the room temperature as the function of ACs temperatures and the Server racks outlet temperatures, based on Eq.(7)(5), the temperature with thresholds ( $T^{red2}$ ) and ( $T^{red3}$ ) will be represented as:

$$T^{red2} \leq T^{room}(t) = \sum_{x=1}^l h_x T_x^{sup}(t) + \sum_{j=1}^m g_j T_j^{out}(t) = \sum_{x=1}^l h_x T_x^{sup}(t) + \sum_{j=1}^m g_j (T_j^{in}(t) + K_j P_j^{Comp}(t)) = \sum_{x=1}^l h_x T_x^{sup}(t) + \sum_{j=1}^m g_j (T_j^{in}(t) +$$

$$K_j (\alpha_j E[W_j(t)] + \beta_j T_j^{in}(t)) = \sum_{x=1}^l h_x T_x^{sup}(t) + \sum_{j=1}^m ((g_j + g_j \beta_j) T_j^{in}(t) + g_j K_j \alpha_j E[W_j(t)]) \leq T^{red3} \quad (14)$$

Moreover, for each AC, there are thresholds ( $T^{red4}$ ) and ( $T^{red5}$ ), which will restrict the AC temperatures to be within the range (0 – 30°C).

$$T^{red4} \leq T_x^{sup}(t) \leq T^{red5} \quad (15)$$

Similarly, by aligning with Eq. (9), we can transfer the constraints functions Eq. (13,14) to the function of  $T_x^{sup}(t)$ .

In short, the operational problem that we will address at the first stage is to optimize the objective function in Eq. (12) by determining the optimal supplied cold temperature of the AC devices, given the constraints in Eq. (13, 14, 15). The optimization would start once we estimate the DC may perform inefficiently by VAR-RL forecasting, and target temperature combination will be set at  $t_1 - \Delta t$  periods on the timeline (Where  $\Delta t$ . is also defined by VAR-RL).

## 4 Future Studies

Future studies will be made on forecasting verification and model adjustment. An application UI (User Interface) will be applied for the DC managers to make sustainable DC management. Field trial studies will also be conducted subsequently, we will modify our models and further studies will be done accordingly.

## References

1. A. Andrae and T. Edler: On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges*, 6(1). 117–157, (2015).
2. A. Capozzoli and G. Primiceri: Cooling systems in data centers: state of art and emerging technologies. *Energy Procedia*, 83, 484–493, (2015).
3. ASHRAE: Data Center Power Equipment Thermal Guidelines and Best Practices. pp.1–60, (2016).
4. M.K. Stansberry: JulianUptime Institute 2012. Data Center Industry Survey, Uptime Institute (2013).
5. Lu, H., Zhang, Z., Yang, L: A review on airflow distribution and management in data center. *Energy & Buildings*, 179, 264–277, (2018).
6. Wang, X., Xing, G., Chen, J., Lin, C. X., Chen, Y. Intelligent sensor placement for hot server detection in data centers. *IEEE Transactions on Parallel and Distributed Systems*, 24(8), 1577–1588, (2013).
7. Volk, E., Rathgeb, D., Oleksiak, A., Volk, E., Rathgeb, D., Rathgeb, D., Oleksiak, A.: CoolE-mAll-optimising cooling efficiency in data centres. *Compute Sci Res Dev*, 29, 253–26, (2014).

8. Fu, Y., Zuo, W., Wetter, M., VanGilder, J. W., Han, X., Plamondon, D.: Equation-Based object-oriented modeling and simulation for data center Cooling: A case study. *Energy and Buildings*, 186, 108–125, (2019).
9. Song, M., Chen, K., & Wang, J.: Numerical Study on the Optimized Control of CRACs in a Data Center Based on a Fast Temperature-Predicting Model. *Journal of Energy Engineering*, 143(5), 1–8, (2017).
10. A. Beghi, L. Cecchinato, G. Dalla, and M. Lionello: Modelling and control of a free cooling system for Data Centers. *Energy Procedia*, vol. 140, 447–457, (2017).
11. J. Athavale, M. Yoda, and Y. Joshi: Comparison of data driven modeling approaches for temperature prediction in data centers. *Int. J. Heat Mass Transf.*, vol. 135, 1039–1052, (2019).
12. S. Y. Lin: Reinforcement learning-based prediction approach for distributed Dynamic Data-Driven Application Systems. *Inf. Technol. Manag.*, 16(4), 313–326, (2015).
13. A. Marinescu, I. Dusparic, and S. Clarke: Prediction-based multi-agent reinforcement learning in inherently non-stationary environments, *ACM Trans. Auton. Adapt. Syst.*, 12(2), 2017.
14. H. Nazaktabar, K. Badie, and M. N. Ahmadabadi: RLSP: a signal prediction algorithm for energy conservation in wireless sensor networks. *Wireless Networks*, 23(3), 919–933, 2017.
15. DCR Homepage, <https://datacenterresources.com/articles/increasing-crac-set-points/>. last accessed 2020/07/11
16. Q. Tang, S. K. S. Gupta, and G. Varsamopoulos: Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. *IEEE Trans. Parallel Distrib. Syst.*, 19 (11), 1458–1472, (2008).
17. R. Zhou and Z. Wang: Modeling and control for cooling management of data centers with hot aisle containment. *HP Lab. Tech. Rep.*, 82, (2011).
18. A. Beghi, L. Cecchinato, G. Dalla, and M. Lionello: Modelling and control of a free cooling system for Data Centers. *Energy Procedia*, 140, pp.447–457, (2017).
19. H. Sun, P. Stolf, J. M. Pierson, and G. Da Costa: Energy-efficient and thermal-aware resource management for heterogeneous datacenters. *Sustain. Comput. Informatics Syst.*, 4(4), 292–306, 2014.

# Escaping the Everyday Chaos: Assessing the Needs for Internal Knowledge Transfer in SMEs via an Escape Room

Irina Ehrlich<sup>1</sup>, Michael Filipenko<sup>3</sup>, Tanja Kranawetleitner<sup>1</sup>, Heike Krebs<sup>1</sup>, Robert Loew<sup>2</sup>, Diana Pistoll<sup>4</sup>, and Julia Thurner-Irmeler<sup>1</sup>

<sup>1</sup> Application Center for Materials and Environmental Research (AMU), Augsburg University, Germany

{irina.ehrlich, tanja.kranawetleitner, heike.krebs, julia.thurner}@amu.uni-augsburg.de

<sup>2</sup> DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany

loew@dipf.de

<sup>3</sup> Institute for Software & Systems Engineering (ISSE), Augsburg University, Germany

filipenko@isse.de

<sup>4</sup> Medical School Hamburg, Germany

diana.pistoll@medicalschooll-hamburg.de

**Abstract.** The implementation of an Escape Room to assess the needs for internal knowledge transfer is a method used in the project “Education 4.0 for SMEs”. The aim of this project is to support Small and Medium-sized Enterprises (SMEs) in their digitization challenges in terms of internal knowledge transfer and to help them collect the company’s knowledge among its employees. For this purpose, a needs analysis using a specially designed Escape Room adapted to a fictional company is applied. Initial results from pilot SMEs provide indications of the effectiveness of this method.

**Keywords:** Knowledge Management · Knowledge Transfer · Serious Game · Escape Room · SMEs · Needs Analysis

## 1 Introduction

Making existing knowledge visible and usable is one of the goals of knowledge management in companies. A meaningful and clear knowledge base, above all one that is actually used, is crucial in today’s world [13]. But reality is often different. There is no structured and thus efficient knowledge exchange, searches for information are time-consuming, strategies and technologies for knowledge documentation are missing, right up to knowledge loss [21].

This dilemma is the starting point of the project “Education 4.0 for SMEs”. It aims at finding out more about possible support for Small and Medium-sized Enterprises (SMEs) in their digitization challenges, especially in terms

of knowledge transfer. The project focuses on knowledge transfer in terms of exchange and distribution of knowledge among employees within a company. The exchange with partner companies is also taken into account in the project.

In the following, the development and obstacles to knowledge management in SMEs will be displayed, teamwork being one possible solution. Communication and teamwork are essential elements of Escape Rooms as well. Therefore, Escape Rooms as a Serious Game and/or as a Serious Gaming tool serving to increase positive (learning) effects are being reviewed. So far, Escape Rooms have mainly been used for educational purposes – their application in a business context has been rather marginal. Hence, the paper shows possible benefits of an employee-centered needs analysis via an Escape Room and discusses methodical issues. To this end, relevant theoretical aspects are presented, followed by a description of the design and implementation of the Escape Room, which is conceived as the starting point for further cooperation with a company in the field of knowledge management, knowledge transfer and digitization. Given the novelty of the method in this context, first results of the evaluation are discussed with regard to the further development of the method.

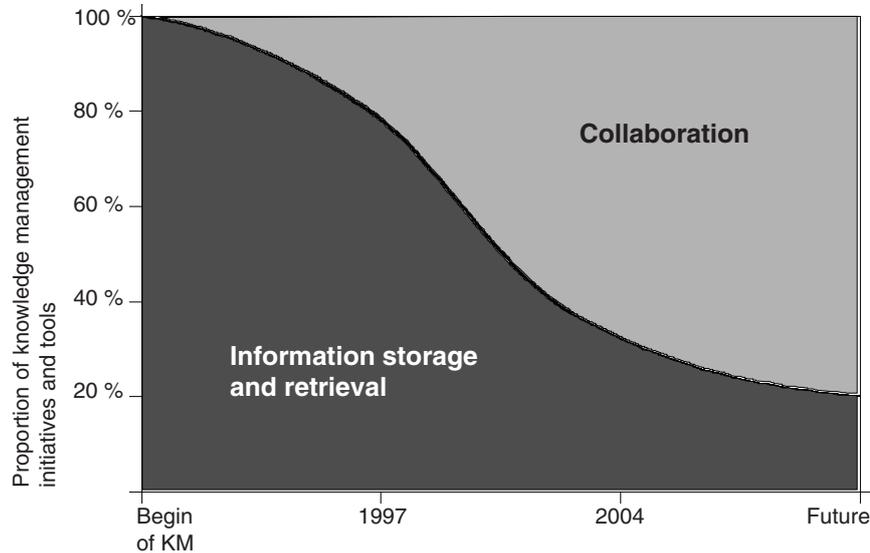
## 2 Knowledge Management in SMEs

Knowledge transfer is an important component of a knowledge management initiative, which among other things has the identification, documentation, classification and storage of important knowledge as its goal [20]. The use of knowledge stored in this way leads – especially in Information Technology (IT) management – to positive effects in a company and to innovation [38], such as the sharing and the common application and further development of knowledge [20].

The use of knowledge management in SMEs rises [25] to increase their competitiveness [36]. Successful knowledge management practices are based on technology, organization and people [43]. Technology is needed for the support of knowledge management processes. Organizational aspects, for instance clear communication structures and transparent responsibilities, lead to a company culture that focuses on the management of its knowledge as a crucial factor for competitiveness. People are the most important factor for knowledge management initiatives – they must accept and “live” knowledge management [24]. An over-estimation of technology and the negligence of people and processes are often reasons for the failure of knowledge management initiatives [15].

A trend towards a decrease in the importance of storage of information and documents in favor of support of knowledge sharing and teamwork can be discovered [20]. This may result in a solid basis of IT systems already used by most companies. In studies from 2000 and earlier, communication and collaboration were not the focus of knowledge management (see [30]). Since 2000, participants in research have asked for communication and collaboration features [24]. For example, a survey from 2015 shows that following process improvement (55 % “very relevant” / 16 % “relevant”) the enhancement of communication flows in the company was most important to the participants (43 % “very relevant” / 29

% “relevant”) [23]. Figure 1 summarizes this trend moving from storage-oriented knowledge management to supporting more and more collaboration oriented knowledge sharing.



**Fig. 1.** Trend from storage-oriented towards collaboration oriented-knowledge management [24].

The focus of SMEs’ knowledge management initiatives concerns cooperation, analysis, search and publish information [16]. The systems concentrate on employees and not on implementing a powerful knowledge management system [1]. In the practical use of knowledge management, various hurdles have been identified which must be overcome in order to use it effectively. These are for example:

- Fear of replaceability (“If I document all my knowledge, I am easily replaceable.”) [33];
- Fear of making mistakes (grammar, spelling, incomprehensible or imprecise formulations) [10];
- Assessment of the relevance of topics (“How much time do I invest in a topic?”, cost-benefit ratio) [26];
- Lack of time.

All these obstacles are mitigated by an open and communicative corporate culture [34]. This is the aim of the offers of the project: Many of the hurdles mentioned can be overcome through teamwork.

### 3 Escape Rooms as Serious Game/Gaming

Everyday life seems to require humans to cooperate with each other in order to be a functioning member of society [39]. This in turn can be seen as an important

motive for introducing cooperative learning situations in education and training, since it equips the learner with an active and more constructive role. For a few decades, collaborative learning has been viewed as an effective and valid teaching method for learning institutions, for example schools [40]. This kind of learning may be defined as a learning situation in which learners engage in a joint learning activity, enabling the construction of knowledge and/or the solution of problems in pursuit of a joint goal (see [41] and [35]).

In recent years, an increasing amount of research has introduced collaborative learning to game-based scenarios – meaning that the given educational content was made accessible through game play (see e.g. [18]). Collaborative game-based learning can be viewed as a “learning by doing” approach, which allows the learner to make mistakes while at the same time avoiding risks in the real-world [35].

Preventing risks but still possibly offering immersion in a situation, so that participants have the feeling of creating and undergoing a (real) experience, are both great benefits of Serious Games [6]. First recorded Serious Games are approaches of simulation, counting and planning games, which linked the acquirement and usage of knowledge in a playful way [2]. Even though the “carefully thought-out educational purpose” [3, p. 1] is focused by such a game, the entertainment factor should not be neglected as both parts support each other to achieve the (learning) goal [28].

### 3.1 Positive Effects in Escape Rooms

Nowadays, most Serious Games are of a digital kind, but lately another format in which the “hedonic nature of the game is highlighted” [4, p. 43] have gained worldwide popularity: Escape Rooms [9]. Escape Rooms – or Escape Games – are defined as “live-action team-based games where players discover clues, solve puzzles, and accomplish tasks in one or more rooms in order to accomplish a specific goal (usually escaping from the room) in a limited amount of time” [29, p. 1]. They focus on a team-based learning approach, in which players have to delve into a scenario and work together as a team [17]. When the design is specifically directed at reaching a certain educational goal, Escape Games are also considered to be called ‘Educational Escape Rooms’ [4] [7]. However, a concrete educational goal does not have to be of a professional or technical nature alone; the development or encouragement of, for example soft skills, can be conceivable as well [4].

In recent years, a growing interest in an industrial, respectively organizational setting can be detected (see e.g. [14]). Escape Rooms no longer seem to be exclusively a leisure activity or an irrelevant variation of everyday school life. An (Educational) Escape Room can be considered a Serious Game – but in addition, the concept of Serious Gaming is possible, too [4]: Through game-based learning – besides achieving the instruction and story objective – diverse positive learning developments and effects can occur, like generating tolerance towards others and their way of thinking and acting [12]. In most Escape Rooms, teamwork and communication are promoted [29], too. Some studies (e.g. [12]) assert that these

games can be applied as an effective method of changing attitudes towards a designated issue. Even an increase of motivation towards approached subjects and situations of Escape Rooms could be determined, as players felt “double motivated in the process of playing” [4, p. 43]. By improving the participants’ motivation, a deeper engagement with the respective learning content can be ensured [8].

### 3.2 Escape Rooms in a Business Setting

An employee-centered approach offers numerous advantages: For one, it can support the entry into digital learning and facilitate employees’ access to novel technologies. The gaming component has both intrinsic value (e.g. enjoyment) and extrinsic value, meaning that there is a learning element, respectively knowledge, to be obtained by the employee \* the latter being the main objective of an Escape Room designed for a business setting [42]. In addition, the innovative learning context enables the employee to use different strategies and “out-of-the-box” thinking, as it fosters the ability for creative problem-solving [8].

Current research on game-based learning seems to focus on learning goals or the cognition and/or meta-cognition of its participants (see e.g. [44]). This is particularly true for Escape Rooms. Aiming to assess the needs of SMEs and its employees, we did not pursue this focus for the present Escape Room. Rather, a setting was needed that put the participants in an (affective) state, so they would want to share personal and/or informal information and not just superficial or socially desired information.

Affective states – forming a superordinate category for emotions and feelings – have an impact on the learning process and educational achievements, precisely on one’s motivation to learn and achieve a certain goal (see e.g. [31] and [22]). Both positive and negative emotions seem to be a substantial part of learning. The subjective feeling that is experienced can be seen as a source of information that, for example, helps the individual to form judgment and steer cognition e.g. [37]. Information processing and the scope of attention can change due to a specific emotion that is experienced [11, p. 314]: “Many positive emotions broaden individuals’ momentary thought-action repertoires, prompting them to pursue a wider range of thoughts [...]”. Additionally, the emotions one experiences during a learning activity can help with the understanding of individual interests and developing a strong and long-lasting motivation [32]. Dampening or activating positive or negative affect can be helpful in the motivational process to gain access to specific neurological systems, thus accessing further or new information (see e.g. [19] and [27]).

Based on the empirical findings described above, one could assume that through a specific learning environment different affective states might be induced, which could in turn lead to the disclosure of new information. In this case, the emotionally charged Escape Room might enable the participants to access information in a much more flexible way. Facing, for example, a frustrating scenario but also having a sense of success in the end when the riddles are

solved, could accordingly make an Escape Room a helpful tool when evaluating the status quo, as it might provide a more detailed and realistic assessment.

## 4 The Escape Room in the Project Education 4.0 for SMEs

In the first instance, it is important for projects related to SMEs to know the current status of a company in order to support it in its further development. The project Education 4.0 for SMEs applies, among other methods, a self-developed Escape Room in order to determine the needs in the area of knowledge transfer. The following sections describe all relevant aspects of the developed Escape Room, from the identified requirements to the storyline, structure and specific elements such as a control mechanism for the Gamemaster. The chapter concludes with the main goals that the project wants to achieve by using the Escape Room.

### 4.1 Requirements

With regard to the objective of a needs analysis, several requirements have to be met: First of all, the target group should be as diverse as possible in order to get a comprehensive insight into all departments and every hierarchical level of the company. Also, it is recommendable that the Escape Room is playable for a variety of people in order to adapt to the respective situation (quantity, diversity of the group) in the company. The realistic range of participants lies between two and six. In addition to the adjustable number of people, the duration of the Escape Room or the difficulty of the riddles also needs to be individually adjusted.

Since the Escape Room serves as an introduction to the needs analysis on the topic of knowledge transfer, it is followed by a feedback discussion with reflective questions and an interactive presentation on the topic combined with individual surveys. The planned time for the entire part is one hour. Therefore the playing time should not exceed 20 to 25 minutes. Correspondingly, it must be possible to control the game from outside if necessary. The mobility of the Escape Room presents the third precondition, as it is played directly on site and therefore has to be adaptable to spatial conditions. The final requirement is to create a realistic working environment with familiar problems and situations in order to motivate and also challenge communication between the participants, in turn encouraging them to make open and realistic statements about the situation in their company. In this vein, the riddles are presented as everyday tasks one faces in an office environment.

### 4.2 Storyline, Structure and Gameplay

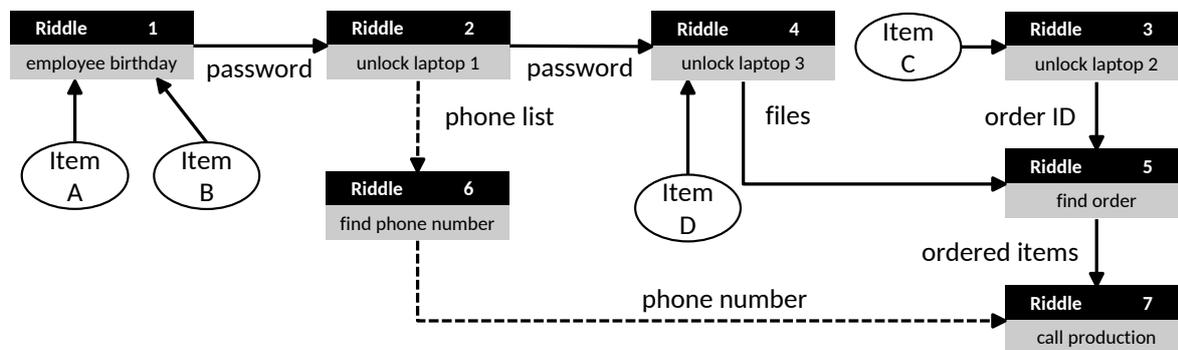
The storyline of the Escape Room is about a company named “DEMMIC”, a fictional, medium-sized company, specializing in the production of components for lightweight construction. One of its sales team members has forgotten to

forward an order to the production department before they went on their well-deserved vacation. The Escape Room takes place right after a customer has called and complained about the missing order. The participants have to gain access to their colleagues' laptops and find it.

The current situation of the company "DEMMIC" serves as a negative example of internal knowledge management:

- No coherent documentation of processes;
- No use of common company standards like file servers etc.;
- Cluttered working environment full of unnecessary stuff (e.g. not work related magazines);
- Outdated or even wrong documents.

The structure of the Escape Room and the riddles and items needed to find the missing order are shown in figure 2.



**Fig. 2.** Riddle-Flow-Chart (as seen in [7]).

After a brief introduction to the Escape Room method, the participants are directly confronted with the starting situation (they are employees of DEMMIC) and sent into the room. Suddenly an actor (the disguised Gamemaster) enters and panic-strickenly talks about the situation of the disappeared order and the angry customer who has just called. The actor also hands over a smartphone as a possibility to call for help and for the Gamemaster to give clues and leaves the room. In addition, the phone is needed for riddle 7 to call the production department with the corresponding order. As it can be seen in figure 2, the participants initially have to find several items (A - C) to solve the starting riddles (1 - 3) and to get access to two of the three laptops. To unlock the third laptop (riddle 4) they have to combine another item (D) with some information they get by solving riddle 3. By unlocking each laptop, the participants get a piece of information which is needed to find the lost order (riddle 5) and afterwards solve the final riddle (7). The dotted arrows to and from riddle 6 in the figure show an optional riddle which can be added or removed by the Gamemaster and will be described in the next section (4.3). After solving riddle 7 by calling the production department the participants can successfully escape

the room. In addition to the riddles 1 to 7, there are a lot of false clues hidden throughout the Escape Room, which is also a common feature in commercial Escape Rooms [29]. This is intended to slow down the participants and simulate a typical situation at work, when they have to search for information.

### 4.3 Control Options for the Gamemaster

The Gamemaster is aware of the progress the participants have achieved and can intervene whenever needed. To this end, participants are observed audio-visually from the outside. To give hints, the smartphone is used in order to preserve the tense atmosphere of the Escape Room. When calling the participants, the actor (the disguised Gamemaster) leaves clues like: “Have you looked into the drawer? They usually put all their stuff there.”

As mentioned in the previous section, riddle 6 is optional. This was introduced to control the playing time. Especially in the context of the Escape Room being only one of several different workshops that take place at the company, this is a valuable option. Riddle 6 is about finding the right phone number of the production department when given several, partly outdated, lists of numbers. In order to solve this riddle, the participants have to call the right number after finding the correct list. Each call is received by the Gamemaster, who, depending on the correctness of the number, can react correspondingly:

- 1) Answer the phone with: “Person X, from production department. How may I help you?”
- 2) Not answer the phone at all.
- 3) Answer the phone with some excuse that they must have called the wrong number.

When skipping the riddle, the Gamemaster just answers the first phone call with option number 1).

### 4.4 Main Goals

As the Escape Room is part of the needs assessment in the project Education 4.0 for SMEs, different objectives for further data collection are linked to the participation of the workforce. One of the main goals is to *raise awareness and sensitivity for the importance of appropriate knowledge transfer within the company*. This ranges from understanding of knowing what one knows, to articulation and transparent and comprehensible documentation of the knowledge in question. *Learning more about the current difficulties in knowledge transfer and the status quo in the respective enterprise* represents a further goal. Finally, by showing the participants a worst-case scenario of how not to do knowledge transfer and documentation, they should *be encouraged to talk openly and freely about the experiences they are facing in their everyday working life*.

## 5 Exploratory Study in two SMEs

Joint work with a company begins with a preliminary discussion with the management to identify starting points. Based on these first results, the so-called practical taster day(s) is individually organized for each company. In addition to a practical insight into the project contents, the format of the practical taster day consists of several workshops including a needs analysis. The needs analysis intends to reify the discrepancy between the desired situation and the actual situation and thus has the aim of “analyzing deficits in education, promotion and organizational development at the strategic, operational and individual level” [5, p. 825]. The management selects the representatives of the workforce who participate in the practical taster day, according to specific criteria. They represent a group as heterogeneous as possible in terms of different departments, hierarchical levels, training, gender, full-time and part-time employment, etc.

In the context of the project Education 4.0 for SMEs, the needs analysis comprises two thematic priorities: digitization and knowledge transfer. While the situation in the field of digitization is primarily evaluated by the method LEGO® SERIOUS PLAY®, the self-developed Escape Room described above primarily serves to assess needs in the area of knowledge transfer. In addition, the technical solutions for knowledge management that may already be available in the company are considered at this point as well.

The collection of data in the area of knowledge transfer is divided into two parts: 1) Participation in the Escape Room is followed by a short feedback round with reflective questions and 2) an interactive presentation about knowledge transfer including individual surveys. Furthermore, at the end of the practical taster day the participants fill in questionnaires in which, among other topics, they also evaluate the experienced Escape Room and the input for the knowledge transfer (12 items with 5 point Likert scale ranging from “fully agree” to “totally disagree”).

### 5.1 Results of the Questionnaire on the Escape Room

In the first step, some insights from the questionnaire on the Escape Room method as a tool for needs analysis will be given. It proved to be an innovative and enjoyable method, as 90.5 % ( $n = 21$ ) of the participants had not visited an Escape Room in their free time yet, while almost all of them had fun in the session during the practical taster day (95.2 % “fully agreed” and “agreed” on this item, while one was “neutral”). The difficulty and number of tasks were furthermore deemed appropriate by the majority of the participants (85.7 % and 81 % respectively for positive answers, only one participant “totally disagreed”, which might be reflected in one of the answers in the feedback session described in 5.2).

Apart from the method as such, its suitability in terms of its goal *raising awareness for the importance of appropriate knowledge transfer within one's company* was rated. To this end, the participants were asked whether they “found the ‘Escape Room’ method helpful to think about knowledge transfer in [their]

everyday work”. Although 14.3 % did not take any side and 4.8 % disagreed on this question, the statement was confirmed explicitly by 80.9 %. Also, most of the respondents (76.2 %) agreed that the method helped them “to understand the topic of knowledge transfer better”. Asked for new insights about knowledge management and transfer in their everyday work, the participants answers diverged more, ranging from “totally agree” (19 %) to “disagree” (4.8 %); the majority however “agreed” to have become aware of new aspects (61.9 %). Considering the “practical implications” of knowledge transfer, most participants were positive (76.2 %) that the Escape Room supported their awareness of the topic. However, almost a quarter (23.8 %) had a neutral stance regarding this question, showing that some issues might not be new.

## 5.2 Results of the Feedback Conversation

In order to contextualize the data from the quantitative and anonymous evaluation that was held at the end of the day, some statements from two feedback rounds held directly after the Escape Room will be drawn on. Asked about the number of people playing the Escape Room, the participants considered it to be “okay”. The participants then pointed out that, while agreeable for this playful context, it was a rather unrealistic scenario compared to their work, in that usually such issues are not solved in a team. At the same time, especially for participants without a special affinity for computers, the team was an essential factor for solving the riddles. This might be an explanation for one participant’s judgment who did not find the task’s difficulty appropriate.

Even this first question relating to the participants’ experiences in the Escape Room, which was mainly meant to serve as feedback for the further development of the method, showed the difficulty of separating methodological feedback on the Escape Room from statements about the Escape Room’s topic of knowledge management. These statements were induced in the context of the second goal, which was for both the participants and the project Education 4.0 for SMEs to *learn about the current difficulties in knowledge transfer and the status quo in the company by encouraging the participants to talk openly and freely about the experiences they are facing in their everyday working life*. To this end, the close connection between the method and the (familiar) topic of knowledge management presents a valuable opportunity to discuss inconvenient issues. In this vein, not only objective aspects were mentioned, such as the applicability of the scenario to various departments, questions of documentation tools and ways of personal knowledge transfer. Rather, these aspects served as a starting point for the possibility of expressing critical issues like the familiar situation of the absence of a colleague as well as possible solutions, for example by official arrangements or unofficial practices.

## 5.3 Results of the Questions about Knowledge Transfer

“Do you know where in your company which knowledge is available or who you can ask about a certain topic?” This question serves as the main motivation for

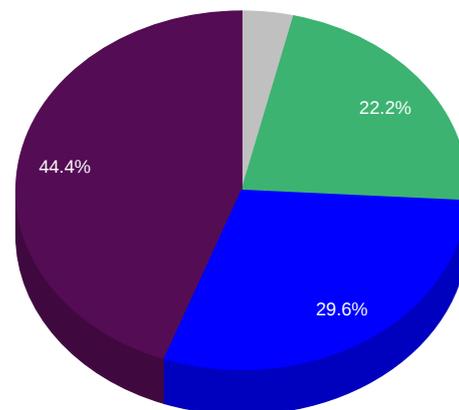
the module on knowledge transfer in the workshops during the practical taster days.

In an interactive session, the systems relevant to knowledge management are identified for the purpose of needs analysis. Based on this list, further action might be identified. The aim of the group experience in the preceding Escape Room Session is, among other things, to increase the the ability to criticize and reflect. The participants usually respond very openly to questions regarding the current situation in the company. This leads to results, such as the answers shown in fig. 3.

In the course of the project several surveys have been conducted on the topic of knowledge management in different companies. Figure 3 shows the answer to the question “How much time do you spend searching for information in your daily work?” given in the participating two pilot SMEs and three online seminars about knowledge transfer. As a result, there is a considerable potential for improvement of knowledge management, which could lead to significant time saving in practice.

### How much time do you spend searching for information in your daily work ?

Answer	%
5 minutes	3.7
10 minutes	22.2
30 minutes	29.6
more than 60 minutes	44.4



Survey results over the years 2019 and 2020.  
 n = 81 (Collected in Project Education 4.0 for SMEs needs analysis sessions and online seminars)

**Fig. 3.** Survey Results.

A similar conclusion was reached in a survey of more than 1600 office workers, which found that 25 % of respondents spent up to 60 minutes per working day searching for knowledge [21]. The needs analysis showed that companies usually have several IT systems in place that can serve as the basis for a knowledge management initiative. To the above-mentioned question whether the participants know where to find information on a topic or who to ask, most respondents answered that they roughly know this, but that there is still a considerable amount of work to be done to find concrete solutions. Participants would like to have a *simple and central means of access to relevant information*, which reduces the above mentioned search effort.

## 5.4 Lessons Learned

Due to tight constraints on the duration of the feedback session, some answers stayed relatively superficial. Giving the participants more time (e.g. about 30 minutes instead of 15-20 minutes) to reflect on the topic of knowledge transfer in their company might lead to more profound results. However, the responses received in this short period were already promising and should be examined further in future studies. A basis for this assumption is the relaxed and positive atmosphere that was indeed induced by the method, which, however, is also closely linked to the corporate culture. In the feedback session, banter on the Gamemaster's "mean" riddles and the admission that a certain riddle was "really annoying", as well as various sessions of laughter showed that the planned affective state of the participants could be achieved. Therefore, participatory observation is highly recommended for further studies in order to find out more details.

In addition to the experience gained for further data collection, certain points can also be noted for the design of the Escape Room itself. One opportunity for improvement is the clue system, in which a smartphone was used to communicate with the participants. Even though no problems have been encountered so far, the dependency on the reception level on site (especially indoors) might cause trouble in future sessions. This might be avoided by adding another way of communication, e.g. using the local company network or other internet sources. A possible further development emerged from the fact that communication can be done remotely using common messengers or video conference tools, thus shifting the whole Escape Room to an online solution. Such a digital Escape Room could be integrated better into the daily work routine, for example by decoupling it from the other workshops, thus requiring less time.

## 6 Conclusion and Further Work

In conclusion, the collected data indicate that the (method) Escape Room has certainly some justification to be used as a tool in the business setting to analyze the needs of in-company knowledge transfer management. This paper does not allow for a clear statement regarding the emotionally inducing impact the Escape Room might have had on its participants. Accordingly, it cannot be ruled out entirely that the collected data from the ensuing feedback conversation would have been generated only on the basis of the preceding Escape Room. Future research could therefore focus on the interim inquiry or retrospective judgment of participants' affective states, in order to examine and understand the effectiveness of these states more closely. In addition, the data collected so far could be further deepened within the project. For example, group discussions or individual interviews with representatives of the workforce could be considered in order to give all participants – the project and the participating SMEs – the opportunity to gain a deeper insight into the topic and especially the problems of knowledge transfer. Finally, the Escape Room method offers various potential fields of application such as Human Resources Development. Working on the

riddles together can provide insight into the team dynamics, especially if the solution is not obvious and various hurdles have to be overcome.

## Acknowledgements

The Project “Bildung 4.0 für KMU” (“Education 4.0 for small and medium-sized manufacturing enterprises”, Grant number 01PA17014) is funded by the German Federal Ministry of Education and Research and the European Social Fund for Germany within the “Digital Media in Vocational Training” program.

## References

1. Abele, E., Kuhn, S.M., Liebeck, T.: Wissensmanagement und kmu. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb **98**(7-8), 375–379 (2003)
2. Abt, C.C.: Ernste Spiele: lernen durch gespielte Wirklichkeit. Kiepenheuer & Witsch (1971)
3. Abt, C.C.: Serious games. University press of America (1987)
4. Bakhsheshi, F.F.: Serious games and serious gaming in escape rooms. In: 2019 International Serious Games Symposium (ISGS). pp. 42–47. IEEE (2019)
5. Becker, M.: Personalentwicklung: Bildung, Förderung und Organisationsentwicklung in Theorie und Praxis. Schäffer-Poeschel, Stuttgart (2013)
6. Birkenbusch, J., Christ, O.: Concepts behind serious games and computer-based trainings in health care: immersion, presence, flow. In: Serious games and virtual worlds in education, professional development, and healthcare, pp. 1–14. IGI Global (2013)
7. Borrego, C., Fernández, C., Blanes, I., Robles, S.: Room escape at class: Escape games activities to facilitate the motivation and learning in computer science. JOTSE **7**(2), 162–171 (2017)
8. Clarke, S.J., Peel, D.J., Arnab, S., Morini, L., Keegan, H., Wood, O.: EscapED: A Framework for Creating Educational Escape Rooms and Interactive Games to For Higher/Further Education. Int. J. Serious Games **4**(3) (2017). <https://doi.org/10.17083/ijsg.v4i3.180>, <http://journal.seriousgamessociety.org/index.php/IJSG/article/view/180>
9. Corkill, E.: Real escape game brings its creator’s wonderment to life. Pozyskano z <http://www.japantimes.co.jp/life/2009/12/20/to-be-sorted/real-escapegame-brings-its-creators-wonderment-to-life/#.V-BG6vCLShd> (2016)
10. Davenport, T., Prusak, L.: Working Knowledge: How Organizations Manage what They Know. EBSCO eBook Collection, Harvard Business School Press (2000), <https://books.google.de/books?id=QIyIWhdYoYC>
11. Fredrickson, B.L., Branigan, C.: Positive Emotions Broaden Thought-Action Repertoires: Evidence for the Broaden-and-Build Model. Cogn. Emot. **19**(3) (2005)
12. Glavas, A., Stascik, A.: Enhancing positive attitude towards mathematics through introducing escape room games. Mathematics education as a science and a profession pp. 281–293 (2017)
13. Gold, A.H., Malhotra, A., Segars, A.H.: Knowledge management: An organizational capabilities perspective. Journal of management information systems **18**(1), 185–214 (2001). <https://doi.org/10.1080/07421222.2001.11045669>, <https://www.tandfonline.com/doi/abs/10.1080/07421222.2001.11045669>

14. Gramß Denise AND Pillath, P.A.N.D.S.C.: Escape Room – Mitarbeitende für digitales Lernen begeistern. In: Dachzelt Raimund AND Weber, G. (ed.) Mensch und Comput. 2018 - Work. Gesellschaft für Informatik e.V., Bonn (2018). <https://doi.org/10.18420/muc2018-demo-0507>
15. Grothe, M., Landenberger, H., Schildhauer, T.: Corporate Knowledge. Business Village GmbH, Goettingen (2003)
16. von Guretzky, B.: Bausteine für eine innovationsorientierte, wissensbasierte unternehmenskultur (2004), <http://www.community-of-knowledge.de/beitrag/bausteine-fuer-eine-innovationsorientiertewissensbasierte-unternehmenskultur>
17. Hermanns, M., Deal, B., Hillhouse, S., Opella, J.B., Faigle, C., Campbell IV, R.H., et al.: Using an "escape room" toolbox approach to enhance pharmacology education (2017)
18. Kinio, A.E., Dufresne, L., Brandys, T., Jetty, P.: Break out of the Classroom: The Use of Escape Rooms as an Alternative Teaching Strategy in Surgical Education. *J. Surg. Educ.* **76**(1), 134–139 (2019). <https://doi.org/10.1016/j.jsurg.2018.06.030>, <http://www.sciencedirect.com/science/article/pii/S1931720418303283>
19. Kuhl, J., Baumann, N.: Self-regulation and rumination: Negative affect and impaired self-accessibility. *Control Hum. Behav. Ment. Process. Conscious. Essays Honor 60th Birthd. August Flammer* pp. 283–305 (2000)
20. Kuhlen, R.: Change of paradigm in knowledge management - framework for the collaborative production and exchange of knowledge. In: *Knowl. Manag.* (2011). <https://doi.org/10.1515/9783110955071.21>
21. KYOCERA-Document-Solutions: Wissensmanagement mit dms: Mit dms unternehmenswissen besser verfügbar machen (2018), <https://www.kyoceradocumentsolutions.de/de/smarter-workspaces/media-center/e-books/e-book-wissensmanagement-dms.html>
22. Lane, A.M., Whyte, G.P., Terry, P.C., Nevill, A.M.: Mood, self-set goals and examination performance: The moderating effect of depressed mood. *Pers. Individ. Dif.* **39**(1), 143–153 (2005)
23. Linde, F.: Barrieren und erfolgskriterien des wissensmanagements–ergebnisse einer onlinebefragung. *HMD* **246**, 20–28 (2005)
24. Loew, R.: Information Brokerage – A new Approach using Knowledge Management. Cork Institute of Technology, Cork, Ireland (2009)
25. Gust von Loh, S.: Wissensmanagement und informationsbedarfsanalyse in kleinen und mittleren unternehmen: Teil 2. wissensmanagement in kmu. *Information Wissenschaft und Praxis* **59**(2), 127–135 (2008)
26. Maier, R.: Knowledge Management Systems: Information and Communication Technologies for Knowledge Management. Springer Berlin Heidelberg (2004), <https://books.google.de/books?id=IQD8JCxsEyEC>
27. Martens, T.: Was ist aus dem Integrierten Handlungsmodell geworden? In: Kempf, W., Langeheine, R. (eds.) *Item-Response-Modelle in der sozialwissenschaftlichen Forschung*. Regener, Berlin (2012), [https://www.pedocs.de/volltexte/2016/12327/pdf/Martens\\_2012\\_Was\\_ist\\_aus\\_dem\\_Integrierten\\_Handlungsmodell\\_geworden.pdf](https://www.pedocs.de/volltexte/2016/12327/pdf/Martens_2012_Was_ist_aus_dem_Integrierten_Handlungsmodell_geworden.pdf)
28. Michael, D.R., Chen, S.L.: Serious games: Games that educate, train, and inform. Muska & Lipman/Premier-Trade (2005)
29. Nicholson, S.: The state of escape: Escape room design and facilities. *Meaningful Play* (2016)
30. Nohr, H., NOHR, H.: Wissensmanagement in stuttgarter unternehmen–ergebnisse einer umfrage/fachhochschule stuttgart. 2000. *Arbeitspapiere Wissensmanagement* (10) (2000)

31. Pekrun, R.: Academic emotions. *Handb. Motiv. Sch.* **2**, 120–144 (2016)
32. Pekrun, R., Stephens, E.J.: Academic emotions. In: *APA Educ. Psychol. handbook, Vol 2 Individ. Differ. Cult. Context. factors.*, pp. 3–31. American Psychological Association (2012)
33. Probst, G., Raub, S., Romhardt, K.: Managing knowledge-Building blocks for success. *Organ. Stud. Gr. Organ. Stud.* **22**(1), 186 (2001)
34. Riege, A.: Three-dozen knowledge-sharing barriers managers must consider (2005). <https://doi.org/10.1108/13673270510602746>
35. Romero, M., Usart, M., Ott, M., Earp, J., de Freitas, S., Arnab, S.: Learning through playing for or against each other? Promoting collaborative learning in digital game based learning (2012)
36. Schauer, H., Frank, U.: Methoden des wissensmanagements. *Wissensmanagement für KMU* pp. 147–166 (2007)
37. Schwarz, N.: Feelings-as-information theory. *Handbook of theories of social psychology* **1**, 289–308 (2011)
38. Soto-Acosta, P., Popa, S., Conesa, I.: Information technology, knowledge management and environmental dynamism as drivers of innovation ambidexterity: a study in smes. *Journal of Knowledge Management* pp. 824–849 (02 2018). <https://doi.org/10.1108/JKM-10-2017-0448>
39. Tomasello, M., Carpenter, M.: Shared intentionality. *Dev. Sci.* **10**(1), 121–125 (2007). <https://doi.org/10.1111/j.1467-7687.2007.00573.x>, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-7687.2007.00573.x>
40. van der Linden Jos and Erkens, G., Henk, S., Peter, R.: Collaborative Learning. In: Simons Robert-Jan and van der Linden, J., Tom, D. (eds.) *New Learn.*, pp. 37–54. Springer Netherlands, Dordrecht (2000). [https://doi.org/10.1007/0-306-47614-2\\_3](https://doi.org/10.1007/0-306-47614-2_3)
41. Vuopala, E., Hyvönen, P., Järvelä, S.: Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education* **17**(1), 25–38 (2016). <https://doi.org/10.1177/1469787415616730>
42. Warmelink, H., Mayer, I., Weber, J., Heijligers, B., Haggis, M., Peters, E., Louwerse, M.: AMELIO: Evaluating the team-building potential of a mixed reality escape room game. In: *Ext. Abstr. Publ. Annu. Symp. Comput. Interact. Play.* pp. 111–123 (2017)
43. Wolf, T., Decker, S., Abecker, A.: Unterstützung des wissensmanagements durch informations- und kommunikationstechnologie. In: Nüttgens, M., Scheer, A.W. (eds.) *Electronic Business Engineering*, pp. 745–765. Physica-Verlag HD, Heidelberg (1999). [https://doi.org/10.1007/978-3-642-58663-7\\_39](https://doi.org/10.1007/978-3-642-58663-7_39)
44. Woo, J.C.: Digital game-based learning supports student motivation, cognitive success, and performance outcomes. *Journal of Educational Technology & Society* **17**(3), 291–307 (2014), <https://www.jstor.org/stable/10.2307/jeductechsoci.17.3.291>

# Determinants for the adoption of Regulatory Technology (RegTech) services by the companies in United Arab Emirates: An MCDM Approach

Mohamed Muzammil<sup>1</sup>[0000-0002-6575-2719] and Nitin Simha Vihari<sup>2</sup>[0000-0001-8501-9795]

<sup>1,2</sup> BITS Pilani, Dubai Campus, Dubai, United Arab Emirates  
f20170258@dubai.bits-pilani.ac.in, nitinvihari@dubai.bits-pilani.ac.in

**Abstract.** The increasing compliance expenditure, identity management and increasing need to secure data constitute risks to current regulatory procedures. The concept of Regulatory Technology, or RegTech, in the United Arab Emirates (UAE) is still in the infant stage due to lack of defined processes and it holds immense potential in employing RegTech to its existing institutions. The purpose of this paper is to understand the various factors that determine the adoption for RegTech services and proposes a multi-criteria decision-making approach by ranking those determinants. The determinants are divided into three main hierarchies which include customers, regulators, and environment-based factors. Survey was conducted to rank each of the determinants within the hierarchy to validate the need for RegTech services. It is observed that customer-centric experience, enhancing data security and ensuring real-time data capabilities play the major factors for the implementation of RegTech services that is validated by constructing an AHP model. The need for RegTech's determinants would help financial institutions, regulators, and stakeholders in the UAE benefit from the research and achieve competitive advantage by understanding and implementing the suggested findings.

**Keywords:** RegTech, Compliance, Regulatory Monitoring, Analytical Hierarchy Process

## 1 Introduction

The dawn of the 21st century has largely changed the landscape of financial services as a whole because of a rapidly changing and development of technology. The evolution of Financial Technology, or FinTech, which collaborates technology with creating financial solutions has long led to changing the financial markets [2]. FinTech has entered an industrialization phase marked by the emergence of start-ups and other newcomers, such as e-commerce and IT enterprises, which have segmented the financial services market. Regulatory technology or RegTech, a booming tech word at this age with the rise of technology and a growing focus on data, has made it the focus of attention. Most of the financial heads agree that the real cause of the 2008 global

housing and financial crisis were due to unscrupulous private mortgage loans and unregulated markets [4]. With the global financial crisis of 2008, many regulators have come together to create regulatory solutions that could help manage the compliance risks of the organization. The term RegTech was first coined by the UK's Financial Conduct Authority (FCA) in 2015 who defined it as: "A subset of FinTech that focuses on technologies that may facilitate the delivery of regulatory requirements more efficiently and effectively than existing capabilities." [22]

Existing financial agencies and their regulations for risk management and compliance were previously the main driving force and sector of RegTech solutions. Although the financial sector has long been such a significant user of algorithmic reporting and compliance tools, which have increased regulatory burdens since 2008, this is an opportunity for success implementation of digitization and automation processes has been strengthened [3]. Regulation on financial services has expanded at a significant rate since the financial crisis and hence the cost of regulatory compliance has expanded [7]. Failures to comply have also become costly. The finance sector has paid \$321 billion in fines and sanctions over the last 10 years. But this does not count the adverse effect of fines on the reputation of a company and on the credibility of the shareholders [10]. With the 2008 Global financial crisis, many regulators have come together into creating regulatory solutions that might help manage the organization's compliance risks [8]. RegTech companies collaborate with various institutions and regulatory bodies in coming up with technological solutions basically covering challenges involved with regulatory monitoring and compliance so as to cut back their regulatory failures thereby saving their time and money. In today's world, where data is of prime importance, RegTech has become a source of competitive advantage and can help secure data which can satisfy the regulator's demands and deliver effective insights about customers and the market [4]. The ongoing COVID-19 pandemic puts an unimaginable strain on markets, governments, corporations, and individuals and has shaken the nation's core around the world and has assessed their infrastructure for healthcare, pummeled their financial markets, and left entire populations in fear and lockdown. Strategies for tackling the financial aspects of the crisis to reduce the economic and human impact team include ensuring sufficient liquidity, intensifying information exchange to ensure accurate information, leveraging digital finance payments to reduce human-to-human contact and many more. At the same time, the digitization of financial services over the past decade offers alternative and more direct ways of stimulating the real economy, which will be critical in minimizing economic impacts and preserving economic stability post-COVID [2]. RegTech was initially regarded as a subset of FinTech's, which evolved to help companies move away from the concept of 'Big data' to digitized 'smart data'. Gradually, with the advent of new technology and increasing compliance protocols, RegTech is rising as a separate industry focusing on providing compliance and regulatory monitoring solutions to various companies. We need to embrace new technology to help clear the fog to steer the financial regulatory atmosphere, transforming what was perceived as a limitation into an attempt that contributes to our customers, and thus to the market. RegTech in the UAE, as of 2020, is currently in the developing stage as not many companies consider RegTech as a separate entity. RegTech is still "nascent" in the Middle East. While

there is considerable diversity in the pace with which the UAE adopts regulatory solutions, overall investment was low compared with other regions. RegTech's adoption in the MENA regions is still underway, but UAE has started organizing regional conferences to explore its prospective and existing adoption restrictions [12]. This is mainly due to the UAE government that has been keen to adopt new technologies and to the UAE's greater national push to become a world leader in FinTech innovation. The MCDM analysis, which is being proposed in this project, might help companies adapt to RegTech to ensure a seamless experience for companies to enact financial services as per compliance.

## 2 Literature Review

The increase in regulatory and supervisory scrutiny after the financial collapse means that financial services firms are spending ever greater amounts of money and resources to manage their compliance risk. That cost has now reached unjustifiable levels in the eyes of some investors in the current low productivity environment [15]. The application of new technologies in the financial services sector can bring clear, distinct advantages. Dubai International Financial Center and Abu Dhabi Global Market, two of the MENA region's leading free zones, each seek to appeal to financial services technology developers by setting up innovative environments where FinTech solutions can be developed and, to a certain extent, into the wilderness [9]. RegTech seeks to use innovative technologies such as artificial intelligence, block chain, analytics of big data, cloud computing and biometrics. Its main purpose is to tackle issues of compliance and risk management and facilitate for more effective management of compliance costs and reduction of operational risks [15]. The need for risk and compliance functions to have more involvement in assessing the impacts of FinTech innovation has increased remarkably. Compliance modelling is a broad policy comprising the direction it enforces the function, the technology, and analytics it uses, the amount and essence of its connections to other parts of the business, the standards it has been assigned and more. Innovating compliance can be the most complex and sophisticated way for an organization to monitor what is going on inside its four walls and what is coming from outside [16]. The percentage of firms that identified their risk and compliance functions as having to participate more in Fintech, RegTech and InsurTech has more than doubled from 15% in 2017 to 32% in 2018. Those reporting that their risk and compliance function is fully engaged and consulted in their company's approach to fintech from 37% in 2017 to 18% in 2018 have a parallel drop [7], [11].

Regulators around the world are encouraging fintech's potential benefits while minimizing potential risks and challenges. This is especially the case where loyal customer outcomes could be compromised [8]. A regulatory approach is that of a sandbox where new solutions can be tested and developed without having to comply fully with

all relevant regulatory requirements from the start. Soon, regulators will be under immense pressure to migrate to the market that includes large banks, emerging tech firms and lean startups. RegTech will not only be used to help authorities control and regulate industry participants but also to identify when to do so. There are advantages for industry and regulators alike. It can enable financial institutions to control economic costs more efficiently for business, release regulatory surplus capital, and offer new potentials for FinTech startups, consultative companies, and tech firms [19], [6], [9]. It helps regulators develop continuous monitoring tools to detect problems as they evolve and minimize the time it takes to investigate compliance violations, also promoting the development of simulation systems and sandboxes that can predict the possible impacts of new approaches and changes. To identify the determinants associated with RegTech, various research articles published in reputed journals were studied. The database of Top RegTech studies of 2019 were looked upon to look for data associated with RegTech. Similarly, the Google Scholar engine was also explored to find out the top driving factors that determine RegTech from various research papers. The study carried out by top financial institutions was observed. After carefully going through the abstracts of the research papers, 10 key determinants were chosen. While analyzing the determinants for RegTech, the focal point must be to identify why companies must adopt RegTech services. The extensive literature study was carried out to identify what exactly is RegTech. The following paragraphs explain the determinants of RegTech in bold to assist companies in U.A.E. understand the benefits of the services RegTech has to offer.

Identifying Financial crime poses a considerable threat to the overall integrity, stability, and development of the financial services industry. They also have the potential for strengthening business performance and effectiveness within financial institutions [17], [20]. Whether it's cyber fraud, terrorist financing or the history of corruption is full of the trail of economic devastation left by financial crimes. In 2012, HSBC Holdings faced a backlash and was fined \$1.92B by the US authorities for allowing laundering drug money flowing in and out of the cartel [22]. With the advent of advanced technologies, the regularity and sophistication of financial crime has significantly increased.

According to a study conducted by Thomson Reuters, one of the leading multinational media conglomerates, compliance budgets continue to be on the rise. There has been a year on year rise on compliance budget with 63% of firms in 2019 compared to 61% in 2018. Lack of compliance budget constitutes to 13% challenge for compliance officers in 2019 [23]. This increasing compliance budget is supported by the fact that more numbers are opting for automation of compliance activities, continuing regulatory change and enhancing the role of compliance within the businesses to assure their stakeholders' trust in the company [18], [14].

RegTech can be used to automate compliance tasks which involve utilization of advanced algorithmic processes. Companies providing technology-based solutions

have been on the rise lately. With the advancement of technologies, there has been an increase in competition from many companies into adopting the latest technologies to satisfy their customer needs [12], [5].

Artificial intelligence will supposedly play a major role in driving RegTech to success. Linking analytics with artificial intelligence will provide agility of data and ensure a complete user-friendly experience while ensuring compliance [7], [14].

Many startups have been emerging over the past couple of years into bringing the best regulatory solutions into the market. This has led to an increase in innovation and startups and consequently, led to an increase in the competition with more startups, there is an increasing possibility of more risks involved. RegTech identifies these risks into creating a more proactive environment. Because of this and many other potential hazards, regulators and firms see significant opportunities to use the latest technologies used by financial firms to manage risks and improve market efficiency, security, and soundness. The main task of RegTech is to help companies to adjust their work to the requirements of legislation. The demand for such services is growing because traditional legal processes are becoming too expensive in terms of resources and time. [13].

We discussed how one of the main objectives for RegTech is moving from ‘*big data*’ to ‘*smart data*’ thereby offering real-time data capabilities to ensure a timely response and have a lower turn-around time. Modern times call for data to be available at our fingertips for quick analysis and a better understanding of the problem to reach a quick solution. This can also help make sure timely reporting, as these reports can be instantly optimized and extracted allowing data-driven compliance and constructive regulation. RegTech offers the possibility for financial firms to clump their raw data to enhance agility through these algorithmic processes [10].

Another driver is the similar complexity of financial institutions served in line with its business concepts, mechanisms for business entities and markets. Tracking, analyzing and compliance to existing and proposed regulations is a challenge even for the major banks. The transition from individual solutions to RegTech is reinforced by the fact that the major portion of the costs are spent on consulting firms, consulting services and IT professionals [14].

Companies are evolving at a rapid pace to highly focus on acquiring smart data for their processes. As our financial system moves from one based on the principles of Know Your Customer through Know Your Data approach, a new regulatory framework must evolve that will have to deal with regulating identity assurance and focus on data privacy in financial industries and also facilitate due diligence and Know Your Customer (KYC) procedures, screening and detection of AML and anti-fraud [16]. There have been significant advancements in KYC procedures with a data-driven process to identify their customers and gain background information to assess the risk factors involved in the transactions.

While tackling the cost of spiraling compliance and avoiding hefty fines may provide banks with plenty of motivation to navigate the rewards of RegTech solutions, there is more to it. By restructuring processes, organizing data appropriately, developing innovative new solutions will be put in place to provide customer-centric experience while also providing regulatory challenges with solutions [1], [5].

RegTech also makes sure financial crime is at the minimum as it also focuses on enhancing data security and protecting individuals' data. This has been a primary key in why many companies adopt RegTech services to gain their customers' as well as their stakeholder's trust. In today's current era, data security has been of utmost importance as it affects protective measures to secure data from security breaches and data mismanagement throughout the data life span [20]. The 10 key determinants for adoption of RegTech services discussed are tabulated in Table 1.

**Table 1.** The selected drivers for adoption of RegTech services

<b>Determinants</b>	<b>Authors</b>	<b>Description based on literature review</b>
Real-time data capabilities (V1)	Anagnostopoulos (2018)	RegTech offers access to real-time data and act instantly based on customer requirements to develop real-time, proactive responses
Relative complexity of financial institutions (V2)	Butler (2019), O'Brien (2019)	Complexity in terms of business designs, mechanisms of legal entities and procedures is a major driving force for RegTech. It is a challenge to monitor, interpret, and comply with current and planned regulations
Automated advanced algorithmic processes (V3)	Anagnostopoulos (2018)	Application of technology can help automate the performance of regulations thereby reducing the time involved
Linking advanced models and analytics with artificial intelligence (V4)	Arner (2017), Barberis, (2017), & Buckley (2017)	RegTech provides potential for constant monitoring and real-time perspectives with the help of deep learning and artificial intelligence
Increasing need to identify financial crime (V5)	F.C. Authority (2015)	Knowing the cause of the issue can help to solve most of the problem. RegTech can help in identifying any crime so that the perpetrators can be caught in real-time and help in reducing financial loss
Customer-Centric Experience (V6)	Lootsma (2017)	Develop a customer experience-based platform that provides a personalized experience and enhance digital screening of customer through the help of KYC
Encourage Innovation and Competitiveness of Business (V7)	Arner (2017) and Buckley (2017)	New approaches to regulation must be developed to attempt the balance with innovative technology and urbanization benefits with economic stability

Increase in compliance expenditure (V8)	English (2019) and Hammond (2019)	RegTech adoption in developed economies is driven mainly by the increasing cost of compliance. The rapid increase of the compliance budget is a clear indication that RegTech is proving to be successful.
Identity management and assurance (V9)	Johansson (2019), Sutinen (2019), Lassila (2019), Lang (2019), Martikainen (2019), MLehner (2019)	RegTech companies aim to focus on identity management, security, and data privacy in financial services. Involve diligence and understanding of AML and anti-fraud procedures.
Enhance Data Security (V10)	Anagnostopoulos (2018)	To prevent unauthorized access to individuals, cyber security becomes a key driving force in developing RegTech

### 3 Methodology

This study introduces the Analytic Hierarchical Process modelling and the Total Interpretive Structural modelling technique to analyze drivers of RegTech in a systematic way and a more comprehensive approach. In this manner, the competitiveness of each determinant is analyzed. The measurement of degree of each RegTech determinant is a complex task which seeks to enhance the adoption of RegTech services. Multi Criteria Decision Making is employed to solve such kinds of problems. MCDM problems are determined as challenges in making decisions that comprise several parameters. In the MCDM approach, a certain problem is expressed in linguistic terms which are then mapped into fuzzy numbers. The collective study of criteria helps us to improve comprehension between them of direct and indirect relationships, rather than studying them individually [2]. There are many approaches into tackling an MCDM problem such as Analytical Hierarchy Process (AHP), Total Interpretive Structural Modelling (TISM). AHP is one such popular methods of decision-making process which is being discussed in this paper.

The analytical hierarchy process (AHP) is one of the most frequently used strategies to decide a priority in various fields. Developed by Saaty in 1980, it breaks up the decision-making problem into a system of hierarchies of determinants. AHP can have as many objectives as possible to fully characterize the problem, thereby helping to understand the logic behind choosing this decision. The prominent feature of this strategy is a check of the consistency of the expert decision making in comparison matrix development. The AHP has qualitative and quantitative factors to it. It can handle several criteria and sub-criteria effectively. The main objective of constructing an AHP model is to prioritize the indicators which can assist with decision making

involving multiple criteria. The AHP method is so chosen in this paper to assist with decision making involving multiple objectives. Due to its pairwise comparisons, AHP requires ratio scales. Saaty argued that the ratio scales in the form of a comparison scale, also known as Saaty's fundamental scale is the only possible measurement to calculate a weighted sum of the objectives. In original Saaty's AHP, the verbal statements are represented by scale with measures from one to nine. The Saaty's fundamental scale is the only scale used in AHP. Since we are doing a pairwise comparison of linguistic objectives, a reliable scale is required. Saaty proposed a 9-point scale for analytic hierarchy studies with the assumption that the objectives are of the same order and magnitude and their relative weights do not differ more than 9. The 9-point scale was so selected since it offers a wide range of levels and was found to be highly reliable in several scale development studies. This paper uses radical root method and devises the AHP as follows [25].

*Step 1:* Identify the problem and its objectives. Define the factors affecting the problem which can be obtained through studying literature reviews, conducting surveys or by having brainstorming sessions.

*Step 2:* Construct a pair-wise comparison and design a matrix for each of the determinant on a scale of 1-9 using Saaty's fundamental scale given in the following Table 2 [23].

**Table 2.** The fundamental scale for pair-wise comparisons

Intensity of importance	Definition	Explanation
1	Same	Neither of the two alternatives are preferred over another
3	Weak	The selected alternative is preferred slightly over the other
5	Clear	The selected alternative is preferred clearly over the other
7	Strong	The selected alternative is preferred strongly over the other
9	Very Strong	The selected alternative is preferred very strongly over the other
2, 4, 6, 8	Compromise	Can be used as a compromise for alternative grading

*Step 3:* Find the relative priority vector ( $w_i$ ) of each determinant by calculating the geometric mean of the  $i^{\text{th}}$  row and normalizing the geometric mean of the rows in the comparison matrix. This is represented as follows:

$$\text{Geometric Mean (GM)} = \prod_{i=1}^N (a_i)^{1/N} \quad (1)$$

where  $a_i$  denotes the comparison value in the given cell and  $N$  refers to the number of determinants in the specified objective.

The relative priority vector ( $w_i$ ) is given by:

$$W_i = \frac{GM}{\sum_{i=1}^N (GM)} \quad (2)$$

*Step 4:* Calculate the respective sum of each of the columns of the obtained matrix and multiply it with the corresponding priority vector in the row.

*Step 5:* Determine the maximum eigenvalue  $\lambda_{max}$  that is the sum of the results in the row obtained from Step 4.

*Step 6:* Calculate the consistency index (CI)

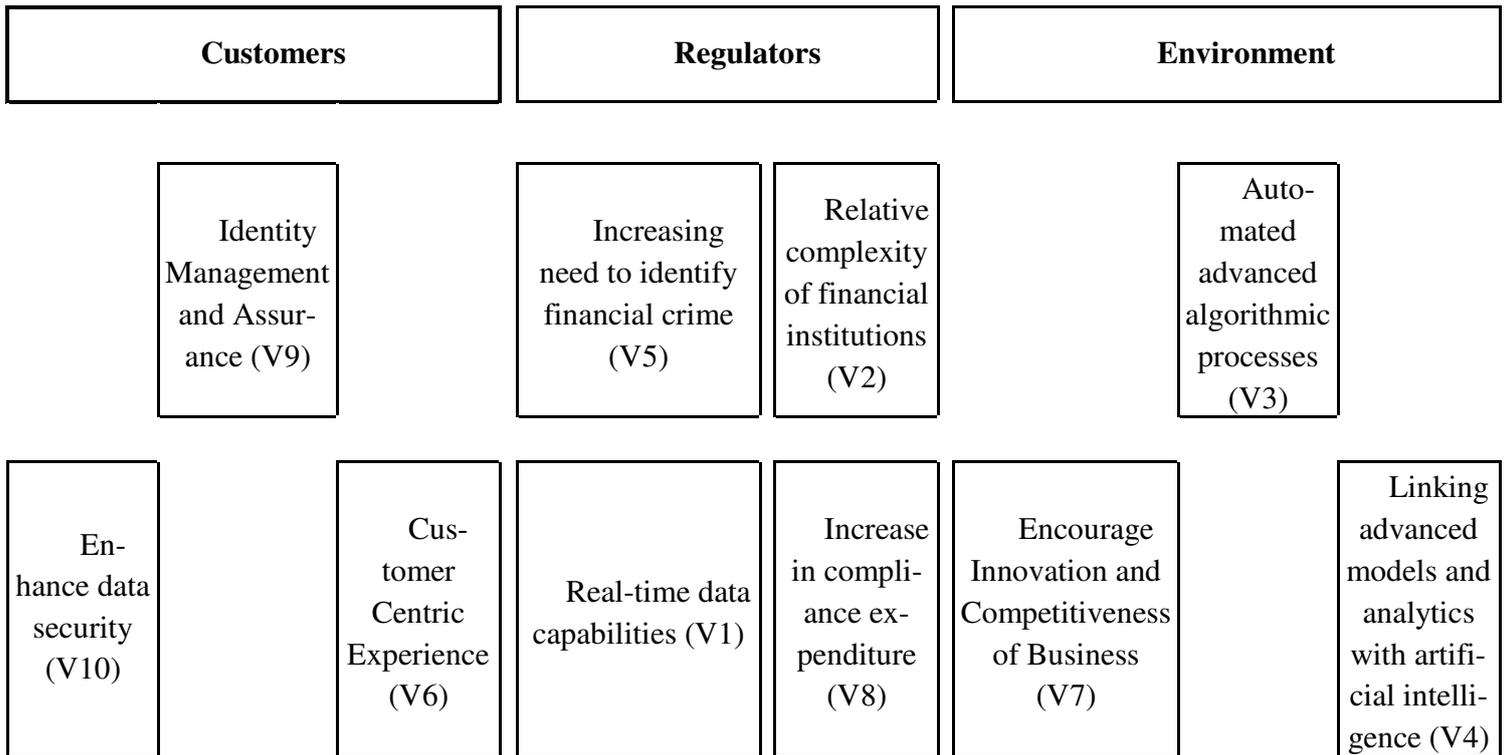
$$CI = \frac{\lambda_{max} - N}{N - 1} \quad (3)$$

*Step 7:* Obtain the random index (RI) for the number of objectives used in decision making from Table 3.

**Table 3.** Random Index (RI) Values

Attributes	3	4	5	6	7	8	9	10
Random Index (RI)	0.52	0.89	1.11	1.25	1.35	1.4	1.45	1.49

*Step 8:* Calculate the consistency ratio  $CR = CI/RI$ . A CR of 0.1 or less is normally considered acceptable and reflects an informed rational judgment attributable to the analyst's knowledge of the issue under study [13]. Based on the determinants, we classified them into the type of stakeholder the determinant belongs to easily be able to analyze it. In a RegTech company, there are mainly three stakeholders that the company can function upon. Based on these stakeholders, we divide our problem into three main hierarchies. These hierarchies are defined to assist the decision maker to address which alternative is best suited for the growth of its company. The determinants which were determined were classified into three hierarchies, namely -- customer based, regulatory based and environment based. A questionnaire was developed with the help of Google forms for pilot testing of construct validation to validate the competitiveness of determinants for RegTech from the literature review to substantiate the competitiveness of RegTech determinants. The survey conducted was primarily used to determine the pairwise comparison of each of the grouped determinants. The survey was sent on various social media platforms to obtain a viewpoint on the determinants selected on a 9-point comparison scale (1 → determinant is equal; 9 → determinant is most important). 75 responses were received after the survey was sent out across the social media platforms. The survey was conducted in the middle of the pandemic during April-May 2020. A standard Google Form was sent to the employees through email and the respondents included customer relationship managers, account managers and sales managers of a top IT firm.



**Fig. 1.** Classification of determinants into three hierarchies (Source: Authors' Representation)

## 4 Results & Discussion

The survey was conducted in three parts -- by pairwise comparing customer-based, regulatory-based and environment-based determinants and then comparing them overall. Based on the responses collected, the following pairwise comparison matrix was obtained.

**Table 4.** Customer-based pairwise comparison

	V6	V9	V10
V6	1	5	3
V9	0.2	1	0.33
V10	0.33	3	1

The eigen value  $\lambda_{max}$  obtained after AHP analysis for customer-based pairwise comparison was 3.051. The consistency index (CI) was calculated to be 0.0193 and the consistency ratio was calculated as 0.037 which is below 0.1 and is an acceptable value.

**Table 5.** Regulatory based pairwise comparison

	V1	V2	V5	V8
V1	1	5	5	3
V2	0.33	1	3	5
V5	0.2	0.33	1	3
V8	0.33	0.2	0.33	1

The eigen value  $\lambda_{max}$  for regulatory-based comparison obtained is 4.375. Consistency index is 0.125 and the Consistency ratio is 0.1404.

**Table 6.** Environment-based pairwise comparison

	V3	V4	V7
V3	1	0.142	0.33
V4	7	1	3
V7	3	0.33	1

The eigen value  $\lambda_{max}$  for environment-based comparison obtained is 3.007. Consistency index (CI) is 0.0035 and the Consistency ratio is 0.00675. Since the CI value is less than 0.1, it implies that the obtained readings are consistent.

Based on the above results, it is evident that the AHP has produced satisfactory results as the consistency ratio was well below 0.1 for customer-based and environment-based pairwise comparison. After obtaining the pairwise comparison of each matrix, the weights of each of the determinant pertaining to the given hierarchy was calculated by first taking the geometric mean of the row of the matrix.

The weights were then calculated by dividing the geometric mean of the given determinant to the sum of the geometric mean. The ranking is determined by giving priority to the highest weights and subsequently ranking each determinant based on the hierarchy.

**Table 7.** Local weights and rankings of customer-based hierarchy

	Local weights	Ranking
V6	0.63699	1
V9	0.10473	3
V10	0.25829	2

**Table 8.** Local weights and rankings of regulatory-based hierarchy

	Local weights	Ranking
<b>V1</b>	0.50387	1
<b>V2</b>	0.29091	2
<b>V5</b>	0.13009	3
<b>V8</b>	0.07511	4

**Table 9.** Local weights and rankings of environment-based hierarchy

	Local weights	Ranking
<b>V3</b>	0.087946	3
<b>V4</b>	0.66941	1
<b>V7</b>	0.24264	2

To determine the overall weight of each of the determinant, the weights of the hierarchies are determined by constructing a pairwise matrix and calculating the weights likewise.

The matrix was obtained from the survey responses as below:

**Table 10.** Hierarchical pairwise comparison

	Customers	Regulators	Environment
<b>Customers</b>	1	3	5
<b>Regulators</b>	0.333	1	3
<b>Environment</b>	0.2	0.333	1

The eigen value  $\lambda_{max}$  obtained after AHP analysis for hierarchical pairwise comparison was 3.039. The consistency index (CI) was calculated to be 0.01925 and the consistency ratio was calculated as 0.037 which is below 0.1 implying consistency and a satisfactory result.

The consistency ratios of all the payoff matrices are well below 0.1 which is the acceptable value. This implies that the comparisons are of good consistencies. The weight of each hierarchy is calculated by dividing the geometric mean of each row to the sum of geometric mean.

**Table 11.** Weight of Hierarchy

	Weights	Ranking
<b>Customers</b>	0.63699	1
<b>Regulators</b>	0.25828	2
<b>Environment</b>	0.10473	3

To calculate the global weight of each of the determinants, each of the determinant's local weight is multiplied by its hierarchical local weight. This will give the ranking of each of the determinants and help in prioritizing the determinant while adopting RegTech services. The priority for ranking is given for the determinant having the largest weight. The above Table 11 shows the weightage of each hierarchy after developing AHP and Table 12 shows the weightage and ranking of each of the determinant. The global weights give the overall weightage of each determinant while the local weights give the weightage of the determinant within its predefined hierarchy.

Keep in mind that the local weights give the weightage of the given determinant against its allocated hierarchy. For example, the local weight of Customer-Centric Experience (V6) is 0.6369 or 63.69%. This is the weight of the given determinant inside the Customer-based hierarchy. The global weight or the overall weight of the determinant is calculated by multiplying the local weight to the weight of Customer-Based hierarchy, i.e.,  $0.6369 * 0.6369 = 0.40575$  or 40.57%. From Table 11 of the hierarchies, customers are ranked 1st in the list with a weightage of 63.7% followed by Regulators at 25.83% and Environment at 10.47%. Of the determinants from Table 12, customer centric experience is getting the highest weightage of 40.58%. This is evident from the fact that the RegTech solutions are tailor made exclusively for the customers to provide a seamless user experience. The UAE is a major hub that prioritizes delivering to its customer and is keen on enhancing and regularly upgrading customer satisfaction to promote happiness and wellbeing. Data security comes second with 16.45% which is a prime factor in gaining customers and the stakeholders' trusts in the business.

**Table 12.** Global weights and rankings of determinants of RegTech

Determinant	Local Weights		Global Weights	
	Weights	Ranking	Weights	Ranking
Customer-Centric Experience (V6)	63.70%	1	40.58%	1
Enhance Data Security (V10)	25.83%	2	16.45%	2
Real-time data capabilities (V1)	50.39%	1	13.01%	3
Relative complexity of financial institutions (V2)	29.09%	2	7.51%	4

Linking advanced models and analytics with artificial intelligence (V4)	66.94%	1	7.01%	5
Identity management and assurance (V9)	10.47%	3	6.67%	6
Increasing need to identify financial crime (V5)	13.01%	3	3.36%	7
Encourage Innovation and Competitiveness of Business (V7)	24.26%	2	2.54%	8
Increase in compliance expenditure (V8)	7.51%	4	1.94%	9
Automated advanced algorithmic processes (V3)	8.79%	3	0.92%	10

The UAE has set in place many legislations to protect the data and the privacy of the persons and the companies. Ensuring that data is provided in real-time becomes another major driver for RegTech services and holds 13.01% weightage. Technology plays a major role in the success of RegTech services. The existing regulatory processes often involve manual and error-prone methods with regards to old management protocols that lead to chaotic and inefficient environments. RegTech solves this issue by ushering into the digital data era using the latest and the most efficient technology. The rest of the determinants hold low weightage and are dependent on other determinants.

## 5 Discussions and Implications

The study examined the importance of RegTech and the driving factors behind RegTech. The findings through an extensive literature review proposed 10 determinants for adoption of RegTech services. With the help of Multi Criteria Decision Making Approach (MCDM), AHP analysis was done to understand the focus of RegTech services and an attempt to understand the major driving factors of RegTech. To further enhance customer satisfaction, companies in the UAE must be ready to adopt RegTech that promises to provide a customer-centric, user friendly experience to its valuable customers. The findings of this study contribute to the existing companies in UAE to heavily focus on adopting RegTech within their businesses [3]. Based on the AHP findings, it is observed that customers play the major driver factor in RegTech determinants. It is evidenced by the fact that the UAE government lays special significance on consumer protection and consumer rights. This paper presents one of many studies that attempt to promote the adoption of RegTech services in the United Arab Emirates. RegTech can greatly favor UAE markets as it brings in a standardized form of the regulatory domain mixed with the vast array of new technologies across the world. The UAE has already made strides into embracing the concept of RegTech within their companies. This has already begun with Abu Dhabi Global Market launching sandbox programs designed to simulate the structure and functioning of RegTech within their businesses. The findings will help regulators and stakeholders

understand the benefits of adopting RegTech which can be further expanded to the MENA region. The findings of this study may lead the way to determine many more drivers due to the dynamic nature of RegTech and help regulators and stakeholders set up many more RegTech firms in the UAE.

## 6 Conclusions and Future Scope

This paper presents the AHP analysis using empirical data and identifies several specific determinants of RegTech which are notably effective in encouraging RegTech services in the UAE. In the report present, the driving factors for RegTech are identified through various research papers and are analyzed through AHP which does a pairwise comparison of each determinant. A survey was taken to establish a pairwise comparison of each of the determinants to develop the AHP analysis. The results of the AHP process imply the weights of each key indicator which contribute to why companies must adopt RegTech. The study finds that providing customer-centric experience as the largest weighted determinant at 40.8%. Furthermore, ensuring data is secure is ranked second from AHP analysis. Industries and financial institutions must ensure that their data is secure to prevent unauthorized access. RegTech is a currently growing field and has a vast potential to reach out too many markets and provide its service [5]. Many UAE banks and stakeholders can benefit from utilizing RegTech services to provide a safe and seamless user experience to its customers. The AHP technique presented in this paper is qualitative in nature and based on one's perception of a given idea. The results cannot be generalized to all companies present in the UAE. The research was conducted only from a sample of 75 participants without involving many experts' opinions. Since RegTech changes dynamically due to its vast potential and currently developing field, the data obtained may not hold true a few years down the line [8]. Further studies may consider a greater number of participants and take into consideration more expert opinions on RegTech to better understand and analyze the priorities of the obtained determinants. The research was conducted keeping in mind the financial industries. The results may vary if the same study is performed in a different market or in a different geographic location. As the United Arab Emirates is not a very prominent RegTech hub, further studies may consider a generic study that involves all markets over a vast geographical area that may include the entire MENA region or the European region where RegTech is very prominent.

## References

1. Anagnostopoulos, I.: Fintech and RegTech: Impact on regulators and banks. *Journal of Economics and Business*, *Journal of Economics and Business* 100, 7–25 (2018).
2. Arner, D. W., Barberis, J. N., Buckley, R. P.: The emergence of RegTech 2.0: from know your customer to know your data. *Journal of Financial Transformation* 44, 79–86 (2016)

3. Arner, D. W., Barberis, J., Buckley, R. P.: FinTech, RegTech, and the reconceptualization of financial regulation. *Northwestern Journal of International Law and Business* 37(3), 371–392 (2016).
4. Arner, D. W., Barberis, J., Buckley, R. P.: FinTech and RegTech in a Nutshell, and the Future in a Sandbox, *CFA Institute Research Foundation* 3(4), 01–20 (2017)
5. Arner, D. W., Zetsche, D. A., Buckley, R. P., & Barberis, J. N.: FinTech and RegTech: Enabling innovation while preserving financial stability, *Georgetown Journal of International Affairs* 18(3), 47–58 (2017)
6. Authority, F. C.: Call for Input: Supporting the development and adoption of RegTech, *Financial Conduct Authority, London* (2015)
7. Bektenova, G. S.: Are Regtech, Fintech, Blockchain the Future? *KnE Social Sciences* 3(2), 61–67 (2018).
8. Butler, T., O'Brien, L.: Understanding RegTech for digital regulatory compliance. In Lynn T., Mooney J., Rosati P., Cummins M. (eds.) *Disrupting Finance, Palgrave Studies in Digital Business & Enabling Technologies*. pp. 85–102. Palgrave Pivot, Cham (2019).
9. Colaert, V.: RegTech as a response to regulatory expansion in the financial sector. In: *Revue internationale des services financiers/International Journal for Financial Services (RISF)*, pp. 54–75. Larcier, Vienna (2017).
10. English, S., Hammond, S.: *Fintech, RegTech and the Role of Compliance Report 2019*. Thomson Reuters, New York (2019).
11. Gomber, P., Kauffman, R. J., Parker, C., Weber, B. W.: On the Fintech Revolution: Interpreting the Forces of Innovation, Disruption, and Transformation in Financial Services. *Journal of Management Information Systems* 35(1), 220–265 (2018)
12. Goyal, P., Rahman, Z., Kazmi, A. A.: Identification and prioritization of corporate sustainability practices using analytical hierarchy process. *Journal of Modelling in Management* 10(1), 23–49 (2015)
13. Helleiner, E.: Understanding the 2007–2008 global financial crisis: Lessons for scholars of international political economy. *Annual Review of Political Science* 14(1), 67–87 (2011).
14. Khatwani, G., Singh, S. P., Trivedi, A., Chauhan, A.: Fuzzy-TISM: A fuzzy extension of TISM for group decision making. *Global Journal of Flexible Systems Management* 16(1): 97–112 (2015).
15. Markham, J. W.: Regulating the “Too Big to Jail” Financial Institutions. *Brooklyn Law Review* 83(1), 517–529 (2017).
16. Martin, E. W.: Deferred Prosecution Agreements: Too Big to Jail and the Potential of Judicial Oversight Combined with Congressional Legislation. *North Carolina Banking Institute* 18(2), 457–471 (2014).
17. Rao, R. V.: *Decision making in the manufacturing environment: using graph theory and fuzzy multiple attribute decision making methods*. 1st edn. Springer Science & Business Media, London (2007).
18. Shahrokhi, M.: The Global Financial Crises of 2007–2010 and the future of capitalism. *Global Finance Journal* 22(2), 193–210 (2011)
19. Singh, M. K., Kumar, H., Gupta, M. P., Madaan, J.: Analyzing the Determinants Affecting the Industrial Competitiveness of Electronics Manufacturing in India by Using TISM and AHP. *Global Journal of Flexible Systems Management* 19(3): 191–207 (2018).
20. Treleven, P.: Financial regulation of FinTech. *Journal of Financial Perspectives* 3(3), 67–79 (2015).
21. Vives, X.: The impact of FinTech on Banking. *European Economy* 2(1), 97–105 (2017).
22. Wind, Y., Saaty, T. L.: Marketing applications of the analytic hierarchy process. *Management Science* 26(7), 641–658 (1980).

23. Wójcik, D., & Ioannou, S.: COVID-19 and Finance: Market Developments So Far and Potential Impacts on the Financial Sector and Centres. *Tijdschrift voor economische en sociale geografie*, 111(3), 387-400 (2020).

# A Future Prospect for European Collaboration on Advanced Analytics in Economy and Society

Nicola Below<sup>[0000-0003-3388-7958]</sup>, Dirk Burkhardt<sup>[0000-0002-6507-7899]</sup>,  
Lukas Kaupp<sup>[0000-0001-9411-6781]</sup>, Kawa Nazemi<sup>[0000-0002-2907-2740]</sup>

Darmstadt University of Applied Sciences, Human-Computer Interaction and  
Visual Analytics group, Germany  
{nicola.below, dirk.burkhardt, lukas.kaupp, kawa.nazemi}@h-da.de

**Abstract.** Analytical Reasoning by applying machine learning approaches, artificial intelligence, NLP and visualizations allow to get deep insights into the different domains of various stakeholders and enable to solve complex tasks. Thereby the tasks are very heterogenous and subject of investigation in the different areas of application. These tasks or challenges should be defined by the stakeholders themselves and lead through a deep investigation to advanced analytical approaches. We therefore set up a strategic alliance of research, enterprises and societal organization with the goal of a strong collaboration to identify in a first step these challenges and workout technological solutions for each application scenario. We give in this paper a first draft of current challenges and technological advancements. The main contribution of this paper is next to an accurate description of the current challenges in the analytics domain, also the description of an agenda how these challenges can be solved. Furthermore, a process is explained, how the strategic alliance should act and organize their work to realize beneficial and useful analytical solutions.

**Keywords:** research collaboration, European network, strategic management, trend analytics, business intelligence

## 1 Introduction

Analytical Reasoning by applying machine learning approaches, artificial intelligence, NLP and visualizations allow to get deep insights into the different domains of various stakeholders and enable to solve complex tasks. An example for such a complex task could be the identification and prediction of new technologies, upcoming technological and methodical trends and innovations in the different domains of application. This is essential for making strategic decisions for economical and societal challenges. The analytical approaches can be supported through trend mining, machine learning, artificial intelligence, visual analytics and simulation.

To face the named aspects, a strategic alliance is necessary that seriously considers a variety of data and focus in particular on huge amount of data, streaming data and unexplored free data from different resources. But even more, the alliance should not

have only the data as basement in mind, it is much more important to follow the entire data processing pipeline. So that also data storing, data processing and data visualization needs to adequately handled for an effective and efficient analysis.

The main goal is to optimize and use technological and methodological innovations to identify required analytical approaches for politics, business, education and research and to develop adequate solutions to the identified challenges. Therefore, an interdisciplinary European network should be initiated that relies on the expertise of the involved partners, who are internationally renowned in their respective fields. From a technological point of view, methods of digitization, artificial intelligence, natural language processing, visual analytics, data analytics and simulation will be optimized and used for different application scenarios. Thereby, the actors in the respective areas define the challenges themselves, whereby data-driven trends will also be identified. The validation of the results is also carried out by the actors of the respective areas. Therefore, a network of interdisciplinary partners has been established who are able to qualitatively define and assess those challenges and solutions. In this paper we describe the major challenges, particularly in Europe, how the network aims to face these challenges and finally how the network actions will be organized.

## **2 Challenges in European Collaboration toward Advanced Analytics in Strategic Management**

The establishment of analytics and analytical solution is a challenge in most countries. This challenge is often aligned to the digitalization challenge, since wide ranges of current business are still analog. But to face nowadays developments, it is necessary to consider enhancements of products and services to have the address the changed needs of customers.

However, the challenges in using and establishing analytics in daily business and processes requires a more precise understanding. In this section, the intention is to outline a variety of challenges that not only cover technological aspects, but also societal and economical challenges.

### **2.1 Technological Challenges**

Recent advances in technologies and especially artificial intelligence (AI) thrive whole new business concepts in digitizing current labor or is the basis of completely new industries. In the Global Risk Report 2019, the WEF identified the “adverse consequences of technological advances” as one of the major risks [1]. These risks can be seen as challenges for the private sector and society. Disruptive technologies such as AI affect the work of the future. Together, with the trend towards fully-automated smart factories driven by industry 4.0 manufacturing [2] will change considerably. So-called “smart manufacturing” with its new interconnected cyber-physical systems is only one of many examples of technological advances that will affect the labor of tomorrow. If society as a whole does not adapt accordingly, a lot of people may be left behind. The technological challenge may then lead to considerable societal and economical risks

[13]. In our current information age, the spread of information and wisdom is tremendous. The widespread global access to information is deemed to accelerate invention of new technologies. The processing and assessment of information is key to success for current and tomorrow's businesses to develop their strategies which will provide new forms of labor for hundreds of workers for the next decade. Assessment of mass information and their presentation in a manner that is accessible, understandable and therefore usable is one of the great technological challenges of our century. Analytics can help in both challenges to access, assess and help to understand such information masses. The first challenge is to help processing, annotating, identifying errors, visualizing dependencies and connections with the aim to accelerate daily work or internal processes. Therefore, analytics has a considerable potential to support workers sticking with time restraints without letting the learning curve [14] become too steep with regard to the adaptation to new technologies. In consequence, it bears the potential to help leaving nobody behind by technological advances. Moreover, analytics supports identifying trends in the mass information that helps to build sustainable business strategies for the company's future well-being and consistency. Analytics in this context contains in particular corporate foresight. An exemplary research question could be: "How can innovations be created and strengthened and future technologies as well as possible scenarios be predicted by the approaches that are currently at the forefront of technology research in order to make strategic and other decisions in a more targeted manner?"

Due to the rise of a variety of machine learning, Artificial Intelligence, data mining and visual analytics methods, it is necessary and important to evaluate the potentials of all kinds of technologies [12], data and approaches for the so called "data markets". The main research questions in this context are:

- Which kind of data or combinations of data are appropriate for strategic decisions to enhance the potentials of enterprises and local authorities?
- Which models (machine learning and artificial intelligence methods) fit best to support the decision-making process for strategic purposes?
- Which simulation techniques are appropriate to predict and simulate future scenarios for both, enterprises and local authorities?
- Which methods can be applied to evaluate the business value of the gained information?
- Which Visual Analytics techniques enable to involve the "human in the loop" and open at least parts of the black-boxed machine-learning methods?
- Which impact can be measured, if the process of learning and predicting is more transparent?
- Which technologies and approaches will enable the "market-creating innovation" to focus on both societal challenges and industrial competitiveness

## 2.2 Societal Challenges

The WEF identified "unemployment or underemployment" as one of the many major global risks in their Global Risk Report 2019 [1]. It is in direct relation to "adverse consequences of technological advances", that we focus on. Through automation of

labor and new technologies such as AI that assist the trend in automation the societal challenge of unemployment or underemployment becomes reality. Automation causes the creation of new higher educated jobs as it is the reason for losses in lower educated jobs. To gain momentum against this trend a parallel development should be fostered. Analytics can help to process and assess new information and technologies in a simplified and therefore consumable manner by its users. Having said this, the technology has a potential to help people in lower educated jobs become qualified and prepared for the requirements coming with the transitioning to new jobs.

The Global Risk Report additionally identified the “spread of infectious diseases” as one of the major risks that we are currently facing [1]. Covid-19 has shown that to solve the problem, we rely heavily on traditional care services, which cannot be digitized. However, a variety of contextual aspects might be digitized to at least support the workers in their jobs. Care service in general can be supported with ambient assisted living technologies so that on the one hand people could do many jobs longer self-confident and therewith the nurses do not need to do the task. The challenge in this regard is not only the development of such novel assisted living technologies, it is also about having technologies that support the identification of beneficial services/technologies. Even today a number of solutions exist that might be helpful in the care service sector, but actually they are not known and identified [7]. Smart solutions can help to find synergies.

In an ever more complex world, which develops at an ever-faster pace, governments are struggling to keep up and to address societal problems in a sound and timely manner. This led to a decline in acceptance of government decisions throughout the democratic world, leading to a rise of citizens and associations who are critically questioning democratic governments to the extent that some groups of people are willing to terminate the social contract on their side. To countermeasure this trend societies place their hopes on e-governance and new forms of policy modelling giving stakeholders, from scientists to business associations over civil society groups down to the single individual possibilities to access information and represent their insights and interests in the decision-making processes. However, current policy making is still a highly analog process and technologies are still barely used or not implemented consequently [8, 9]. In fact, most governments miss a plan how new technologies and solutions can be considered and embedded in the entire policy making process. And finally, even if new technologies which might assist the predictive tasks of a decision-making process (for example simulations), they are still uncommon or poorly developed. Analytics has considerable potential to support the decision-making processes by enhancing the understandability of the information at hand, both on the side of the decision-makers and on the side of the stakeholders and therefore accelerating the process and coming to better and more acceptable decisions. Moreover, analytics might help in identifying upcoming technologies (for example simulations, but also transparency and participation enhancing technologies) which might help in coming to better decisions behind the background of uncertainties in an accelerating world.

### 2.3 Economic Challenges

More and more data are getting digitized and allows new insights into market, technologies, competitors and more [3, 4, 7, 22]. This digitalization trend is particularly important for small and medium sized companies, since a change enables other market players to modernize businesses and can therewith be a high risk for traditional approaches. For enterprises, municipalities and citizens (in terms of consumers), digitalization requires solutions that enable them to identify, understand and apply such new trends. Demand and competition force an increase in market and price pressure. The pressure will enforce the development of even more automation to keep costs short. The challenge to automate everything in every industry to spare costs also accelerates the other challenges (societal and technological) and is an example for the strong interplay between challenges in society, technology and economy.

## 3 Proposed Objectives & Advancements

Analytics is a way to simplify the assessment of complex data and has the potential to enable, both, the private sector and society to assess, understand and process unstructured data to develop new strategies, in areas, such as corporate foresight and the future of employment with regard to their challenges [10]. In a first attempt of a definition “analytics” contains the entire technologies and models for deciding in a more appropriate fashion for diverse tasks in heterogeneous fields of application. This could be corporate foresight, societal challenges, manufacturing etc.

### 3.1 Technological Objectives

#### **Data Sources, Processing and Transformation for Analytics**

The analysis process requires a number of different steps to allow the application of analytical technologies, e.g. artificial intelligence, machine learning and simulation. It is necessary to investigate the entire digital data transformation, e.g. integrate data, extract information from the integrated data and provide effective and interactive analysis tools. Scientific open access data is maybe one resource that the European Commission will probably address in the new Framework as the current suggestions imply [5]. State of the art in analytical technologies should therefore be investigated. Further the use of various data in particular huge amount of data, streaming data and unexplored free data from social networks may provide better analysis.

#### **Data Mining for Enhanced Analysis and Data Insights**

Given data in the web is often incomplete or not well defined. In consequence it is still difficult to identify a given entity in a full-automated and sufficient way. Also, the categorization of data is often not given. To bridge this knowledge gap, data mining techniques are suitable approaches to mine such information with regard to accessible data [15]. It enables the mining of categories, topics or helps to identify entities by consideration of a number of features that are additionally given with the data itself.

### **Simulation, Prediction and Forecasting through Analytics.**

Besides the pre-processing of data, data mining, simulation, foresight and prediction approaches will enable to evaluate the underlying methods and enable considering different approaches of machine learning and artificial intelligence to identify the emerging technologies and trends. It is necessary to evaluate the approaches from the business-perspective [11]. Therefore, in particular approaches from technology and innovation management, strategic management and business analytics and administrations will play an essential role. Here, the diverse methods stemming from diverse scientific fields, for example technology impact assessment and technology related regulatory impact assessment, should be investigated, as well.

### **Visual Analytics to Perform Trend Analytics**

For trend identification as an example for analytics, an encompassing view on the data is essential to “see” upcoming or manifesting trends [22, 4]. The challenge in trend analysis is that the analytical procedures follow no strict procedure. In fact, this means flexible, interactive and mathematical analysis approaches are required to extract and visualize those parts of data that seem promising [3]. Visual analytics enables the coupling of data, (mathematical) models and visualizations for such an advanced analytical environment to facilitate analysts and decision makers the extraction of the required knowledge from the data [10]. Here, the diverse approaches from different fields of sciences should be considered, as well.

### **Visual Analytics for the Future of Work**

Visual analytics helps to view complex information in a simplified manner. Depicting the smart manufacturing example, a lot of technologies are awaiting at the edge, that will disrupt the training, the daily work and the processes within the domain. Trainees have to understand more and more complex cyber-physical systems that are also interconnected and fully-automated. This machinery produces a lot of complex information that is difficult to assess and therewith processes and procedures are difficult to follow and learn from. Besides the training process, the daily work as surveillance and maintenance is also directly affected by this complexity, followed by the internal factory processes. Visual Analytics can help process, annotate, identify errors, visualize dependencies and connections to accelerate training, daily work or the internal factory processes by mitigating complexity and lower the entropy to make systems and processes more accessible. Therefore, Visual Analytics helps trainees and professionals to stick with time restraints without letting the learning curve become too steep with all the new technologies. Consequently, leaving nobody behind by technological advances.

The strategic alliance owns a variety of technologies and expertise in the named fields, however, many of these technologies focus on only a few aspects, yet. Within the action these technologies and expertise should be merged to advance the power and effectiveness and enable a broader usage. The technological outcome will finally lead to new scientific solutions that are fruitful for a wide number of societal and economical use cases, e.g. “visual analytics for policy modelling”, “visual analytics for medical

purposes", "visual analytics for smart businesses" or - to frame it more broadly - "visual analytics along specific use cases".

### 3.2 Societal Impacts

The general topic of analytics is a multidisciplinary area with methods, models and technologies from computer science, mathematics, economics, law, social sciences and more.

**Education** is the key to withstand the losses in jobs through automation in e.g. smart manufacturing. Analysis helps to assess information simplified and provide suitable information in time in the right proportion. Consequently, the mitigation in complexity helps in education by accelerating training and further education [16]. This enables a faster integration of new technologies.

Technological advances are able to transform **public and societal services** such as health care [17]. Enabling the support of elderly to stay at home for a longer duration or enabling a new level of care-taking. The potential of these new technologies may be hidden in the available mass information. Analytics can cover these traces of information and enable the judgment by the rulers, that insure a faster integration of new technologies in the long run.

Analysis comprises both technology impact assessment and regulatory impact assessment, as they are a major concern of **governmental administration**. The first aims at anticipating possible positive or negative impacts of an analytical result on society [18, 19]. Since laws are an expression of values of a specific society, a concretization of the legal requirements helps to set up criteria to allow an evaluation of societal risks and opportunities stemming from a technology trend [20]. The latter aims at anticipating the current and foreseeable legal framework applying to a technology trend to evaluate risks and opportunities for the technology trend itself.

Moreover, analytics may be applied in a single decision-making process to improve insights to the related problems and questions raised with the aim to improve transparency of such processes for decision-makers and stakeholders, as well, by improving understandability of the data. Moreover, given the fact that governmental decisions are flanked strongly with participation processes aiming to achieve acceptable and therefore better decisions, analytics may also play a key role in improving the participation process itself with regard to the stakeholder involvement, by improving the analytic capacities with regard to the given comments and their uptake, their relation to scientific or political documents, for both the decision makers and the public, as well.

Knowledge from management, such as innovation, technology and information management, but also strategic planning and predictions in companies (corporate foresight) should be considered.

### 3.3 Economic Impacts

Right on **time** is key in industry. Either if it is the right business decision in the business strategy or if it is the opportunity to educate the professionals to strengthen their personal journey. Analytics has impacts on both sides of industry, society and businesses.

Analytics helps to unveil formerly unknown technologies and their potential to deduct better revised business decisions in terms of long planning business goals. Therefore, help to build up a sustainable business strategy in shorter time is a huge advantage on the market. While the development of enterprise strategies covers longer periods of time (often month or years), there are a number of scenarios where analysis should help to act within hours or month. The transportation and logistics domain have to react in short times if a duty is cancelled or a partner resigns or cancels a contract. Only with smart analytical technologies it is possible to react on those situations and find alternative duties or immediately calculate cheaper or more efficient routes [21]. Finally, analytics affects the leverage of market and price pressures by visualizing and providing information that can help companies stay ahead of the market.

Furthermore, analytics helps train professionals to build faith in their companies as a source of motivation, which affects the company's economic side. So, **effectiveness** is another major aspect of economics. May it be from perspectives of highly skilled and professional employees or may it be with regard to retrieved insights in certain analysis. Only a high degree of professionalism and quality ensures decisions that support enterprises in being profitable. In perspective of data analytics this means that the entire processing pipeline has to be designed towards high data quality in perspective of completeness, cleansing and data interlinking, but also the visualization has to follow strict requirements in perspective of traceability, accuracy and clarity.

Lastly, the **costs** and **earnings** are essential criteria. To achieve these criteria, the impact of an analytical solution must lay on generating higher profits or lower the costs. From the research perspective the intention to increase the profits is often the smarter more creative way since this often comes along with building or entering new markets with new products. And this means, analytical solutions have to show, which markets this could be and what the major entering criteria will be. Since any market entering has a risk, even more analytical solutions are essential that are able to highlight risks (e.g. possible patent or legal problems in the entered market or country). In sum, there is wide range of scenarios where analytical solutions could help to lower the costs and/or increase incomes, and all of them could help companies to stay productive.

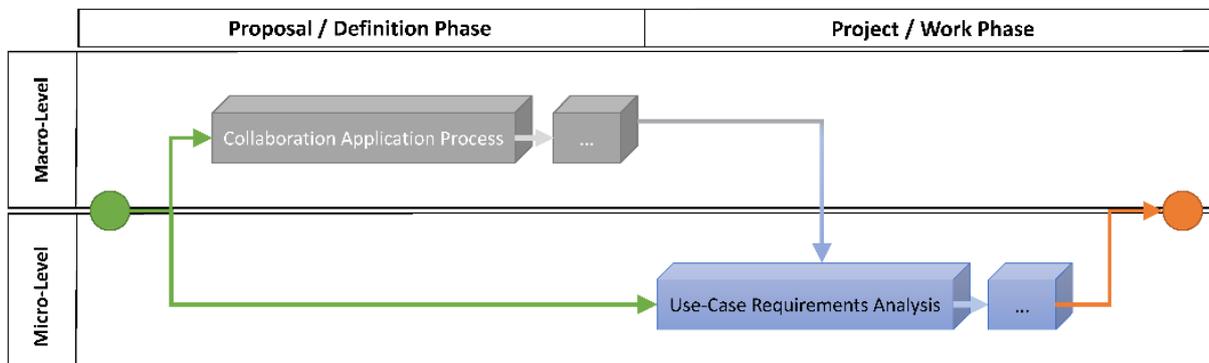
## 4 Macro- and Micro-Level Organization to Enable Advanced Analytics

The organization of the planned analytics collaboration action consists of two main different phases (see Fig. 1).

The first is the proposal and negotiation phase, which also covers the team building that will work together on the specific goals. This first phase can be named as macro-level perspective, since it majorly focuses on the definition and work out the basic goals and objectives.

The second phase is the concrete collaborative work, with the main goal to achieve new insights and technologies. Since the analytics collaboration action is majorly ICT driven, some outcomes will be in form of software prototypes. Due to the heterogeneity of proposed consortium, it is rather challenging to collaborate as far as no common

understanding of use-case requirements are given. This perspective can also be named as micro-level perspective, since it aims to specify concrete use-cases and use-case requirements.



**Fig. 1.** Overview about the entire project pipeline including different project phases and perspective levels.

The figured process model (Fig. 1) consists about a hard-distinguished macro-level and micro-level perspectives, but it is to mention that in praxis it is usually not possible to split them that strict. Due to changes on project objectives and goals over time, e.g. due to partner changes, contextual changes, new research insights etc. some objectives at the macro-level can change with an impact as well as on the micro-level. So even when the micro-level phase already started, there could be changes on the macro-level as well. In sum, the process model in the presented form shows an ideal process model, which in praxis can regularly underlay some modifications. Furthermore, the process can be seen as an iterative process with the aim to refine analytics over many years in various iterations.

#### 4.1 Macro-Level Perspective

The macro-level perspective is particularly at the beginning of such as proposal phase and at the beginning of an accepted analytics collaboration action of relevance. The intention is to define common goals and objectives and teams to collaborate with to achieve these. At this level specific procedures, technologies, algorithms etc. are of interest and the focus lays only on the intended result and main outcome. This means also, that the knowledge/expertise and collaboration is important.

This principle intention is also considered in our defined analytics collaboration action application process, which is divided into four phases (see also Fig. 2).

**Phase 1** aims at promoting discussions on Part B of the present document. The coordinator sends around the present document to inform the partners about the process and gives its proposition on the common understanding (Part B). The common understanding will be updated after comments, additions and requested changes of the proposal. Phase 1 concludes with an online-workshop to discuss and finalize the common understanding, to elect a management board, which will facilitate the drafting of the

technical annex (Phase 2; see Part C) and to give further instructions on the application process (e.g. mode of further operation, answering questions etc.).

**Phase 2** aims at drafting the technical annex (Part C), which forms the most sensitive part of the application. The drafting of the technical annex will be highly collaborative and the single members of the management board will be assigned specific tasks to fulfill. The draft will be finalized by the end of September, to allow for an assessment period (Phase 4).

**Phase 3**, which will be carried out in parallel to phase 2, aims at starting the analytics collaboration action application. This is considered to be the formal side of the application, which gives to the association to assess the conformity of the action and possible overlaps of the action with other actions on basis of a summary. Alterations are still possible in this phase. The coordinator will start the process shortly after and on the basis of the outcomes of phase 1. The partners of the analytics collaboration action will then be electronically invited to provide their contact and institutional information to the analytics collaboration action application. Phase 3 will iteratively inform phase 2 and vice versa.

**Phase 4** aims at giving the partners enough time for a last internal assessment phase before the final submission. Here, a generous time buffer is implemented, if one of the other phases might require a longer period.



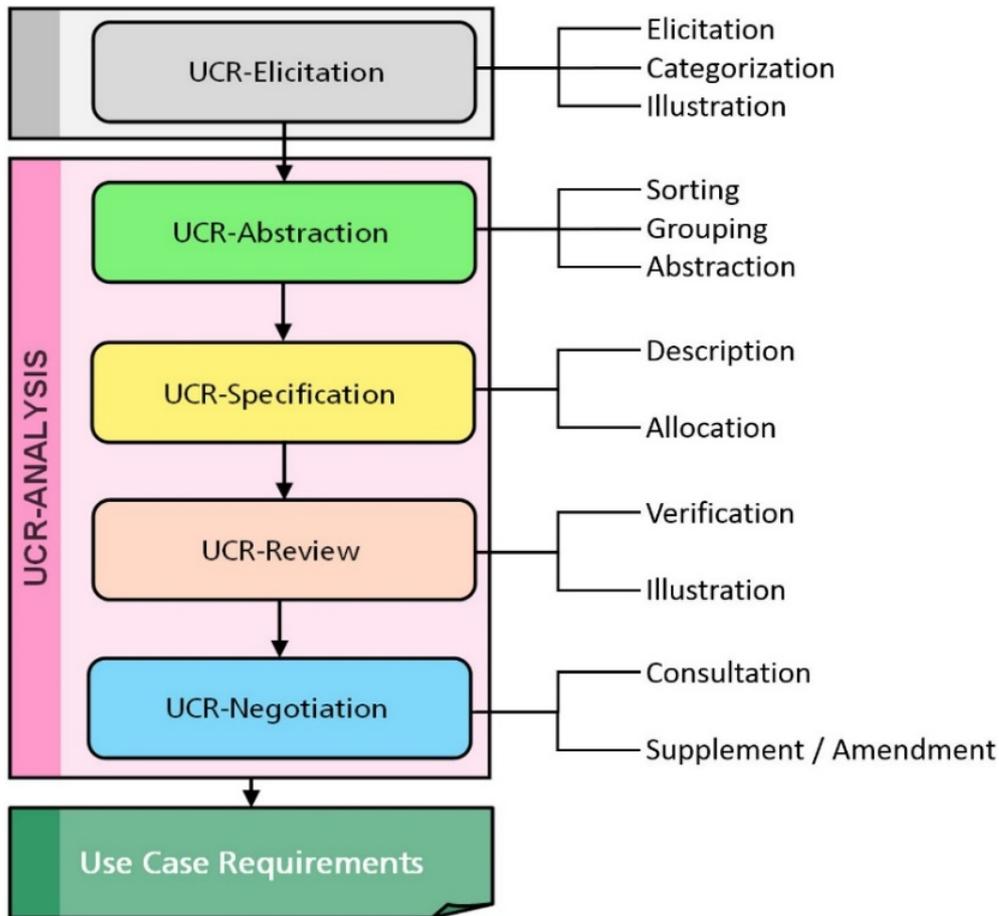
**Fig. 2.** Macro-level process perspective

## 4.2 Micro-Level Perspective

While the macro-level perspective is almost considering major goals and objectives and to define principal aspects, the micro-level perspective is to achieve and a final realization, especially in larger collaborating teams. Since in the work within heterogenous teams do collaborate, different opinions and (technical) understandings are a common challenge that we aim to phase with this micro-level process – also named as Use Case Requirement Analysis Model (UCRA) [6]. The UCRA covers all relevant parts to find a common agreement and at the end a successful solution, but before that use-case requirement analysis is elementary to begin a work with common result understanding. The procedure comprises: domain identification, elicitation (categorization, illustration), abstraction (sorting, grouping), specification (description, allocation), review (verification, illustration) and negotiation (consultation, validation, supplement/amendment) (see Fig. 3).

Overall this model can be seen as an iterative approach. After the first round of the UCR analysis and implementation of discussed features, another round can be initiated to refine and extend the system.

One important fact is that this approach should be performed before major developments in perspective of user related features are in progress.



**Fig. 3.** Hierarchical model of use-case requirements analysis as main part of the micro-level project perspective [6].

## 5 Conclusion

The paper aimed on describing how a strategic network should be founded that aims on optimizing and using technological and methodological innovations to identify analytical approaches in the areas of politics, business, education and research and to develop adequate solutions to the identified challenges. That strategic network of partners should rely on the expertise of the involved partners, who are internationally renowned in their respective fields. From a technological point of view, methods of digitization, artificial intelligence, visual analytics, data analytics and simulation should be considered to optimize and use for different application scenarios. Thereby, the actors have to respect the named challenges, but can precise and complete them further by themselves, whereby data-driven trends will also be identified. A following validation of the results is also carried out by the actors of the respective areas later on.

## Acknowledgment

This work was partially funded by the Federal Ministry of Education and Research under the grant number 13FH020AN9 and was conducted within the research group on Human-Computer Interaction and Visual Analytics (<http://www.vis.h-da.de>). For further information about the strategic alliance see <https://www.vis.h-da.de/network.html>.

## References

1. World Economic Forum, 2019. The Global Risks Report 2019: 14th Edition. [Online]. Available: [http://www3.weforum.org/docs/WEF\\_Global\\_Risks\\_Report\\_2019.pdf](http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf)
2. BMWi, 2019. Leitbild 2030 für Industrie 4 .0: Souveränität, Interoperabilität und Nachhaltigkeit sind die zukünftigen strategischen Handlungsfelder bei Industrie 4 .0. Das Ziel: Digitale Ökosysteme global gestalten. [Online]. Available: [https://www.bmwi.de/Redaktion/DE/Downloads/Monatsbericht/Monatsbericht-Themen/2019-10-leitbild-2030-fuer-industrie-40.pdf?\\_\\_blob=publicationFile&v=6](https://www.bmwi.de/Redaktion/DE/Downloads/Monatsbericht/Monatsbericht-Themen/2019-10-leitbild-2030-fuer-industrie-40.pdf?__blob=publicationFile&v=6)
3. Kawa Nazemi, Dirk Burkhardt, 2019. Visual Analytics for Analyzing Technological Trends from Text. In *Proceedings of 23rd International Conference Information Visualisation (IV)*, pp. 191-200, Springer, Cham. doi:10.1109/IV.2019.00041
4. Kawa Nazemi, Dirk Burkhardt, Lukas Kaupp, Till Dannewald, Matthias Kowald, Egils Ginters, 2020. Visual Analytics in Mobility, Transportation and Logistics. In *Proceedings in ICTE in Transportation and Logistics 2019*, pp. 82–89, Springer International Publishing, Cham. doi:10.1007/978-3-030-39688-6\_12
5. European Commission, 2019. Shaping Europe’s digital future: Open Science. [Online]. Available: <https://ec.europa.eu/digital-single-market/en/open-science>
6. Dirk Burkhardt, Kawa Nazemi, Silvana Tomic, Egils Ginters, 2015. Best-practice Piloting of Integrated Social Media Analysis Solution for E-Participation in Cities. In *Procedia Computer Science: ICTE in regional Development 2015*, Valmiera, Latvia, 77, pp. 11–21. doi:10.1016/j.procs.2015.12.354
7. Reiner Wichert, Beate Mand, 2016. *Ambient Assisted Living*. Springer, Cham. doi:10.1007/978-3-319-52322-4
8. Dirk Burkhardt, Kawa Nazemi, Jörn Kohlhammer, 2014. Policy modeling methodologies. In book *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*, pp. 48–60, IGI Global, Hershey PA, USA. doi:10.4018/978-1-4666-6236-0.ch004
9. Susanne Sonntagbauer, Peter Sonntagbauer, Kawa Nazemi, Dirk Burkhardt, 2014. The FUPOL Policy Lifecycle. In *Book Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*, IGI Global, Hershey PA, USA. doi:10.4018/978-1-4666-6236-0.ch005
10. Pak Chung Wong, J. Thomas, 2004. Visual Analytics. In *IEEE Computer Graphics and Applications*, vol. 24, no. 5, pp. 20-21, IEEE. doi: 10.1109/MCG.2004.39
11. Dace Aizstrauta, Egils Ginters, 2013. Introducing Integrated Acceptance and Sustainability Assessment of Technologies: A Model Based on System Dynamics Simulation. In *Modeling and Simulation in Engineering, Economics, and Management*. MS 2013. *Lecture Notes in Business Information Processing*, pp. 23-24, vol 145. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-38279-6\_3

12. Detcharat Sumrit, Pongpun Anuntavoranich, 2013. Using DEMATEL Method to Analyze the Causal Relations on Technological Innovation Capability Evaluation Factors in Thai Technology-Based Firms. In *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 4 (2).
13. Tomas Hellström, 2003. Systemic innovation and risk: technology assessment and the challenge of responsible innovation. In *Technology in Society*, 25 (3), pp. 369-384, Elsevier. doi: 10.1016/S0160-791X(03)00041-1
14. Alec Levenson, 2018. Using workforce analytics to improve strategy execution. In *Human Resource Management*, 57, pp.685–700, Wiley. doi: 10.1002/hrm.21850
15. Hongyi Peng, Siming Zhu Peng, 2007. Handling of incomplete data sets using ICA and SOM in data mining. In *Neural Computing and Application*, 16, pp. 167–172, Springer. doi: 10.1007/s00521-006-0058-6
16. Eyal Sela, Yesha Sivan, 2009. Enterprise E-Learning Success Factors: An Analysis of Practitioners' Perspective (with a Downturn Addendum). In *Interdisciplinary Journal of E-Learning and Learning Objects*, 5 (1), pp. 335-343. Informing Science Institute.
17. Fatemeh Mariam Zahedi, Nitin Walia, Hemant Jain, 2016. Augmented Virtual Doctor Office: Theory-based Design and Assessment. In *Journal of Management Information Systems*, 33(3), pp. 776-808, Taylor & Francis. doi: 10.1080/07421222.2016.1243952
18. Dirk Burkhardt, Kawa Nazemi, Jörn Kohlhammer, 2014. Policy Modeling Methodologies. In *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*, pp. 48-60. Hershey, PA: IGI Global. doi: 10.4018/978-1-4666-6236-0.ch004
19. Dirk Burkhardt, Kawa Nazemi, Jan Ruben Zilke, Jörn Kohlhammer, Arjan Kuijper, 2014. Fundamental Aspects for E-Government. In *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*, pp. 1-18. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6236-0.ch001
20. Dirk Burkhardt, Kawa Nazemi, 2019. Visual legal analytics – A visual approach to analyze law-conflicts of e-Services for e-Mobility and transportation domain. In *ICTE in Transportation and Logistics 2018*, 149, pp. 515-524, Elsevier. doi: 10.1016/j.procs.2019.01.170
21. Barbara H. Wixom, Bruce Yen, Michael Relich, 2013. Maximizing Value from Business Analytics. *MIS Quarterly Executive*, 12 (2), pp. 111-123, EBSCO.
22. Kawa Nazemi, Dirk Burkhardt, 2018. Visual analytical dashboards for comparative analytical tasks – a case study on mobility and transportation. In *ICTE in Transportation and Logistics 2018*, 149, pp. 138-150, Elsevier. doi: 10.1016/j.procs.2019.01.117

# Millennial's CSR Perception, Social Influences, And Intention to Buy Social Responsible Products: A Conceptual Framework

En-Chi Chang<sup>1</sup>, Stefanie Regier<sup>2</sup> and Ingo Stengel<sup>3</sup>

<sup>1</sup> Allmarketing.com.tw, Taichung, Taiwan  
enchi.chang@gmail.com

<sup>2</sup> Karlsruhe University of Applied Sciences, 76133 Karlsruhe, Germany  
stefanie.regier@hs-karlsruhe.de

<sup>3</sup> Karlsruhe University of Applied Sciences, 76133 Karlsruhe, Germany  
ingo.stengel@hs-karlsruhe.de

**Abstract.** This paper aims to develop a conceptual framework presenting how social influences, including technology-based influence, and millennials' inherent corporate social responsibility (CSR) perception affect millennials' attitudes and purchase intention related to socially responsible (SR) products. Millennials' CSR perception and behaviour and the relevant antecedents are considered in developing the conceptual framework based on the theory of planned behaviour and social conformity. The discussion about subjective norms considers social conformity and millennials' reliance on technology-based media, while the attitude towards SR products consists of cognitive and affective attitudes. The framework and the subsequent empirical study will help companies to evaluate in which CSR initiatives they should get involved and which social influences would be the most effective for delivering CSR-related messages.

**Keywords:** Millennial, CSR, Social Conformity, Theory of Planned Behaviour

## 1 Introduction

Millennials are 'people born between 1986 and 2002' [1][2]. This generation grew up with computers at home [2] and is comfortable with connecting with others, e.g., with their family, friends or even strangers, through various channels, such as cellphones, computers, and mobile devices [1][2]. Millennials are also the first generation growing up with CSR [3]. They believe they should make the world a better place, and companies have the responsibility of pursuing the same endeavor [4][5].

Marketers consider millennials a promising generational segment [6][1]. Market reports and academic research have reported how a company's corporate social responsibility (CSR) initiatives influence the millennial's job-seeking, e.g., [7][8] and brand perception, e.g., [9][10][11]. However, millennial consumer studies still call for an un-

derstanding of the relationship between millennials' characteristics and CSR perception. In addition, they are interested in how this relationship influences millennials' consumption behaviour.

Though market reports mention that the millennial is a CSR conscious generation [3][4][12], it is not clear to what extent CSR motivates and influences these consumers' purchases and which CSR dimension is these consumers' primary concern. Understanding such influences would help companies to evaluate in which CSR initiatives they should get involved and which social influences would be the most effective for delivering CSR-related messages. Therefore, this paper aims to develop a comprehensive framework for an empirical study investigating how CSR perceptions and social influences shape millennials' attitudes and purchase intention of socially responsible (SR) products.

## 2 The Conceptual Framework

According to the theory of planned behaviour (TPB, [11]), subjective norm, attitude towards the behaviour, and perceived behavioral control influence a consumer's behavioral intention and, in turn, actual behaviour. In this paper's context, millennials' social influences, CSR perceptions, and behaviour will be considered in the development of the TPB-based conceptual framework (Figure 1).

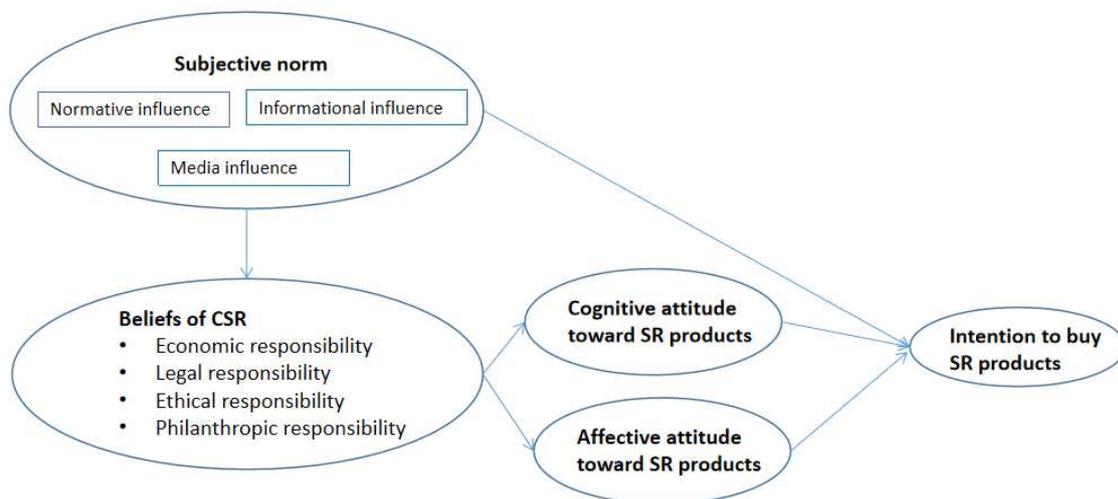


Figure 1. Conceptual Framework

### 2.1 CSR Beliefs And Attitudes towards SR Products

Because millennials grew up with information on cause-related marketing and CSR, such information is internalised as millennials' beliefs and influences how they evaluate companies and products [14][2].

Ben Brik et al. [15] argue that CSR enables a company to build a positive relationship with its customers and other stakeholders. Therefore, a consumer with CSR in mind would consider an SR product positive. Past studies in consumers' CSR perception have considered the CSR sub-dimensions separately, e.g., [16][17] [18][19][20]. In this paper, CSR belief is considered a holistic construct consisting of beliefs about different CSR sub-dimensions, i.e., economic, legal, ethical, and philanthropic CSRs [21], and influencing millennials' evaluation of SR products. Such an evaluation is termed 'customer value' by Papista & Krystallis [19] and can be cognitive and affective [22][19]. The discussion here leads to the following propositions.

**P1.** Millennials' CSR beliefs, including economic, ethical, legal, and philanthropic responsibilities, affect millennials' cognitive attitude towards socially responsible products.

**P2.** Millennials' CSR beliefs, including economic, ethical, legal, and philanthropic responsibilities, affect millennials' affective attitude towards socially responsible products.

## 2.2 CSR Beliefs And Attitudes towards SR Products

When shopping for products or services, millennials often ask opinions from family and friends [23][1] and want to conform with peer groups and social trends [24][25]. Such a tendency can be explained by social conformity theory[26]. Social conformity includes normative conformity and informational conformity. The former is motivated by the concern of gaining others' approval or maintaining social harmony with others. The latter relates to 'the desire to be correct' and the concern of making an accurate and valid judgment [26][27].

In the context of millennials, research [24][28][29] also shows that various media, mainly technology-based/social media, have been a part of millennials' daily life. These media are critical channels for delivering marketing messages, and the information from these channels could even influence millennials' prosocial emotion towards charitable causes [30]. Therefore, social influences, both from their social circles and from online sources, e.g., company websites and blogs, play an essential role in assisting millennials' decision-making [23]. The discussion here leads to the following proposition.

**P3.** Subjective norms, including normative influence, informational influence, and media influence, affect millennials' willingness to purchase responsible products.

In addition, millennials grew up with the information of cause-related marketing and CSR [3]. They have been subject to the influences of CSR information from various sources. Therefore, the subjective norm defined here could also affect millennials' beliefs about CSR. This inference leads to the following proposition.

**P4.** Subjective norms, including normative influence, informational influence, and media influence, affect millennials' CSR beliefs.

### **2.3 Attitudes towards SR Products And Purchase Intention**

Past studies [31][18] have shown that attitudes towards a brand's socially responsible behaviour could influence purchase intention. Millennials' attitudes towards SR products will affect these consumers' willingness to purchase such products [3][32]. Since there are two types of attitudes, namely cognitive and affective, [22][19] and both exert different levels of influence on purchase intention [33], their influences on willingness to purchase should be considered separately. Particularly when millennials consider both the functionality and 'feel good' aspects of a product [23], the effects of both their cognitive and affective attitudes on purchase intention of SR products should be considered.

**P5.** Millennials' cognitive attitudes towards socially responsible products influence the willingness to purchase socially responsible products.

**P6.** Millennials' affective attitudes towards socially responsible products influence the willingness to purchase socially responsible products.

## **3 Conclusion and Potential Contributions**

Millennials are a CSR-conscious generation and concern whether companies play an active role in fulfilling their social responsibility [4][5]. Therefore, integrating CSR in a company's core business is essential for attracting potential millennial job seekers and consumers. However, studies exploring millennials' intrinsic motivation for supporting CSR are still scant.

This study aims to seek a theoretical explanation and empirically investigate how such motivation and social influences affect millennials' intention to purchase socially responsible products. As the first step, this paper establishes a TPB-based framework demonstrating the possible antecedents of millennials' attitudes towards and purchase intention of socially responsible products. The next step would be to survey millennials to examine the relationships proposed in this paper empirically. The data will be analysed through structural equation modeling in order to elucidate and verify the proposed framework.

Theoretically, it is expected that the empirical study will make the following contributions. First, this study provides a modified TPB framework for explaining millennials' disposition for evaluating and purchasing socially responsible products. Second, including both social conformity and the influence of technology-based media, subjective norms in the modified framework will help to examine whether a consumer's tendency towards socially responsible products is due to an individual's pure desire to be

correct or due to external influences such as the pressure of maintaining social harmony and immersion in the media message. Third, unlike past studies often examining separate CSR sub-dimensions, the framework here considers a consumer's belief about CSR, a composite construct consisting of various sub-dimensions. Fourth, unlike the original TPB viewing attitude as a single construct, attitude in this framework consists of cognitive and affective attitudes; such design will reflect the view on attitude in the extant literature.

Managerially, the empirical study extended from this paper will help managers to understand the primary source of social influences affecting millennials' purchase intention of socially responsible products and design more impactful marketing messages targeting millennials. Besides, the potential study results will show which CSR sub-dimension is millennials' primary concern so that marketers could consider the fit with their core businesses and should accentuate in their marketing messages.

Finally, the discussion in this paper is subject to several limitations. First, when discussing the theoretical framework, this paper does not consider different cohorts that might exist in the millennial segment [23]. Second, this paper does not consider the difference between product categories in the discussed relationships [34].

## References

1. PR Newswire: Study refutes myths about millennials and reveals the motivations behind their brand loyalty and buying power. PR Newswire, April 01. (2014)
2. Solomon, M.R.: Consumer behavior: buying, having, and being. 11th edn. Pearson, Upper Saddle River .(2015)
3. Cone Communication: Perceptions, millennials, and CSR: How to engage the new leaders of tomorrow, <http://www.conecomm.com/csr-and-millennials>, last accessed 2020/07/11.
4. Koch, L.: Millennials want brands to be socially responsible, <https://www.emarketer.com/content/millennials-want-brands-to-be-socially-responsible>, last accessed 2020/07/08.
5. Taylor, C.: What makes a CSR message resonate with millennials and Generation Z? <https://www.forbes.com/sites/charlesrtaylor/2019/04/30/what-makes-a-csr-message-resonate-with-millennials-and-generation-z/#9e81cec7fc71>, last accessed 2020/07/08.
6. Barton, .C, Kowslow, L., Beauchamp, C. How millennials are changing the face of marketing forever: The reciprocity principle, [https://www.bcgperspectives.com/content/articles/marketing\\_center\\_consumer\\_customer\\_insight\\_how\\_millennials\\_changing\\_marketing\\_forever/](https://www.bcgperspectives.com/content/articles/marketing_center_consumer_customer_insight_how_millennials_changing_marketing_forever/), last accessed 2020/07/11.
7. Leveson, L. A., Joiner, T.: Exploring corporate social responsibility values of millennial job-seeking students. *Education + Training*, 56(1), 21-34 (2014).
8. McGlone, T., Spain, J. W., McGlone, V.: Corporate social responsibility and the millennials. *Journal of Education for Business*, 86, 195-200 (2011).
9. Eveland, V.B., Crutchfield, T.N. and Rynarzewska, A.I.: Developing a consumer relationship model of corporate social performance", *Journal of Consumer Marketing*, 35(5), 543-554 (2018).

10. Guzmán, F. and Davis, D.: The impact of corporate social responsibility on brand equity: consumer responses to two types of fit, *Journal of Product & Brand Management*, 26 (5), 435-446 (2017).
11. Zayyad, H.M. A., et al.: Corporate social responsibility and patronage intentions: The mediating effect of brand credibility, *Journal of Marketing Communications*, <https://doi.org/10.1080/13527266.2020.1728565>, last accessed 2020/07/11
12. Nielsen: Global consumers are willing to put their money where their heart is when it comes to goods and services from companies committed to social responsibility, <http://www.nielsen.com/us/en/press-room/2014/global-consumers-are-willing-to-put-their-money-where-their-heart-is.html>, last accessed 2020/07/08.
13. Ajzen, I.: The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50 (2), 179–211(1991)
14. Chong, K.: Millennials and the rising demand for corporate social responsibility, <https://cmr.berkeley.edu/blog/2017/1/millennials-and-csr/>, last accessed 2020/07/11.
15. Ben Brik, A., Rettab, B., Mellahi, K.: Market orientation, corporate social responsibility, and business performance. *Journal of Business Ethics*, 99, 307-324 (2011).
16. Choi, S., Ng, A.: Environmental and economic dimensions of sustainability and price effects on consumer responses, *Journal of Business Ethics*, 104, 269-282 (2011)
17. de Leeuw, Valois, P., Ajzen & Schmidt, P.: Using the theory of planned behavior to identify key beliefs underlying pro-environmental behavior in high-school students: Implications for educational interventions, *Journal of Environmental Psychology*, 42, 128-138 (2015)
18. Mohr, L. A., Webb, D. J.: The effects of corporate social responsibility and price on consumer responses. *The Journal of Consumer Affairs*, 39(1), 121-146 (2005)
19. Papista, E., Kystallis, A.: Investigating the types of value and cost of green brands: Proposition of a conceptual framework. *Journal of Business Ethics*, 115, 75-92 (2013).
20. Pomarici, E. and Vecchio, R.: 2014 Millennial generation attitudes to sustainable wine: An exploratory study on Italian consumers, *Journal of Cleaner Production*, 66, 537-545
21. Carroll, A. B.: Managing ethically with global stakeholders: A present and future challenge. *Academy of Management Executive*, 18(2), 114-120 (2004)
22. Bagozzi, R. P., Burnkrant, R. E.: Attitude measurement and behavior change: A reconsideration of attitude organization and its relationship to behavior. In Wilkie, W. L. (eds.) *NA - Advances in Consumer Research*, Vol. 6, 295-302. Association for Consumer Research, Ann Arbor, MI (1979).
23. Fromm, J., Garton, C. *Marketing to millennials: Reach the largest and most influential generation of consumers ever.* AMACOM Books, New York(2013).
24. Circella, G., Tiedeman, K., Handy, S., Alemi, F., Mokhtarian, P.: What affects millennials' mobility? Part I: Investigating the environmental concerns, lifestyles, mobility-related attitudes and adoption of technology of young adults in California, <https://ncst.ucdavis.edu/research-product/what-affects-millennials-mobility-part-i> last accessed 2020/07/08
25. Noble, S. M., Haytko, D. L., Philips, J.: What drives college-age generation Y consumers? *Journal of Business Research*, 62, 617- 628 (2009).
26. Cialdini, R. B., Trost, M. R: Social influence: Social norms, conformity and compliance. In Gilbert, D. T., Fiske, S. T., Lindzey, G. (ed). *The handbook of social psychology*, Vols. 1 and 2 (4th ed.). , McGraw-Hill, New York, N.US: 151-192 (1998).
27. Myers, D., Abell, J., Kolstad, A., Sani, F.: *Social Psychology, European Edition.* McGraw-Hill Higher Education, London (2010)
28. Moore, M.: Interactive media usage among millennial consumers. *Journal of Consumer Marketing*, 29(6), 436-444 (2012)

29. Smith, B. G., Gallicano, T. D.: Terms of engagement: Analyzing public engagement with organizations through social media. *Computers in Human Behavior*, 53, 82-90 (2015)
30. Paulin, M., Ferguson, R. J., Schattke, K., Jost, N.: Millennials, social media, prosocial emotions, and charitable causes: The paradox of gender differences. *Journal of Nonprofit & Public Sector Marketing*, 26(4), 335-353 (2014).
31. Klein, J., Dawar, N.: Corporate social responsibility and consumers' attributions and brand evaluations in a product-harm crisis. *International Journal of Research in Marketing*, 21, 203-217.(2004).
32. Lerro, M., Raimondo, M., Stanco, M., Nazzaro, C., Marotta, G.: Cause related marketing among millennial consumers: The role of trust and loyalty in the food industry, sustainability, 11, 535-551 (2019).
33. Trafimow, D., Sheeran, P.: Some Tests of the Distinction between Cognitive and Affective Beliefs, *Journal of Experimental Social Psychology*, 34 (4), 378-397 (1998).
34. Brown, T. J., Dacin, P. A.: The company and the product: Corporate associations and consumer product responses. *Journal of Marketing*, 61(1), 68-84 (1997).

# Award Winners

## Best Paper Award

Application of statistical functions for microstructure characterization and determination of elastic properties of ceramic foam

Vinit Vijay Deshpande, Romana Piat, Kay Andre Weidenmann

## Presentation Award A

A Close Look on the Corona Impact on Surveillance Radar Channel Loads

Roman Raekow, Michael Kuhn, Bernd-Ludwig Wenning

## Presentation Award B

A new metric for assessing the performance of 2D Lidar SLAMs

Bingxin Zi, Haiying Wang, Jose Santos, Huiru Zheng

CERC 2020  
Collaborative European  
Research Conference  
Belfast, UK  
10 - 11 September 2020  
<https://www.cerc-conf.eu>

