DigitalExposome: A Multimodal Sensor Fusion Approach to Study the Impact of the Environment on Momentary Mental Wellbeing

Thomas William Johnson

N0502790

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

2023

Abstract

In recent years, the field of environmental sciences has gained considerable attention, driven by increases in global population and rapid urbanisation. The issues have been widely recognised, as has the need for solutions to address it. Previous work to explore the impact relationship has shown poor air quality is harmful not only for health, mental health, and wellbeing but also in recent years as serious as death with the first landmark case of 'air pollution' as a cause of death.

DigitalExposome, a novel conceptual framework is introduced to quantify the impact of environment and mental wellbeing. The investigation uses real-time air quality with the approach of making inferences based on an individual's personal characteristics, behaviour and momentary wellbeing within urban spaces. Using a multimodal sensor-fusion approach in this work with the purpose of utilising miniaturised sensing and smartphone technologies aims to acquire environmental, human on-body physiological and mental wellbeing data, specifically labelled at the point of collection. This has entailed the creation of an affordable, sensor-based environmental monitoring station incorporating Internet of Things (IoT) technologies.

To address this, a practical approach is explored of three stages to unravel and understand the impact of the environment on wellbeing. Firstly, to observe a more human-based personalised approach, the use of trajectories were studied alongside the addition of semantics to collect environmental air quality and on-body physiological data. As a result, semantic-enriched trajectories combined with episodes supports the limitation to quantifying the impact at the point of exposure. Secondly, a study involving 40 participants in the real-world

is conducted in a novel multimodal sensor fusion approach involving real-time data collection using self-labelled wellbeing, air quality characteristics and on-body physiological data. The study extends previous literature by quantifying multiple sensors and self-labelled wellbeing using a more digital approach through low-cost, affordable sensors and mobile technology. The aggregated approach supported a higher accuracy level and produces a more comprehensive relationship impact between the environment, human physiology, behaviour and wellbeing.

Thirdly, this work explores data analysis used to quantify the impact between air quality factors and wellbeing. To observe variable importance, statistical approaches such as Principle Component Analysis and Multiple Variant Regression, results in Particulate Matter and Nitrogen Dioxide having considerable negative impact to human wellbeing. Various models such as Dynamic Time Warping (DTW), Deep Belief Network (DBN) and Convolutional Neural Networks (CNN) have created new opportunities for real-world inference of mental wellbeing using environmental and on-body physiological sensor data. A personalised approach using DTW is proposed as a way to observe changes in wellbeing at a personal human-interaction level which in this work demonstrates a high level of accuracy achieving an F1-Score of 0.88 using a DTW network classifying on a 5-point wellbeing scale. To leverage the concept in quantifying an individual's exposure to the environment using technology combined with artificial intelligence (AI) detailed in this thesis gains a deeper understanding into the negative impact air quality exposures can have towards mental wellbeing.

This thesis offers the first attempt towards assessing the relationship of air quality and mental wellbeing incorporating innovative methods of digital technology and artificial intelligence for the first time. This work has the potential to shed light on how individuals breathe, feel and interact with their environment in different surroundings.

Acknowledgements

The work described and undertook in this thesis was carried out at Nottingham Trent University between October 2019 and July 2023, whilst I was working as a part-time doctoral researcher and Academic Associate.

As a result, I would like to thank the following individuals who have made an impact personally and academically to this work:

Firstly, I would like to share sincere gratitude and thanks to my director of studies, Prof. Eiman Kanjo for her unequivocal support of guiding me through this PhD study. Your consistent motivation, enthusiasm and opportunities that you have given will never be forgotten. For the creative conversations and ideas you have always had to offer and the many publications leading to this thesis. In addition, I would like to extend a thanks to the team of co-supervisors I have had at different stages of my PhD journey including Prof. Saeid Sanei and Dr. Omprakash Kaiwartya. I am grateful for your support, encouragement and guidance in completing this study.

Working towards this PhD has been a significant milestone in my life, and I owe much of my success to the unconditional support and love of my family, which is something that I shall always remember and be very grateful for. My thanks and appreciation to you all for being part of this journey and making this thesis possible.

I would also like to acknowledge the support of my loyal friends for helping me get this far, particularly those past and present of the Smart Sensing Lab. Furthermore, I would like to dedicate this thesis to those individuals who ignited my passion for Computer Science at a young age. Their unwavering enthusiasm and dedication played a pivotal role in shaping this journey, and I owe much of my current position and achievements to their inspiration.

Finally, I would like to express my heartfelt appreciation to the numerous friends and colleagues within the Department of Computer Science. Nottingham Trent University holds a special place in my heart as a chapter in my life that I will always remember with fondness, having completed my BSc. (Hons) and MSc in 2016 and 2017 respectively. These academic achievements have not only shaped my knowledge and skills but have also enriched my personal growth and left lasting memories. Specifically, I want to thank Dr. Jo Hartley for her continuous support and guidance throughout this study and whilst in the role of Academic Associate.

This work was supported by the Department of Computer Science whilst in the role of Academic Associate for the duration of (2019-2022).

Thomas William Johnson
2023

Contents

Abstract i

Acknowledgments iii

Contents vi

List of Figures xi

List of Tables xvi

1 Introduction 1

1.1 Background and Motivation 1
1.2 Research Gap 5
1.3 Research Questions 6
1.4 Aim and Objectives 7
1.5 Major Contributions 9
1.6 Publications 11
1.6.1 Magazines 12
1.6.2 Presentations 12
1.7 Thesis Outline 13

2 Literature Review 15

2.1 Background 15
2.2 Quantitative Assessment of Exposures in Urban Environments . . 17
2.2.1 Unveiling the Influence of the Environment on Human Health 17
2.2.2 The Exposome Concept 19

2.2.3 Limitations to Quantify the Exposome Concept 22
2.2.4 Location Time-Series Data as Trajectories to Quantify the Personal-interaction Level 23
2.2.4.1 The Opportunity of Semantic Enriched Trajectories 24
2.2.5 Methods to Monitor the Environment 27
2.2.5.1 Particulate Matter 30
2.2.5.2 Noise 30
2.2.5.3 Gas Sensors 31
2.2.5.4 Nitrogen Dioxide 31
2.2.5.5 Ammonia 31
2.2.5.6 Outdoor Temperature 32
2.2.5.7 Weather Monitoring 32
2.3 The Models of Affective States 33
2.3.1 Traditional Approaches to Quantify Affective State 35
2.4 Monitoring Affective State 37
2.4.1 Physiological Sensor Measurement 37
2.4.1.1 Heart-Rate 40
2.4.1.2 ElectroDermal Activity 40
2.4.1.3 Blood Volume Pulse 40
2.4.1.4 Skin (Body) Temperature 41
2.4.1.5 Motion 41
2.5 Deep Learning Networks 42
2.5.1 1 Dimensional Convolutional Neural Network 42
2.5.2 Long Short-term Memory 43
2.5.3 Deep Belief Networks 46
2.5.4 Dynamic Time Warping 48
2.6 Reflection and Research Challenges 50
2.6.1 Challenges 50
2.6.1.1 Privacy and Ethics 50
2.6.1.2 Data Collection 50
2.6.1.3 Portability 51
2.6.1.4 Battery Life 51
2.6.2 Current Gaps and Research Opportunity 52

3 DigitalExposome: Wearable and Mobile Sensing Technologies for Transforming Wellbeing **54**
3.1 Introduction . . . 54
3.2 Introduction to the DigitalExposome Conceptual Framework . . . 56
3.2.1 Conceptual Framework . . . 58
3.2.2 Opportunities with DigitalExposome . . . 62
3.2.2.1 Statistical Analysis . . . 62
3.2.2.2 Machine Learning . . . 63
3.3 Data Collection . . . 64
3.3.1 Wellbeing Smartphone Mobile Application . . . 65
3.3.2 Environmental Monitoring System . . . 66
3.3.2.1 Enviro-IoT: A Low-Cost Alternative to Assessing Air Quality Levels . . . 71
3.3.2.2 Deployment . . . 72
3.3.3 Urban Wellbeing: Design of an Ecological Momentary Assessment Tool . . . 73
3.3.3.1 User Testing . . . 79
3.4 Datasets . . . 80
3.4.1 Dataset 1: Semantic-Enriched Trajectories . . . 81
3.4.2 Dataset 2: Quantifying DigitalExposome . . . 84
3.4.3 Dataset 3: Urban Wellbeing: Ecological Wellbeing Assessment . . . 86
3.4.4 Data Storage Protection and Management . . . 87
3.4.5 Data Pre-Processing . . . 88
3.5 Conclusion . . . 89

4 Multimodal Environmental Sensing for Mental Wellbeing Prediction **90**
4.1 Introduction . . . 90
4.2 Semantic Trajectories to assess the Impact at the Point-of-Exposure 92
4.2.1 Experimental Setup . . . 92
4.2.1.1 Identification of Trajectory Episodes . . . 93
4.2.2 Statistical Analysis . . . 96

4.2.3 Classification of Emotion Change using Semantic Trajectory Episodes . . . 97
4.2.4 Dynamic Time Warping . . . 97
4.2.4.1 Classification Results . . . 98
4.3 DigitalExposome Aggregated Study . . . 100
4.3.1 Experimental Setup . . . 100
4.4 Results . . . 101
4.4.1 Statistical Factorial and Variable Importance Analysis . . . 101
4.4.2 Multi-Variant Regression Analysis . . . 106
4.4.2.1 Multiple Regression Model for EDA . . . 106
4.4.2.2 Multiple Regression Model for HR . . . 108
4.5 Classification of Aggregated Mental Wellbeing . . . 109
4.5.1 Deep Belief Network (*DBN*) Analysis . . . 110
4.5.1.1 Classification Results using a DBN . . . 110
4.5.2 Convolutional Neural Network (*CNN*) Analysis . . . 113
4.5.2.1 Classification Results using a 1D-CNN . . . 113
4.6 Discussion . . . 117
4.7 Conclusion . . . 119

5 Exploration of Research Applications and Real-World Interventions 121
5.1 Introduction . . . 121
5.2 Applications . . . 122
5.2.1 Visualising Space and Time: Exploring Spatial and Temporal Patterns . . . 122
5.2.1.1 Heat Maps . . . 123
5.2.1.2 Dirichlet tessellation: Application of Voronoi . . . 125
5.2.2 Urban Wellbeing . . . 130
5.2.3 Enviro-IoT . . . 134
5.3 Conclusion . . . 139

6 Conclusion and Future Work 141
6.1 Conclusion . . . 141

6.2 Summary of Major Thesis Novel Contributions . . . 145
6.2.1 A Novel Framework for Quantifying the relationship between Environment and Mental Wellbeing . . . 145
6.2.2 Real-World Data Collection . . . 146
6.2.3 Deep Learning Classification on MultiModal Sensor Fusion 147
6.2.4 Real-Time Research Application Interventions . . . 147
6.3 Future Directions, Challenges and Recommendations . . . 148

References **150**

List of Figures

1.1 System Architecture showing design, multimodal data collection, classification and research application stages. 4

2.1 The structure of Chapter 2 showing the organisation and their respective dependencies of background overview. 16
2.2 The Exposome Concept Framework depicting the three stages of the approach (including examples) and how each section plays an active part in calculating the health assessment risk for an individual. 19
2.3 A spatio-temporal trajectory in its simplest form depicting an individual point combining x,y as GPS (latitude and longitude) and t as timestamp. 23
2.4 The Russel's Circumplex Model of Affect [190]. 34
2.5 SAM Model of Affect (top) measurement of valence displeasure to pleasant and (bottom) Arousal very active to sleepy. 35
2.6 An example of a 7-item scale utilising the approach of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). 36
2.7 Architecture of a Convolutional Neural Network. 42
2.8 The architecture of an LSTM cell with input vector (X), cell output (h), cell memory (c), input (I) forget (f) and output (o) gates [135]. 44
2.9 Schematic Architecture of a Deep Belief Network: comprised of three-stacked RBMs including an input and output layer. 46
2.10 Differences between calculating the Euclidean Distance and Dynamic Time Warping of a trajectory [220]. 48

3.1 DigitalExposome Concept: The quantification step to unravel the relationship of environment and physiology on mental wellbeing . 57
3.2 The Conceptual Framework for DigitalExposome on the Quantification Step for Mental Wellbeing. 59
3.3 Several data collection types for DigitalExposome. 64
3.4 Screenshot of smartphone wellbeing application made up of the five well-known emojis from crying to very happy. 65
3.5 Electronic circuit of the Enviro-IoT and Enviro-Rucksack depicting a Raspberry pi, power through a battery and several environmental air quality low-cost sensors. 72
3.6 Two locations of the Air Quality Monitors (Enviro-IoT) deployed in Nottingham City Centre and Western Boulevard alongside the DEFRA AURN. 73
3.7 Urban Wellbeing Mobile Application and the six sub-processes. . 74
3.8 Urban Wellbeing system processes: The combined data collection and processing workflow of Urban Wellbeing involving Smartphone application, AURN Air Quality Monitoring Station and Database. 75
3.9 (Left) User view of Urban Wellbeing demonstrating the wellbeing assessment page, (Right) table equivalence of emoji vs scoring comparison. 76
3.10 Image capture page demonstrating the option for users' to understand terms of use accompanies with four example images taken by participants. 77
3.11 Screenshot of Urban Wellbeing assessment in process as air quality and noise sampling is obtained from the environment matching the location of the participant. 78
3.12 Urban Wellbeing mobile application three landing screens to give a general overview of the application before starting the individual assessment. 80
3.13 Experimental setup and tasks associated with this research encompassing a (1) Enviro-IoT Custom-built environmental sensing kit, (2) E4 Empatica to measure physiological responses, (3) Custom-built smartphone mobile application. 81

3.14 Selected route taken by participants demonstrating the journey through a several green spaces (left picture) and a busy, polluted environment (right picture). 82
3.15 The pre-specified route made up of busy, polluted (right) and green spaces (top-left) taken by all participants in the experiment. . . . 84
3.16 Map overlay from all participants that labelled. Each dot represents a location trace using the longitudinal and latitude values recorded by the application. 87

4.1 Wellbeing overlay from one participant's trajectory. Each dot represents a change in wellbeing, tagged with a trajectory (x, y) containing environmental and physiological factors. 93
4.2 An individual trajectory combining semantic features; Environmental, physiological factors depicting each single episode made up of the changes in emotions across the entire trajectory. 95
4.3 Pearson's Correlation Coefficient Matrix made up using the collected environmental, physiological and self-labelled wellbeing. . . 96
4.4 Comparison of classification models used with combined environmental and physiological data to classify wellbeing across episodes. 98
4.5 Comparison of F1-Scores for both environmental and physiological data separated. 99
4.6 Correlation Matrix for Environmental, Physiological and self-reported label. 103
4.7 PCA Analysis - (A) Variance between the different variables, (B) Variance between the different variables without EDA. 104
4.8 Depicts the relationship between the self-reported Participant's wellbeing (Label) and PM2.5. 105
4.9 EDA Regression: Residuals Vs Fitted values curve (A) and Q-Q Plot (B). 107
4.10 HR Regression: Residuals VS Fitted values curve (A) and Q-Q Plot (B). 109
4.11 Comparison of classification models trained using statistical features and features extracted from DBN. 111

4.12 Comparison of Random Forest combined with DBN and CNN when trained using only pollution or physiological data. 112
4.13 Comparison of classification models trained using statistical features and raw data from a CNN. 114
4.14 Comparison of Random Forest combined with CNN when trained using only the environmental or physiological data. 115
4.15 Radar charts showing the accuracy levels of three models (top) CNN, (bottom) CNN-RF, based on ten users data in ad-hoc and fused modes. 116

5.1 Nine heat maps using the collected environment and physiological data applied from the DigitalExposome Concept pre-specified route. 124
5.2 The four step process in constructing a Voronoi diagram using sub-sections as segments. 126
5.3 (left) Voronoi overlay from one participant data. Each polygon represents one location trace tagged with a wellbeing label while collecting the data in a specified route (the map layer from Microsoft Bing), (right collected label data from start to end). . . . 128
5.4 A Correlation Matrix depicting the relationship between the collected environment air quality and physiological factors. 131
5.5 Graph overlay from three participant's data. Each column represents a location trace tagged with the level of noise, Air Quality Index, self-reported wellbeing label and an image of the current environment demonstrating a range of impacting factors within the environment. 132
5.6 Several images obtained from photographs taken during an individual assessment from Urban Wellbeing where self-report wellbeing was labelled as 1 (very negative impact) or 2 (negative impact) as indicated at each image. 133
5.7 Several images obtained from photographs taken during individual assessment from Urban Wellbeing where self-report wellbeing was labelled as 5 (very positive impact) or 4 (positive impact) as indicated at each image. 134

5.8 The Enviro-IoT devices in-situ at the DEFRA owned AURN Site in Nottingham City Centre. 135
5.9 The impact of PM 2.5 on air quality of readings from low-cost sensors (Enviro-IoT) and Industry standard equipment (AURN). . 137
5.10 The impact of PM 10 on air quality of readings from low-cost sensors (Enviro-IoT) and Industry standard equipment (AURN). . 137
5.11 The impact of NO2 on air quality of readings from low-cost sensors (Enviro-IoT) and Industry standard equipment (AURN). 138

List of Tables

2.1 Summary of Popular Environmental Monitoring Stations. 27
2.2 Several assessment technologies and tools to observe changes in physiology. 38

3.1 Four environmental sensing prototypes with embedded electronics and low-cost sensors, developed to support the DigitalExposome process. 67
3.2 The Semantic Trajectory dataset involving factors including: (1) raw trajectory, (2) semantic and (3) mental wellbeing data. . . . 83
3.3 The DigitalExposome dataset involving factors including: (1) environmental, (2) physiological and (3) self-report wellbeing. 85

4.1 Summary of the Descriptive Statistics of the collected Environment and Physiological Data. 102
4.2 Multiple Regression Analysis between EDA and Environmental variables. 106
4.3 Multiple Regression Analysis between HR and Environmental variables. 108

5.1 Statistical Analysis of Environmental Pollution Variables Comparison between AURN and Enviro-IoT. 136

Chapter 1

Introduction

1.1 Background and Motivation

As the global population continues to rapidly grow the places in which we live, work and the environment is impacted greatly by increased urbanisation with busier places for people to meet, things to touch, see, hear and the air we breathe [109], [45]. The World Health Organisation (WHO) have recently found that over 55% of the world's population live within urbanised areas and this is expected to greatly increase to 68% by 2050 [167]. The Office for National Statistics (ONS) conducted research on behalf of the Department for Transport and found that over the last decade emissions into the air have increased from cars by 29% [69], with the study also identifying that in the last five years (from 2018) saw a steeper incline of use. With over 91% of individuals living and working around within these situations, breathing polluted and harmful air which often exceeds the national and international guidelines limit is a reminder of the serious situation still on going for human health [165].

For this reason, the issues surrounding poor air quality is one of the most prevalent risks to public health in the UK and more widely impacts across the entire world [66]. According to Public Health England, a typical human is exposed to poor air quality each day of their lives through either exposure to industry, residential combustion, road transport, agriculture, manufacturing or construction [62].

In recent years there have been vast amounts of work and efforts by studies looking to develop environmental monitoring systems to understand, measure and monitor air quality levels [38], [2]. The Exposome Concept developed in 2005, [237] is an emerging approach that aims to unravel the impact of environmental exposures humans encounter towards their health from birth to death [64]. The results in an impact health assessment report for each individual are typically made up from multiple datasets [163], which despite a useful approach, often with large datasets [134] and time required for analysis it can be cumbersome and incredibly difficult to calculate and understand. Previous research is showing that in order to progress this concept the use of digital technology is required [226], [36]. As a consequence, air quality has significantly contributed to an increased risk of developing health, behavioural-related issues and mental health concerns that between 2017 and 2025 is expected to cost £1.6 billion for the National Health Service (NHS) [62], indicating the need for this to be at the forefront of any approach to understand the impact of the environment.

At the same time, mental health disorders have been widely recognised as a concern to both the economic and societal perspective to individuals [71]. In particular, evidence through research studies are demonstrating that short and long-term exposures to poor air quality in Europe alone is resulting in more people diagnosed with psychotic and mood disorders having an impact of £1.575 trillion spent in healthcare admission costs [152]. On the other hand, previous research so far shows that mental wellbeing is impacted by environmental noise [102], [8] but there is more work to be carried out to fully understand the impact on wider aspects of the environment.

To quantify and ascertain human behaviour through emotional states, the term Affective Computing [218] refers to the development of computing technologies to interpret and measure impact on humans in order to make people's lives better . An essential element of gauging and understanding affective state is the use of technological physiological measurement approaches that can measure the impact through responses of on-body sensors. Previous research on the impact of affective computing highlights the causes of a negative wellbeing: depicting an elevated ElectroDermal Activity (EDA) and decrease of Blood Volume Pulse (BVP) [63] and a reduced heart-rate variability (HRV) [25] as these stimulate an activation for the sympathetic nervous system [194], [94], which is a series of nerves that helps the body in activating its 'flight or fight' responses. Although the many advances in technology and mobile computing are extensive, there remains little research conducted to facilitate the provision of cutting edge monitoring of affective states within the real-world.

Many previous works are continuing to build and explore multimodal sensor fusion approaches which with the ever prevalent growing advances in wearable and smartphone technologies allow for much more data to be obtained from a wider variety of on-board sensors [40]. In a recent study, smartphone technology was used through a mobile application to incorporate the built-in GPS and microphone for noise detection to understand the impact of noise as participants' moved around a city centre environment [100]. Results demonstrated the usefulness of obtaining multiple data variables from the smartphone that could help to infer mental wellbeing states across environments and understand physical behaviours. However, although this is a positive step in being able to quantify the impact there is still little focus on the wider impact of factors in the environment, particularly to air quality such as particulates and gases.

Building on these successes, studies are beginning to utilise on-body physiological wearable devices and smartphone applications in a sensor-fusion approach to unravel the link between environmental noise and wellbeing [102]. The ever increasing capabilities of technology in this field, indicates the need for multimodal sensor fusion approaches to be at the forefront of studies to gain a full perspective of the environment and to utilise the majority, if not all sensors on devices to greater understand the impact.

There have been limited efforts in recent years to use environmental data mostly focused around noise [102], applying Artificial Intelligence (AI) such as deep learning to infer mental wellbeing states in the real-world. The major advances and prevalence of wearable and mobile sensing health options in the area of deep learning provide a new potential of inference of wellbeing whilst at the point-of-exposure. Adapting these advances into this work, the thesis explores through 4 stages of designing, developing and quantifying a new framework to quantify the relationship between the urban environment and mental wellbeing.

Figure 1.1 details the overall system architecture for mental wellbeing through the design of a low-cost environmental air quality monitoring system and self-labelled ecological momentary assessment mobile application to quantify wellbeing at the point of exposure within the environment. This continues to develop a multimodal sensor fusion approach, combining environmental air quality, physiological sensor and wellbeing data for classification in the real-world. Finally, this approach concludes with the many research applications of this work which has been discussed later in this thesis.

Figure 1.1: System Architecture showing design, multimodal data collection, classification and research application stages.

1.2 Research Gap

Despite the interesting proposition of multimodal sensor-fusion approaches in environmental studies, there are limitations in terms of collecting, analysing and inferring real-world data that is fused at the point-of-exposure. So far, little work previously has focused towards gaining an understanding between the relationship of environmental noise and human physiology to infer the impact to emotions [100] and wellbeing [102]. The research study *Urban Mind* [22] utilises mobile technology to obtain multiple data entries fused at the point of exposure to understand the impact of the natural environment to wellbeing, although this is carried out through a subjective survey-based question approach. Alternative work *ExpoApp* [57] demonstrates the potential of employing smartphone technology to assess the exposure in the environment through physical activity and location tracking. In all cases throughout the literature, there remains a lack of consideration involving the wider environmental aspects; such as real-world, real-time obtained air quality and physiological factors for example; (HR, HRV, EDA and BVP) using a sensor fusion approach.

The development of a more low-cost sensing technology presents a novel and exciting opportunity to support the collection of real-world sensor data particularly environmental context within the urban environment. At present, many commercial and industry standard environmental monitoring systems are large in size in a fixed location [68], cumbersome and have high costs associated [161] which limit the scalability to carry out real-world experiments [138], [114], [130]. In particular, among the most popular available to the industry, the AQMesh costs around £3-5,000 [16] and DustTrak [225] can cost in excess of £5,000 depending on requirements for the customer. Furthermore, sensing systems incorporating low-cost sensors may not be adequately reliable to provide real-time air quality levels [209], [208]. Table 2.1, explores a range of environmental monitoring air quality systems that are the most popular in everyday use. As such, there remains a gap in the literature of utilising low-cost sensing technology that provides accurate air quality levels for both portable and fixed sensing stations.

The use of AI is frequently utilised in helping to infer wellbeing states, however many previous work examples are limited in terms classifying the combined multiple data variables through a multimodal sensor-fusion approach. In addition, many previous studies have only focused on a limited number of variables, specifically either physiological or environmental noise (i.e. BVP, EDA, EEG, noise) to train with [94], [207], [102]. Although classifying multimodal sensor-fusion data has been carried out before [124], there is few examples of work involving the combination of environmental factors such as air quality, human physiology and real-world self-labelled mental wellbeing.

1.3 Research Questions

After reviewing the research overview, the thesis primarily aims to explore and unravel the relationship between the environment (focused on air quality), human physiology and momentary mental wellbeing. In particular, this thesis attempts to answer the following questions:

1. How can we quantify the person-environment interaction to help explore an urban environment that promotes a positive wellbeing?

2. How we monitor, fuse and model the relationship of a multimodal approach to understand the impact between urban environment, human physiology and mental wellbeing?

3. Can real-world sensor data be quantified and leveraged to infer wellbeing through personal and community approaches?

4. What are the best approaches for validation and application of DigitalExposome for real-world use cases?

In order to address these questions, the next section explores the aims and objectives of the research performed in this thesis.

1.4 Aim and Objectives

The overarching aim of this research is to explore and investigate the relationship between the urban environment and momentary mental wellbeing using a multimodal sensor fusion approach through wearable, smartphone technology and low-cost sensing devices. In particular, this research focuses towards air quality data in the environment such as noise, particulates and gases to determine the impact of mental wellbeing in a real-life setting. This work goes beyond the conventional 'off-the-shelf' and industry standard devices by developing custom-built devices using low-cost sensors and Internet-of-Things Technology. Initially, environmental monitoring (fixed and portable) systems are designed and physiological on-body devices consisting of miniaturised sensors to aid the collection of real-world labelled sensor data that is required to train classification models. To explore this and in relevance to the research questions and aims for the study, the following objectives have been derived:

1. To conduct extensive research through existing literature, sources and current applications for environmental assessment and momentary mental wellbeing and the impact on health, mental health and wellbeing.
2. Propose and develop a novel conceptual system framework to accurately support in the process of quantifying an individuals' exposure to the environment, utilising digital technologies.
3. To explore how the design of portable and fixed environment monitoring systems, physiological on-body and smartphone technology can be used better to incorporate real-world self-labelled wellbeing.
4. To investigate ways in which multimodal sensor fusion approaches can be used to infer wellbeing with the use of deep learning neural network architectures for feature extraction and classification.
5. To investigate and implement a trajectory modelling approach with the potential of employing semantic enrichment to unravel the impact of mobility and movement-patter behaviour using deep learning neural network architectures.

6. To summarise and explore further the research applications, related to the conceptual framework 'DigitalExposome', mobile sensing applications and environmental monitoring systems.

1.5 Major Contributions

The major contributions of this thesis are summarised below with this work marking an initial attempt at developing a step of quantifying the relationship between the environment, physiology and wellbeing using a sensor fusion approach. The six main contributions of this work are as follows:

1. Propose the design of a novel conceptual framework *DigitalExposome* on quantification of the impact between environment, human physiology, behaviour and momentary mental wellbeing utilising digital technologies in the form of mobile sensing and smartphones.
2. The creation of a novel mobile application by leveraging smartphone technologies for the real-time collection of self-labelled mental wellbeing data from users at the point of exposure, offering a unique and comprehensive approach to obtaining the differences of individual wellbeing in the urban environment.
3. The design and development of a low-cost more affordable solution for real-time air quality monitoring in the urban environments which has demonstrated a high level of accuracy through correlation with industrial standardised equipment.
4. The exploration of human trajectories and how time-series data in the form of a multimodal sensor fusion semantic trajectories enriched with environmental and physiological data whilst extracting self-labelled emotion can be used to measure the impact more directly to the individual.
5. The collection of real-time self-labelled mental wellbeing data used across three datasets obtained in this research to train classification models. In particular:
    (a) Multimodal sensor fusion dataset made up from a real-world study with 43 participants (resulting in 41,037 samples recorded) collecting

in an environment: air quality data, on-body physiological and real-time self-labelled mental wellbeing.

(b) The use of semantics as a way to enrich time-series trajectories with the use of environmental, physiological data and real-time self-labelled wellbeing which is used as episodes to segment the data and understand the impact more at a direct level (point of exposure). A total of 3,953 samples were recorded from 6 users.

(c) Ecological momentary assessment tool using smartphone technologies in combining real-world environmental sensor data to understand the impact of environment on wellbeing through: live sensor data, images and perceived self-labelled wellbeing. In total there were over 50 downloads and 100 assessments carried out.

6. The exploration of deep learning architectures to classify real-world mental wellbeing. A varied range of deep learning models have been explored such as Deep-Belief Networks (DBNs), Convolutional Neural Network (CNN) and Dynamic Time Warping (DTW) have been trained using real-world sensor data to explore the impact of performance.

7. The exploration of real-time of three intervention applications: Spatio-temporal Visualisations (e.g. HeatMaps and Voronoi), Fusing real-world environmental sensor data and mometary mental wellbeing to calculate impact at the point of exposure and the use of low-cost sensors to aid environmental monitoring.

1.6 Publications

During the study of this research, the following manuscripts made up of journals, conferences, workshops were published.

Johnson, T., Kanjo, E. & Woodward, K. DigitalExposome: quantifying impact of urban environment on wellbeing using sensor fusion and deep learning. Computational Urban Science. 3, 14 (2023). https://doi.org/10.1007/s43762-023-00088-9

Johnson T., Kanjo E. (2023) Designing an Interactive Mobile Assessment Tool to Quantify Impact of the Environment on Wellbeing. The 25th International Conference on Human-Computer Interaction 2023 (HCII 2023).

Johnson, T., Kanjo, E. "Urban Wellbeing: A Portable Sensing Approach to Unravel the Link Between Environment and Mental Wellbeing," in IEEE Sensors Letters, vol. 7, no. 3, pp. 1-4, March 2023, Art no. 5500704, doi: 10.1109/LSENS.2023.3243790.

Johnson T., Kanjo E. "Episodes of Change: Emotion Change in Semantic Trajectories of Multimodal Sensor Data," 2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Atlanta, GA, USA, 2023, pp. 178-183, doi: 10.1109/PerComWorkshops56833.2023.10150220.

Johnson, T., Kanjo, E. "Sensor Fusion and The City: Visualisation and Aggregation of Environmental & Wellbeing Data," 2021 IEEE International Smart Cities Conference (ISC2), Manchester, United Kingdom, 2021, pp. 1-4, doi: 10.1109/ISC253183.2021.9562852.

1.6.1 Magazines

Johnson T., Kanjo E., & Woodward, K. (2021) Real-time Environmental Changes Impacts Mental Wellbeing. Air Quality News.

1.6.2 Presentations

Finalist at Stem for Britain 2023 at the Houses of Parliament (Engineering Category).

1.7 Thesis Outline

This thesis is organised in the following chapters:

Chapter 2: presents a literature review focused on environments within a place and the relative impact it can have on health, behaviour, mental health and wellbeing. The Exposome concept as a system to understand the impact of the environment over a human lifetime on health is described and opportunities to develop this further. The chapter also reviews methods to monitor affective states including physiological technological sensors. Finally, considerations are made towards several deep learning architectures employed within this thesis and concludes with a reflection of research challenges.

Chapter 3: presents the new DigitalExposome Concept and conceptual framework as a principle method for quantifying the relationship between the environment, physiology, behaviour and mental wellbeing. The concept attempts to propose a new way of calculating the Exposome through the use of digital technologies in the form of mobile sensing and smartphone technology. The chapter also presents several devices made up of environmental monitoring system (fixed and portable) and mobile sensing technologies that have been custom-built as a viable step in supporting the DigitalExposome concept. These devices have been demonstrated as viable solutions within DigitalExposome through the use of analysis and scholarly output. As a result, this chapter also discusses the three datasets that have been curated as part of this research study.

Chapter 4: of the thesis explores the classification of mental wellbeing using a range of deep learning classifiers based on two of the three datasets discussed in the previous chapter. Two studies in the chapter are presented; one investigates the use of human trajectories and the combination of semantics to contain environmental and physiological data with the use of episodes to understand the impact at a more direct level, (at the point of exposure). The second study establishes comparisons through mathematical statistical tools and deep learning models of the obtained environment and physiological data and the impact to mental wellbeing at an aggregated group level. Across the two studies, deep learning models including deep belief network (DBN), Convolutional Neural Network (CNN) and Dynamic Time Warping (DTW) is trained from the obtained features.

Chapter 5: presents the three potential research applications of this study by exploring mobile ecological assessment tools to quantify wellbeing based on the environment in real-time and at the point of exposure, in addition to low-cost environmental monitoring systems as a reliable solution to long-term monitoring.

Chapter 6: concludes the work with a summary of the contributions made in this research study and presents a direction of potential work for future research based on what has been undertaken within this thesis.

Chapter 2

Literature Review

2.1 Background

The substantial effects resulting from the rapid global expansion in recent years have raised significant concerns, notably regarding impact towards human health [147], behavior [82], mental well-being [102] and death [117]. Urbanisation is the phenomenon caused by the impact of changes in the environment that can be a result of weather conditions, temperature, increase in air quality, growth in crowdedness and high levels of noise, which in turn can all lead to poor health and physical activities [45].

In line with this, the chapter provides a comprehensive review of previous work in the fields, as depicted at Figure 2.1, including quantifying exposure to the environment, in relation to air quality exposures which this thesis has particular focus on. The impact to health and wellbeing is also discussed with methods to monitor these exposures analysed using approaches such as delving into The Exposome Concept and technological sensors. The discussion then leads on to mental wellbeing exploring affective models, followed by methods to monitor wellbeing using physiological wearables and individual on-body sensors. Finally, a series of deep learning architectures and networks to perform classification are reviewed followed by exploring this current gap and identifying the research opportunity.

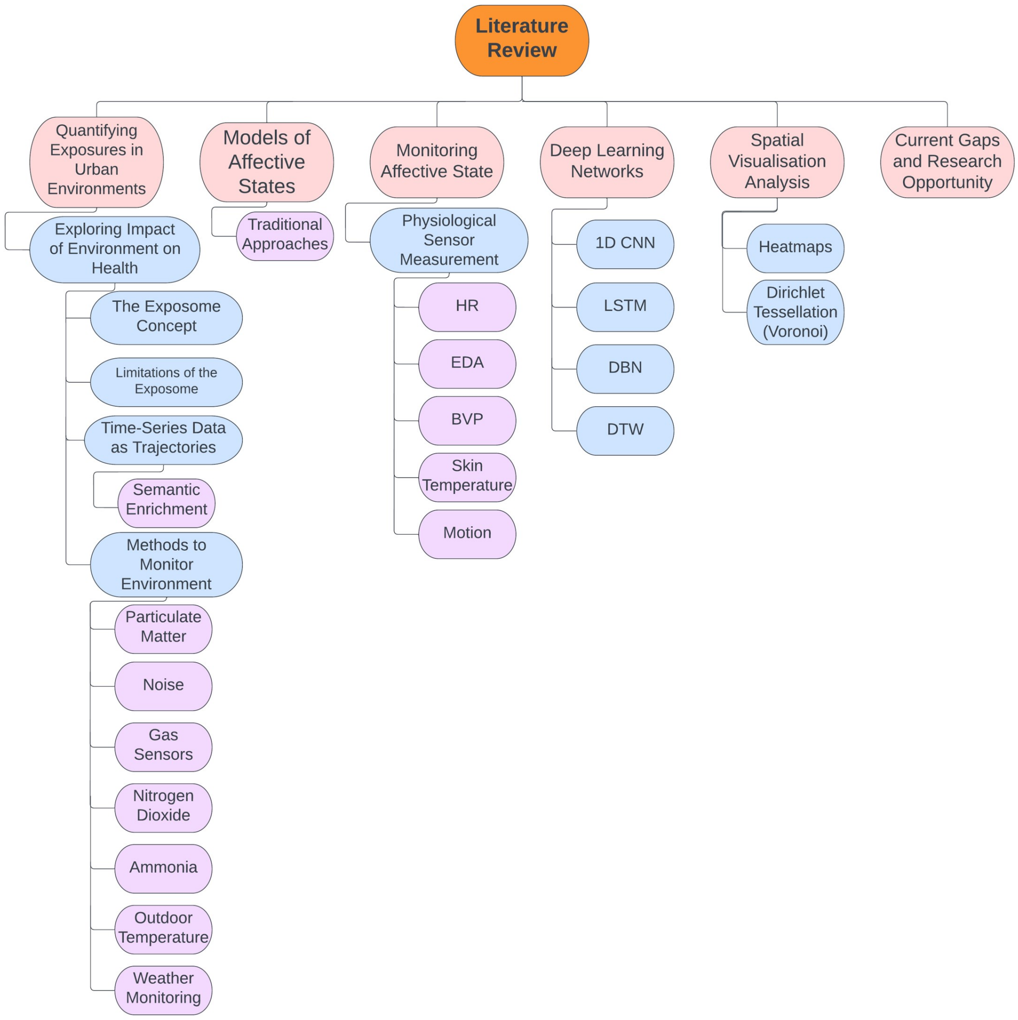


Figure 2.1: The structure of Chapter 2 showing the organisation and their respective dependencies of background overview.

2.2 Quantitative Assessment of Exposures in Urban Environments

Exposure to poor air quality within urban environments is a continual problem that significantly contributes to the rising health and mental wellbeing challenges of our world. Short, long-term and repeated exposures to pollutants such as (noise, air particulates and gases) have all been found to increase the risk of developing health related issues [210], [216] often more serious conditions such as respiratory and cardiovascular diseases [123], [119], [147], behavioural [137], [139], physiological health [102], [82], [28], mental health [186], [24] and more recently death following the first UK citizen to have 'air quality' labelled as the cause of death on the certificate [118].

The World Health Organisation have long studied the relationship between air quality, health and their impact. Research demonstrates that over 91% of people living in urban environments struggle to have satisfactory air quality levels that meet national and international guidelines and the continued use of non-clean fuels and household emissions cause around 4.2 million deaths each year further adding to the concern [97]. As a result, those living in locations whereby air quality levels exceed the guidelines, particularly people in the UK have a higher chance of developing serious health conditions such as higher heart rate, asthma and cardio-cerebrovascular disease [4].

2.2.1 Unveiling the Influence of the Environment on Human Health

ExpoApp [57] used a sensor fusion approach (environmental and on-body factors) to model the short term health impact of high air pollution. By utilising a variety of sensors, the authors were able to calculate the time an individual spent in a specific location whereby particulate matter was at a consistently high concentration. This in turn provides a clear approach to understanding the associated risks high level exposures to the environment can result in. The study analysis revealed that those individuals who didn't have access to green spaces inhaled a higher rate of air pollution. In addition 'Project Helix' [134] studied the total environmental exposure and impact on individuals living in urban environments. As a result increased levels of blood pressure, asthma, allergy related illnesses and behaviour issues are a common concern for those living in urban environments [134], [82], [202].

Personal sensors to measure individual exposures such as air pollution, noise, outdoor temperature, physical activity and blood pressure have been a positive way forward in monitoring due to their ability to collect data continually and in real-time [52] helping to reveal early health conditions [155]. With the high prevalence of combined sensor data streams and the possibility for an individual to continuously wear sensors, the data has the potential to show the exposure an individual encounters as well as predict early health conditions.

Utilising technologies within smartphones such as activity and GPS provides the opportunity to predict an individual's location and the potential exposure that can be encountered. Specifically, a recent investigation showed a positive impact on using sensors and a smartphone to assess an individual's exposure [209]. Similarly, the capabilities that using a smartphone and the built-in microphone has to detect noise as individual's walk around a city centre [100].

There has been a significant interest in the role of 'data driven approaches' in order to obtain reliable, real-time and real-world data [13]. A recent study, utilising a range of on-body sensors and smartphone technology to capture changes within the surrounding environment has demonstrated how an increased level of air pressure can have an impact on body temperature, motion and heart rate [102].

Previously, electroencephalography (EEG) has been used in studies to assess the environmental impact to groups of participants on journey's through urban spaces in order to observe the changes in neural activity [19]. A 2015 study analysed a group of participants that walked around three different urban environments whilst tracking in real-time the EEG of individuals to analyse the emotional experience of the journey. Results found that participants had lower levels of frustration within green spaces and higher levels when entering a busy, polluted urban environment [19]. In addition it was noted that green spaces stimulate a better engagement [96]. Furthermore, walking in urban environments with green spaces has been found to promote relaxation [148].

2.2.2 The Exposome Concept

The Exposome Concept, developed in 2005 [237] is a traditional approach to assess environmental exposure and the impact towards human health within epidemiological studies. This approach has the potential to measure the totality of exposure a human would be subjected to from birth to death [238], [12]. The calculation of the concept in practice has been shown to be highly effective in being able to investigate the impact of environment on human health, mainly due down to the extensive assessment method and data that the concept utilises [85], [213]. Also, increased health outcomes from calculating the exposome has been shown because of the increased knowledge that the concept produces [131].

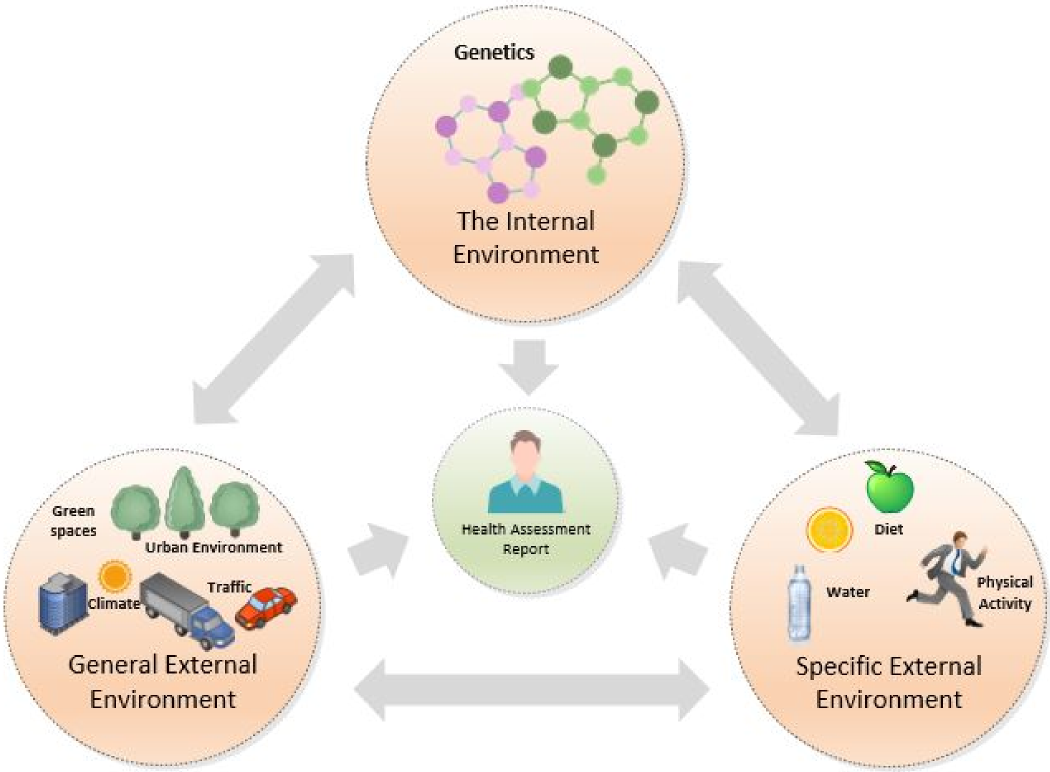


Figure 2.2: The Exposome Concept Framework depicting the three stages of the approach (including examples) and how each section plays an active part in calculating the health assessment risk for an individual.

Figure 2.2, presents the Exposome Concept in its simplest form and highlights the significant amount of data (e.g. Climate, Urban environment, Social, Diet, Physical Activity, Genetics) that is required in order to calculate exposure impact across an individual's lifetime. There are three stages associated with the Exposome; (1) internal, (2) general exposome and (3) specific external before being fused together to develop a health outcome for an individual [231].

Each stage is done independently of the other with the first stage of calculating the exposome is, 'internal', that measures the body's biological response to exposures such as ageing and stress. Next, the second stage of the exposome, 'general exposome' considers the wider impact on our lives and the influences on individuals such as educational background and financial situation. The final stage is 'specific external' which examines effects outside of the body such as radiation, pollution and diet. Once all three stages have been combined, the exact impact of exposure can be calculated [228], [237], [238].

The concept is currently being utilised towards monitoring the environmental impact to human health which in studies has shown the relational impact of air pollution to an individual as discussed in Section 2.2.1. The Project Helix study used the Exposome concept to understand the impact of children (from birth to age 6 and 11) who remained living in urban environments that went on to develop health and development concerns [134]. Out of 5 other countries explored the UK was found to be the highest risk for children developing health-related conditions such as asthma, food allergy concerns, obesity and some showing clear signs of ADHD [134]. A major concern in calculating the exposome, particularly in the case of this study is the amount of data required, whereby the authors relied on questionnaires for the majority of analysed data and existing datasets that were freely available.

Recent developments on Exposome-based studies have shown to identify the exposures early through using the concept, which is leading to improved understanding of diseases and leading to better risk assessments [52]. In addition research has clearly shown that by assessing the Exposome early on can delay a potential life-changing disease or condition [163]. One of the main reason behind this is due to the more extensive assessment methods that the Exposome concept can use in order to measure the impact of exposure [85], [213]. Ultimately, this allows for better interventions to be put into place more easily in order to better protect the individual.

Some studies are starting to utilise the Exposome concept to capture an individual's specific external by incorporating sensors to capture a range of factors within the environment. The calculation reveals that poor levels of air quality can have a detrimental impact to human health, which has been shown to bring an increased risk of developing conditions such as asthma and cardio-cerebrovascular diseases [228], [229], [130]. Although a considerable amount of data was required for the calculation it shows the potential in acquiring technological sensing devices to quicken the process of calculating the Exposome.

Combination of the Exposome concept and personal sensors in previous work has discovered early health conditions diagnosed quicker [52], [163]. A study using a range of sensors including air pollution, noise, outdoor temperature, accelerometer and blood pressure to measure the impact across a 24 hours a day period, with the data showing the potential in gathering substantial amount of exposures and individual can encounter as well as identify early on serious health conditions [155]. In addition, this shows potential in gaining a deeper understanding by collecting a wider range of real-world data more easily to help understand the impact of the environment [226].

Wearable technologies incorporating built-in sensors are increasingly being introduced to monitor an individual's internal Exposome because of their ability to measure continually where ever they go and the wealth of data that can be extracted [227]. The investigation used a variety of sensors, such as: pollution, noise, temperature, particle matter and location. Tracking the individual's location created a map which once the data has been analysed would reveal areas where there was an increased level of exposure to the individual. Although the potential is beneficial, there still remains the gap of ascertaining the true impact value of environmental factors.

2.2.3 Limitations to Quantify the Exposome Concept

Although the Exposome Concept is appealing and has potential to determine the relationship and cause of repeated exposures over a long period of time, in its current form there are many limitations and challenges that researchers face. Since the aim of the Exposome is to assess the totality of the exposures [238], it can take significant time to fully understand the impact. The Centers for Disease Control and Prevention (CDC) [70] identify two research areas that currently need to be explored in order to further the work of the Exposome into everyday life and fully quantify the impact, which includes:

1. Can investment and development of new technology and tools measure the external and internal factors?

2. How can we validate the techniques for response monitoring?

Additionally, the vast quantity of data required for the three sections (General External, Internal External and Specific External) continues to be a cause of concern for quantifying the complete impact and producing the health assessment report either due to the quality of data, size [130] and issues associated with gathering data in the form of agreements, guidelines and ethics [178], [64]. As a result many researchers face concerns in the way of analysing the data on selecting the correct tools and techniques to accurately quantify the relationship of the environment and health [203]. There are other considerable concerns regarding the significant computational power required in order to process the obtained data through statistical and analytical models [251].

Furthermore, accurately assessing each part of the Exposome Concept separately before the calculation is often a limitation due to factors such as equipment cost for gathering data and being able to perform statistical analysis [204]. A recent study starting to utilise technology has shown this to be a limitation on the basis of knowledge and feasibility of acquiring the types of technology [251].

This thorough analysis of the Exposome concept reveals that the there are many opportunities associated with developing the Exposome Concept further particularly in automating data collection to reduce time and the use of technology to gather a wider range of exposure data in the hope to fully quantify the relationship and produce a more accurate health assessment report.

2.2.4 Location Time-Series Data as Trajectories to Quantify the Personal-interaction Level

Location data is ubiquitous in many aspects of our digital lives and has the potential of being able to gain a greater understanding into individual human behaviour and movement patterns. This next section delves deeper into the area at a personal-interaction level to explore spatio-temporal time-series data to investigate how the use of trajectories and semantics can be combined to understand how the environment can impact individuals' directly.

Trajectory modelling is becoming increasingly common in order to explore spatio-temporal patterns in mobility and the movement of multiple objects. A spatial-temporal trajectory can be defined as the observation of a moving object in geographical spaces recorded chronologically in ordered points [252], [116], [244].

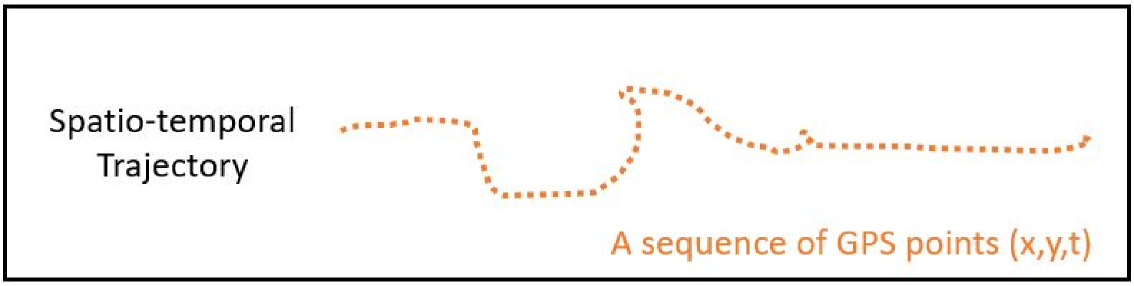


Figure 2.3: A spatio-temporal trajectory in its simplest form depicting an individual point combining x,y as GPS (latitude and longitude) and t as timestamp.

In other words, trajectory data is described as the collection of geo-tagged data points that are ordered by a timestamp [49], [233]. A raw trajectory (T) is a sequence of time-stamped location points depicted at Equation 2.1 [244].

$$T = (P_i \mid i = 1, 2, \cdots, n) \tag{2.1}$$

where Pi is the ith sampling point at spatial positions in geographical space, as depicted at Equation 2.2.

$$P_i = \left(P_{i_{\text{Latitude}}}, P_{\text{Longitude}}, P_{i_{\text{Timestamp}}}\right) \tag{2.2}$$

The observation of a trajectory collected by a smartphone GPS has long been shown to be useful in starting to understand human behaviour [100]. As such, a study in 2014 uses classification to observe trajectories through pedestrian movement and the impact of walking through a busy, polluted urban environment [137]. Geo-visual analytics using visualisations such as how space-time cube (STC) and heat-maps can help to understand how each participant moves around a city environment. However, this research is lacking in exploring changes within the environment and how the participants are moving within could impact their physiology, behaviour and personal characteristics.

2.2.4.1 The Opportunity of Semantic Enriched Trajectories

The concept of Semantics can be defined as the contextualised information that is added to a trajectory enhanced with additional detail which tells us more about the moving object and impacts of mobility directly [185], [7]. Semantics are ultimately used to enrich the content of trajectories beyond latitude, longitude and timestamp information. It can involve situations, such as the weather conditions during a trip, the POIs visited by a tourist (museums, hotels), the activities (shopping, eating). In other words, a primary concern of semantic trajectories is how to present them alongside the additions of the semantics and to facilitate the understanding of behaviours during a single movement. In order to achieve this, novel approaches for modeling trajectories have to go one step further on expressively.

A semantic trajectory involves the re-organisation of a trajectory as depicted at Equation 2.3.

$$T^* = (SubT_i^* \mid i = 1, 2, \cdots, l) \tag{2.3}$$

where:

$$SubT_i^* \tag{2.4}$$

is the ith sub-trajectory associated with particular semantic information.

Many sensor-based studies are utilising semantics in addition to trajectories to help understand the impact towards personal characteristics, daily and human behaviours within indoors and outdoor settings [92], [246]. Previous work showed the opportunities of collecting accelerometer data alongside GPS to create a life-cycle trajectory of daily behaviours [246].

Specific parts of a trajectory or several trajectories can be grouped into a series or several groups which have otherwise been know as an episode [49], [159]. In other words, episodes of a trajectory may follow the scenario such as: period of time (morning, afternoon, night); movement (stopping and moving); category of a particular location region (residence, tourism, commercial, recreation) [49].

The concept 'FrameSTEP' (Framework for annotating Semantic Trajectory Episodes) [158] is a useful approach for segmenting semantic trajectories based on representing generic spatio-temporal episodes. The concept can represent spatio-temporal phenomena at different levels of granularity when focusing directly into the individual episodes. Using FrameSTEP has been shown to give an understanding into large trajectory datasets.

Additionally the concept 'CONSTAnT' [29] is an interesting propositions as it aims to develop a raw trajectory by the addition of semantics to increase the knowledge gained from a journey. The studies use of sub-trajectories (a subset of a trajectory [169]) highlights how a more in-depth understanding on parts of the trajectory can be understood. The results showed how by creating episodes and sub-trajectories of a trajectory tourists could be monitored to understand how they journeyed through a city centre. The work also goes on to suggest how in future work, additional semantics could be added such as the weather and environment monitoring as they moved to understand how this could impact their walk.

Recently, semantic information in the outdoor environment has received increased attention with Electroencephalography (EEG) used to assess the environmental impact to groups of participants on journey's through urban spaces to observe changes in neural activity. A study in 2015 analysed a group of participants that walked around three different urban environments whilst tracking in real-time the EEG to analyse the emotional experience of the journey. Results found that participants had lower levels of frustration and engagement within green spaces and higher levels when entering a busy, polluted urban environment [19]. However, these studies are limited in understanding the true impact of environmental exposure to human physiology, behaviour and in real world environments.

Data mining is an important step in analysing the trajectories obtained from moving objects to discover patterns in the data collected [129]. In addition this method has been shown to help extract important information from a large volume of data [198].

2.2.5 Methods to Monitor the Environment

The next section provides an overview of various environmental monitoring stations commonly employed in industrial settings, as well as individual sensors used to measure specific elements such as Carbon Dioxide and Particulate Matter.

Table 2.1: Summary of Popular Environmental Monitoring Stations.

Sensing System	Specification of Device
AQMesh	Developed in 2012, the AQMesh is one of the smallest devices on the market for air quality monitoring in a fixed location [15]. The device is extremely flexibile with its ability to target as little as 1 pollutant and up to a possible 13 different environmental factors at any time including, gases, particles, noise, wind speed, humidity and pressure [46]. These include: NO, NO2, CO, SO2, H2S, CO2, PM1, PM2.5, PM4, PM10, Noise and temperature [14]. The ability to customise pollutants within the device is useful and a feature that other environmental monitoring systems do not offer. Typically the cost to purchase the device is around £3,000 and £5,000 and an additional monthly maintenance cost which should be taken into account. The AQMesh system is currently being utilised at the Urban Observatory in Newcastle to monitor air quality across the entire city [16]

Continued on next page

Table 2.1: Summary of Popular Environmental Monitoring Stations. (Continued)

Sensing System	Specification of Device
TSI DustTrak Environmental 8543 MCERTS Outdoor Dust & Aerosol Monitor	The DustTrak device is suitable for a more long-term approach to outdoor monitoring of the environment [58], [225]. The device at present allows the measurement of PM Total including PM2.5 and PM1 [222]. Although a certified device of collection for PM, the device is limited as it does not take into account of other environmental pollutants. Each device can cost anything around £5,000 depending on customer requirements for monitoring. Although there are opportunities to hire the device from selected air quality specialists at a reduced cost.
ReliaSENS 19-15	The ReliaSENS device is small, compact and easy to install within the environment to monitor air quality. Each device is able to monitor a range of different pollutants such as CO, CO2, NO, NO2, O3, SO2, H2S, VOC, outdoor temperature, pressure and humidity in real-time [184]. The system also comes with a web interface which allows for the collected air quality to be visualised using cloud storage.
Aeroqual - AQM Series	The AQM series environmental monitoring device offers long-term monitoring with the potential of sensing up to 20 different gaseous and particulate pollutants [2]. The system boasts an interactive dashboard for monitoring of air quality levels and options to visualise the collected data. The whole system and device to purchase is costly and requires a large space for setup which limits the device in research activities.

Continued on next page

Table 2.1: Summary of Popular Environmental Monitoring Stations. (Continued)

Sensing System	Specification of Device
Automatic Urban and Rural Network (AURN)	The AURN is considered to be the UKs largest environmental pollution monitoring system being able to capture pollution types such as NOx, SO2, O3, CO and PM10, PM2.5, weather and temperature conditions [67]. The aim of the network is to ensure that pollution levels are compliant with the agreed national and regional guidelines. The public are given free access to the data of each pollutant level in the area as well as a general air quality index which is a averaged reading of all pollutants calculated together. Although, there are many benefits with the system, they are large and sometimes only one per city meaning pollution levels are only taken into account within that area. The cost with these systems is extremely large with one of the sensors costing around £25,000 [67].

In summary, Table 2.1 demonstrates that there are many limitations when selecting a suitable environmental monitoring system. These include the individual sensor cost as many systems allow a basic device setup which can be expanded on that require additional resources. In addition, many discussed here are fixed systems (such as AQMesh, AURN and ReliaSENS), which cannot be easily portable which reduces the options for using such a system. Considerations as to how the data is collected and then shared is a concern which can limit the analysis carried out and therefore not gain a full understanding of how the environmental is having on human life.

To understand the exposures of pollution within the environment, technological mobile sensing devices and sensors pose the greatest opportunity to collect a range of different environmental factors such as air pollution: made up of noise, particulates and gases. Low-cost sensors are a significant growth in this area with many studies identifying the same results through calibration experiments (tested in the real-world alongside industrial experiments), particularly when testing Ozone and Nitrogen Dioxide sensors [208].

Generally, improved measurement accuracy in sensors has enabled more characteristics within the environment to be collected. The growth in recent years towards custom-built monitoring system with technological sensors has demonstrated the impact of measuring exposures more effectively [211]. As sensors are becoming more reliable, affordable for research they offer exciting opportunities to measure multiple exposures at the same time [114], [47], [160].

The following sub sections summarise the most popular environmental factors that can be monitored within urban environments, in addition to an exploration description.

2.2.5.1 Particulate Matter

As one of the most harmful environmental pollutants, Particulate Matter (PM) [196], is commonly built-in within sensing systems [164]. The structure of the pollutant is extremely small particles and liquid droplets that are produced from acids, chemicals, metals or dust [11]. Within the environment the level of PM can be the result on the combustion of mechanical and industrial processes and vehicle emissions [174]. PM particles are generalized as 1.0, 2.5 and 10 micrometers, with the smaller size considered the most dangerous, as they are invisible [193]. The pollutant is usually measured in terms of micrograms per cubic meter (g/m3) [164]. High levels of PM both daily and overtime have been shown to correlate and cause serious health conditions [119], [123], mental health [179], behavioural issues and death [117]. A landmark study in 2021 demonstrated that high levels in PM2.5 resulted in 'Air Pollution' added to a death certificate for the first time, declaring air pollution and the part it played in causing a child's death.

2.2.5.2 Noise

Sensors to capture sound are often used within environmental sensing systems. Primarily, these sensors work by detecting the overall intensity of sound waves by using an in-built microphone, peak detector and an amplifier [133]. Previous work has demonstrated that the microphones integrated into mobile smartphones are highly effective in directly measuring exposures at the direct level to an individual. This research has revealed that increased noise levels can significantly affect mental well-being [102], [82] and health conditions such as cardiovascular diseases [147].

2.2.5.3 Gas Sensors

Gas sensors are commonly used within environmental monitoring systems as they can be highly sensitive and offer low power consumption, enabling them to be embedded into fixed and portable devices due to their small footprint [78]. Research conducted using both Carbon Dioxide (CO2) and Carbon Monoxide (CO3) sensors highlight significant issues for human health with respiratory and cardiovascular illnesses at the forefront [54]. In addition, more recently the impacts of these gases in the short-term demonstrate physiological changes and behavioural issues [21], although more work is needed to understand the impact of long-term exposures.

2.2.5.4 Nitrogen Dioxide

Nitrogen Dioxide (NO2) is one of the most highly reactive gases as a result from aerosols and combustion processes for fossil fuels within the environment [65]. Previous work on (NO2) has been shown to cause many issues in both short and long-term exposure such as respiratory symptoms [115], stress [188] and to result in cardiovascular illnesses [54]. There is little research showing impact directly to the individual in terms of physiological health and behaviour.

2.2.5.5 Ammonia

Within the environment, Ammonia (NH3) is thought to be one of the most common gases in the atmosphere [83]. Research has shown the impact of Ammonia depends on the level of exposure with low causing irritation to the eyes, nose and throat and high exposure resulting in swelling in the airways and long-term issues in the respiratory system [77]. There is little research in the literature to suggest that ammonia could impact physiological responses.

2.2.5.6 Outdoor Temperature

These types of sensors are able to measure the temperature of the atmosphere where an individual is located. Recent studies have show the impact of low/ high temperature in different seasons which have been shown to be associated with increase health outcomes [248]. In addition, increases in temperature above 21 degrees has been shown to decrease overall emotional mental wellbeing [157].

2.2.5.7 Weather Monitoring

These monitoring systems are made up of different sensors which can include variables such as: air pressure, rainfall gauge, anemometer, sunshine and UV intensity levels. Sensors to measure the humidity aim to measure the quantity of water vapour in the atmosphere. Many studies have used these sensors to measure and monitor the environment [232]. Previous research has shown that air pressure influences physiological responses such as HR [102].

2.3 The Models of Affective States

The term *Affective states* in psychology is often defined as an underlying emotional state [10] and is used as a collective term for encompassing several different states such as emotion, mood, and feelings [1], [23]. Previous physiologists have measured affective state through exploring emotions and stress levels being felt [26].

The concept of *Emotion* as a physiological state has been studied for a long time. Emotions involve a collection of responses triggered from parts of the brain to the body [48], otherwise known as neurophysiological changes that can be associated with feelings, behavioural responses and thoughts [72]. Although generally similar, moods can be interpreted as either a positive or negative valence, which in comparison to Emotion are less specific, intense and unlikely to be provoked by a specific event or stimulus [1]. On the other hand, *Mental Wellbeing* refers to how an individual determines their own potential to cope with normal day life and everyday stresses and that they know their own abilities to contribute to their work and community [166]. Physiologists have often defined the measurement of wellbeing as associated with the intensity of how people feel positive and negative affect [235], [156].

In recent years, there have been a range of theories and different approaches to try to address the challenges of emotion classification. Although there are no common practices, the most popular is Russel's Circumplex Model of Affect [190] as depicted at Figure 2.4. The model represents these emotional states through a spatial model in the form of a circle and measures emotions dimensionally from Arousal to Valence.

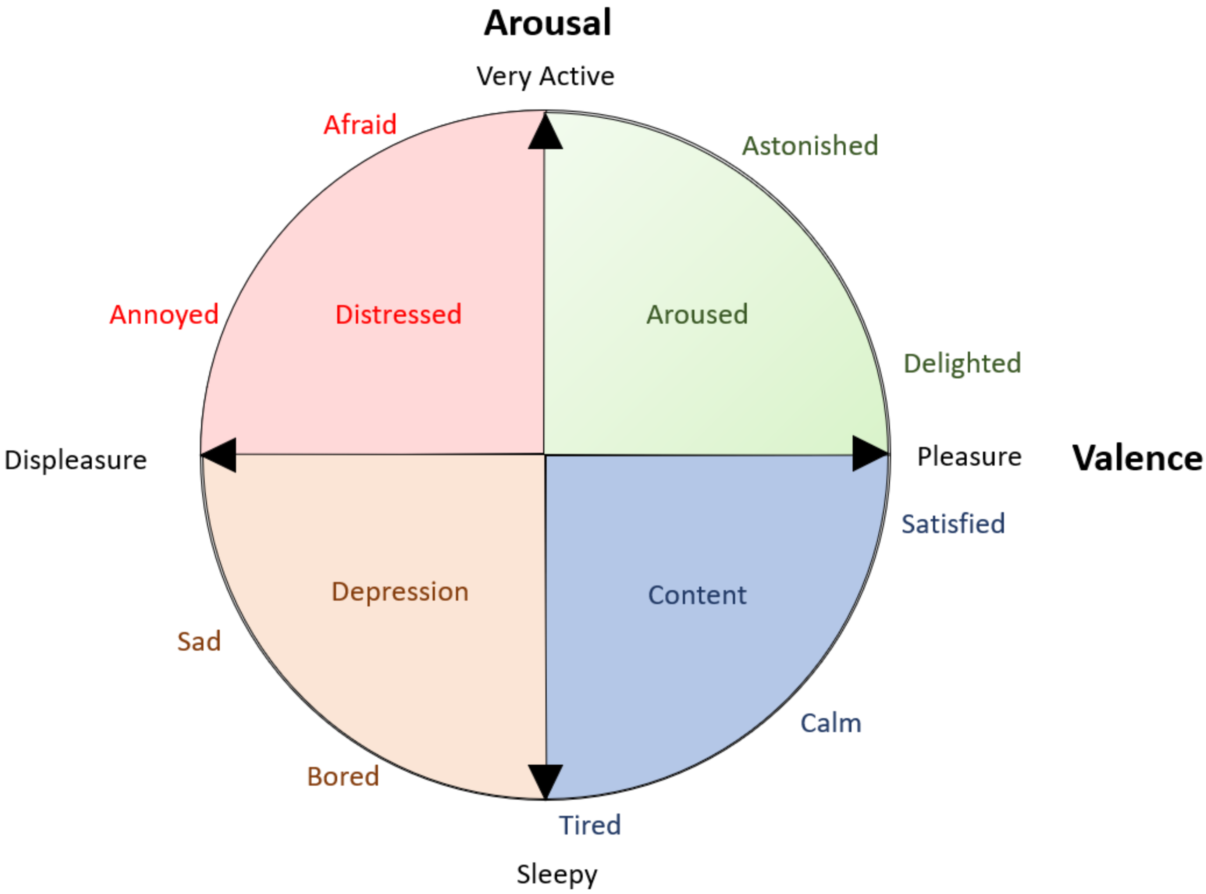


Figure 2.4: The Russel's Circumplex Model of Affect [190].

Additionally, the Ekman Model [60] is made up of 6 individual emotions consisting of Happiness, Sadness, Anger, Surprise, Disgust and Fear which can be expressed through facial expressions. Furthermore, the Self-Assessment Manikin (SAM) [30] is an effective method as depicted at 2.5. The model uses 9 individual pictorial images of humans to measure subjective valence, arousal, and dominance. Using a simpler approach makes this technique easier to apply enabling a quicker assessment and understanding into the impact of affective state. However, consideration must be taken into account when using in a real-world, real-time scenario where the assessment needs to be completed on-the-spot when a change in emotion is felt.

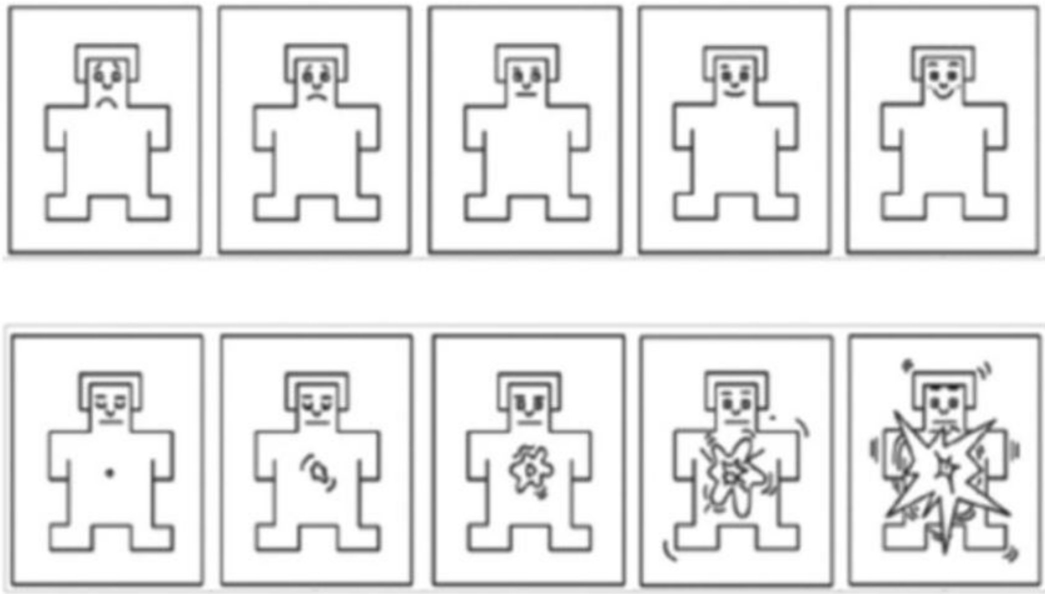


Figure 2.5: SAM Model of Affect (top) measurement of valence displeasure to pleasant and (bottom) Arousal very active to sleepy.

2.3.1 Traditional Approaches to Quantify Affective State

In recent years, there have been a range of traditional approaches to aid in quantifying mental wellbeing which include standardised questionnaires, experience sampling (ESM) [145] and the ecological momentary assessment technique (EMA) [112]. The use of self-report in these approaches has been a primary method of enabling individual's to document their lives (behaviours, thoughts and emotions) which can be used to assess and understand direct triggers [171].

Firstly, the validated scale 'Positive and Negative Affect Schedule (PANAS) has been shown to be an effective questionnaire to measure both mood or emotions of affective states [235]. The scale is made up of 20 items; 10 items measuring positive and 10 items measuring negative affect. In addition, a five-point Likert scale is used to measure the impact from very slightly to extremely likely.

In addition, the Warwick-Edinburgh mental well-being scale (WEMWBS) [162] is an alternative approach to measuring mental wellbeing which consists of 14 positively phrased Likert-style items [221] and sometimes 7 items to reduce the time spent in quantifying mental wellbeing [212] as depicted at Figure 2.6.

Analysing both approaches when quantifying mental wellbeing demonstrates a 0.95 correlation in results indicating that both the 7 and 14 item-approach can be equally as effective [132]. However, self-reporting in this way takes a considerable amount of time since each time a change is encountered an assessment must be completed and scores in the table must be calculated.

Statements	None of the time	Rarely	Some of the time	Often	All of the time
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been clearly thinking	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been able to make up my own mind about things.	1	2	3	4	5

Figure 2.6: An example of a 7-item scale utilising the approach of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS).

Furthermore, an Ecological Momentary Assessment (EMA) involves repeated sampling (at periodic intervals) of individuals' behaviour in real-time [199] which has been shown to minimise bias [200]. Traditionally, an EMA has been completed by pen and paper and also often is a slow process in calculating impact to mental wellbeing [51]. In recent years, smartphone technology has started to be used for EMA, such as the Urban Mind mobile application used to assess the environment through a series of questions [22]. The work starts to highlight how smartphone-based assessment tools can be used to gain a perspective into the natural features within the environment and the impact it can demonstrate towards mental wellbeing. Although useful, this area requires further work with the investigation of real-world sensor data playing a part on the EMA through smartphone applications.

The use of self-reporting across these different approaches have many benefits to accurately capture affective states and mental wellbeing. However, to gain a clear understanding into the data, self-report must be collected for a long period which can be very time consuming [214]. In addition, the application of approaches to real-time experiments has not been suitable since the large resources required. Following exploration of the work to classify emotions, this thesis aims to investigate categorical affective state representation since other studies have shown a guarantee of assigning a physiological reading to a state category [127], [101].

2.4 Monitoring Affective State

2.4.1 Physiological Sensor Measurement

In this section explores the non-invasive technologies and mobile sensors that can be utilised in the real-world and provide the greatest opportunities to capture and assess affective states in real-time. Non-invasive technologies incorporating miniature sensors are increasingly being used in physiological sensing due to their ease of embedding inconspicuously [227], [52], measurement in real-time [243], improved affordability [114] and overall portability [102], [47], [160]. As a result, these technologies have the greatest opportunity to understand mental health issues, providing the mechanisms to collect and physiological changes [101] and behaviour markers of mental well-being [241]. Previous work in the area of machine learning classification and deep learning models to obtain features from the collected sensor data can be used to analyse an individual's affective states [172], [17].

Several physiological assessment technologies readily available in the form of wearables that incorporate miniature sensors have been considered at Table 2.2 which detail the most popular devices available.

Table 2.2: Several assessment technologies and tools to observe changes in physiology.

Image	Device Description
	The E4 Real-Time Empatica [61] is a real-time physiological wearable wristband that enables continuous measurement of the sympathetic nervous systems activity and captures data such as: BVP, Accelerometer, EDA, Body Temperature, HR and HRV. The wearable connects to an app in order to control recordings and data can be easily downloaded as individual variables from the online portal. Although the E4 is currently being phased out for a more up to date version, the device is available to purchase at around £1,690. Previous studies have utilised Empatica's because of its ability to measure and quantify the impact of wearable technology in health and also within clinical trials [42], [217], [113].
	The Apollo Neuro wearable [151], is typically worn either on the wrist, ankle or as a clip on clothing. The device is able to measure heart-rate variability, sleep patterns and activity. Utilised more in clinical studies, the wearable has been shown to help in modulating heart rate variability in order to reduce stress levels and blood pressure [182]. The device costs around £349 and includes a free mobile application as an interactive dashboard to track progress and monitoring levels of variables collected.

Continued on next page

Table 2.2: Several assessment technologies and tools to observe changes in physiology. (Continued)

Image	Device Description
	Movisens [143] offer a range of devices that can be added to a wearable wristband which is able to quantify on-body data such as accelerometer, air pressure, temperature, ECG, DEA, HR and HRV in real life conditions. A limitation to the Movisens devices are that each variable is individual such as temperature and accelerometer have to be purchased separately. The cost for each device can range from £1,000 for an individual (used) device or several £1,000s depending on the variables added for monitoring. Several studies have shown how effective the Movisens devices can be in real-life experiments to quantify health, emotion and clinical studies [91], [146], [122].
	OURA ring [87] is a wearable device that monitors sleep, HR, recognise activity and body temperature in real-time. The device is able to connect to a mobile application that displays the data in different visualisations to the user, also with options to incorporate the data into the respective built-in health applications already installed on a smartphone. For each days' recording, the data is able to be downloaded as a CSV or JSON Format for further analysis. The ring devices roughly cost around £299 [168]. Previous studies have demonstrated the devices ability to quantify HR and HRV and activity quantification [106], [215].

2.4.1.1 Heart-Rate

Heart rate (HR) is measured by the speed of the heartbeat which is referred to as the number of beats per minute (BPM). Typically a normal adult's resting heart rate is between 60 and 90 beats per minute (bpm) [20]. HR sensors due to their small size are often embedded within wearable devices and offer opportunities to accurately assess the autonomic nervous system [103]. This system in previous work has been shown to understand and model positive and negative mood [108]. Heart-Rate Variability can is measured from collecting HR, by understanding the changes in time between heartbeats, which are often referred to as inter-beat intervals (IBI) [195]. Research has shown that a variation in environmental factors such as air pressure and noise have been shown to impact HR [102]. Lower HRV has been shown to be associated with anxiety disorders [76], [37] and result in higher stress [105].

2.4.1.2 ElectroDermal Activity

ElectroDermal Activity (EDA) is measured by assessing the resistance between two-electrodes, typically where sweat glands are located on the human body [34]. These type of sensors are considered the most effective in monitoring the human nervous system. The measurement of EDA changes depending on the level of emotional arousal [74]. The data from collecting EDA can be used towards training affective models in order to classify mental wellbeing as it directly correlates to the sympathetic nervous system which controls the rapid responses to different situations [194]. Several previous studies have utilised EDA sensors to explore the impact of the environment, although so far focusing on noise levels [102].

2.4.1.3 Blood Volume Pulse

Blood Volume Pulse (BVP) is usually measured from a finger or hand and calculated by assessing the changes in volume of blood across vessels and commonly used to indicate physiological arousal in affective computing [110] and to monitor mental wellbeing [253]. The process works by using HR to capture the volume of blood that passes through a particular location within a single heart-beat [126].

Previous studies have shown BVP to be a factor in developing an understanding to mental wellbeing [111].

2.4.1.4 Skin (Body) Temperature

Skin temperature sensors can be used in addition to physiological sensor data to measure mental wellbeing [121]. To measure skin temperature, electrical signals in the form of resistance between two diode terminals is measured where most commonly there is an increase in voltage as it results in a higher temperature [56]. Previous work has attempted to use body temperature sensors to classify mood and emotions achieving 72.3% and 75.0% [128]. However, more recent work focused on multimodal has found increased classification ranging from 0.60 and 0.988 (f1-score) [102], [207], [172], [95]. In many of these studies, skin temperature data has been combined with other physiological data (such as HR and Acceleromter) to increase classification of mental wellbeing.

2.4.1.5 Motion

These sensors can be used to help detect overcrowding in a specific location. For example, working at a 3m distance and a 120 degree detecting angle provides a good detection of motion [79]. Used to capture and measure human movement and physical activity an individual does. Many studies have used accelerometers to monitor the impact movement factors have on health and well-being [9], [24].

2.5 Deep Learning Networks

In recent years the use of Deep Learning has gained significant traction for the opportunities of classifying raw sensor data using classification models such as Convolutional Neural Networks (CNNs) which is explored practically at a later stage in this thesis. There have been a range of classification models previously explored in affective computing to accurately classify physiological data to achieve the best result. Over the following section, a range of deep learning classification networks are explored that demonstrate the potential to classify real-world physiological collected sensor data.

2.5.1 1 Dimensional Convolutional Neural Network

Perhaps one of the most popular algorithms, a Convolutional Neural Network is based on a 'deep feedforward' approach that involves the addition of numerous layers [43]. CNNs are more commonly used with image processing but in recent years have been repurposed and research using 1-dimensional CNNs are found to be effective in modelling time-series data [240]. This is composed of an input, output and several hidden layers which includes the convolutional layer that makes use of a set of learnable filters, polling layers, fully connected and normalisation layers [245]. A 1D CNN forms the same structure as a 2D CNN, although the only difference are the input dimensions and filters. Depicted at Figure 2.7 illustrates the basic blocks of a typical sample CNN configuration.

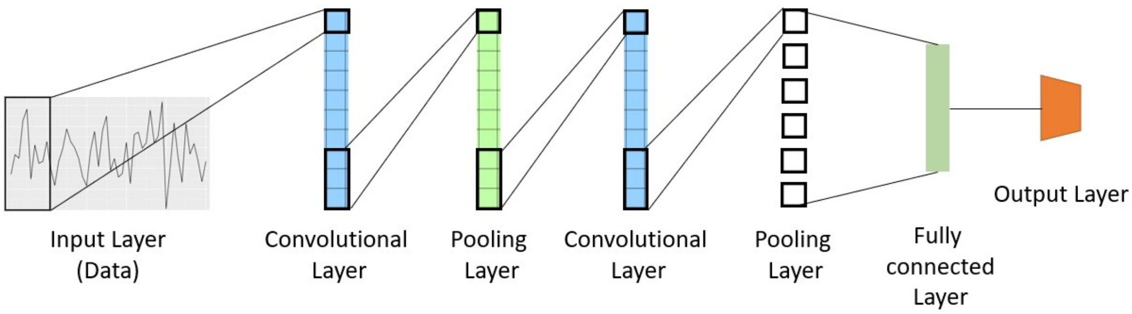


Figure 2.7: Architecture of a Convolutional Neural Network.

More formally, the training input dataset can be represented as: $x = [x_1, x_2, \ldots x_j]$ whereby the number of training samples is j [86]. The dataset is passed through a series of CNN layers that scan for sequences with 1D windows and learn filters. The feed forward process can be further denoted as:

$$p = \sigma\left(w_1 x + b_1\right] \tag{2.5}$$

$$y = \sigma\left(w_2 p + b_2\right) \tag{2.6}$$

Equations 2.5 and 2.6 present the sigmoid otherwise know as the activation function. This involves two weight matrixs made up of w_1, positioned between the input and hidden layer along with w_2 between the hidden and output layer. Finally, b_1 and b_2 refer to the bias vectors [75].

2.5.2 Long Short-term Memory

A long short-term memory (LSTM) network is a specific type of recurrent neural-network (RNN) often used in the field of deep learning classification [89]. There are three specific parts of an LSTM cell which involve a forget (f), input (I) and output (o) gate, as depicted at Figure 2.8.

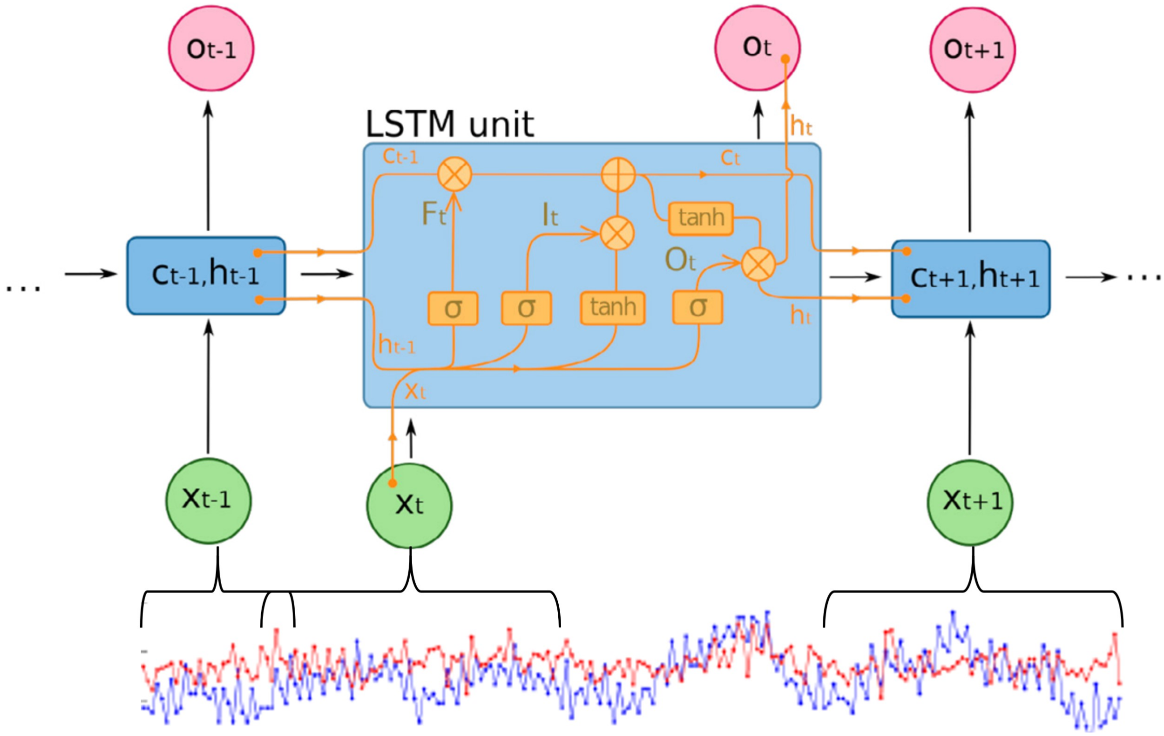


Figure 2.8: The architecture of an LSTM cell with input vector (X), cell output (h), cell memory (c), input (I) forget (f) and output (o) gates [135].

Using an LSTM model requires several steps, started by using the sigmoid layer (forget layer input) to decide what information to throw away at the start, as depicted at Equation 2.7.

$$f_t = \sigma \left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2.7}$$

Secondly, a decision over what new information to store in the cell state is considered which uses the input gate layer to decide which value to update (Equation 2.8), followed by a tanh layer to create a vector of the new candidate value $\tilde{C}_t$ as depicted at Equation 2.9.

$$i_t = \sigma \left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2.8}$$

$$\tilde{C}_t = \tanh \left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{2.9}$$

Thirdly, as depicted at Equation 2.10, the old cell state *(C_{t-1})* is updated into the new cell *(C_t)*. Next, to forget the data previously from the cell, the old cell state is multiplied by (f) and then $i_t * \tilde{C}_t$ is added.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{2.10}$$

Next, Equation 2.11 depicts the sigmoid layer which is utilised to determine which part of the cell state that is needed to output.

$$o_t = \sigma \left(W_o \left[h_{t-1}, x_t \right] + b_o \right) \tag{2.11}$$

Equation 2.12 utilises a tanh on the cell state and multiplying the output of a sigmoid gate results in:

$$h_t = o_t * \tanh \left(C_t \right) \tag{2.12}$$

To finish, a SoftMax layer follows an LSTM cell using cross entropy loss function to result in an prediction output from the classes.

2.5.3 Deep Belief Networks

Unsupervised Deep Belief Networks *(DBNs)* are beneficial as they learn to extract a deep hierarchical representation of the training data which can then be used as features within a supervised machine learning classifier. *DBNs* are generative models and are a composition of stacked Restricted Boltzmann Machines (*RBM*) and Sigmoid Belief Networks [142], [206]. Depicted at Figure 2.9 illustrates the basic blocks of a typical DBN configuration that can consist of several RBMs with an input and output layer.

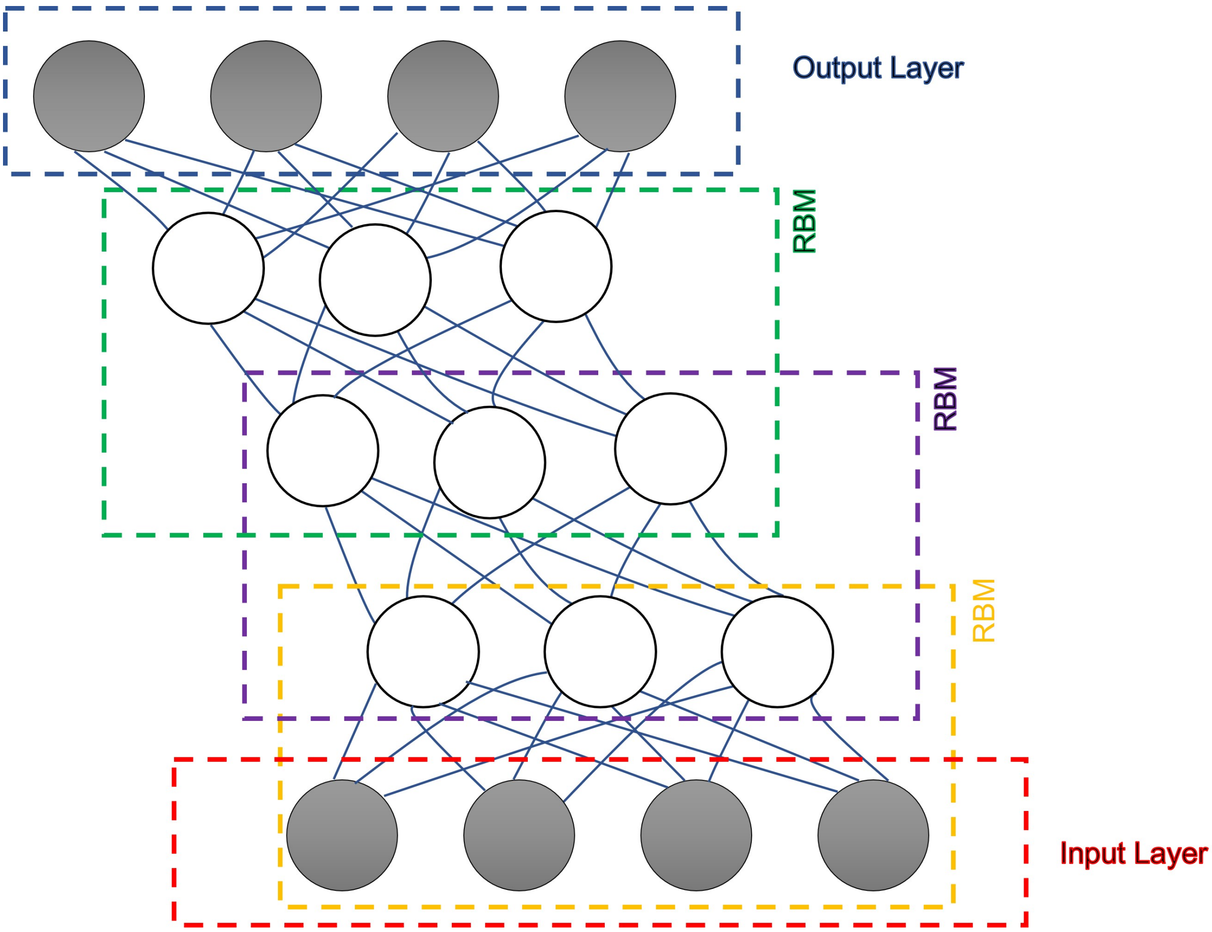


Figure 2.9: Schematic Architecture of a Deep Belief Network: comprised of three-stacked RBMs including an input and output layer.

RBMs are stacked and trained in a greedy manner by training in a sequential way, feeding lower layers' results to the upper layers to form *DBNs* [249]. They model the joint distribution between the observed vector x and the ℓ hidden layers h^k where $x = h^0, P\left(h^{k-1} \mid h^k\right)$ is the distribution of the units conditioned on the hidden units of the *RBM* at level k, and $P\left(h^{\ell-1}, h^{\ell}\right)$ is the visible-hidden joint distribution in the top *RBM* as demonstrated at Equation 2.13.

$$P\left(x, h^1, \ldots, h^{\ell}\right) = \left(\prod_{k=0}^{\ell-2} P\left(h^k \mid h^{k+1}\right)\right) P\left(h^{\ell-1}, h^{\ell}\right) \tag{2.13}$$

Unsupervised learning is used to train the *RBMs* of the *DBN* to automatically construct features and reconstruct inputs. The Gibbs Sampling based contrastive divergence method [35] is used to train the *RBM* as shown below:

1. Typically, the combined data is fed into the RBM as the input $x = h^{(0)}$ of the first layer.

2. Next the activation probabilities of the hidden layers are calculated using Equation 2.14:
$$P\left(h_j \mid X\right) = \sigma\left(b_j + \sum_{i=1}^{m} W_{ij} X_i\right) \tag{2.14}$$

3. Then the activation probabilities of input layers are calculated using Equation 2.15:
$$P\left(X_i \mid h\right) = \sigma\left(a_i + \sum_{j=1}^{n} W_{ij} h_j\right) \tag{2.15}$$

4. The edge weights are updated where α is the learning rate using Equation 2.16:
$$Wij = Wij + \alpha\left(P\left(h_j \mid X\right) - P\left(X_i \mid h\right)\right) \tag{2.16}$$

After training the first *RBM* the edge weights are frozen and the remaining *RBMs* are trained using the same contrastive divergence method with the output of previous trained *RBM* being used as the input of the next *RBM*. After training has completed, the *DBN* features are extracted from the top hidden layer and a hidden unit of the learned network structure is used as the input layer for a supervised *ML* models. The *DBN* is essentially used as a feature selection mechanism for the machine learning models as it is used as a representation learner compressing the original input vector for the *ML* models to use.

2.5.4 Dynamic Time Warping

Dynamic Time Warping *(DTW)* has been shown as an effective algorithm in many areas, particularly time-series data [192] in being able to measure the similarity between two temporal sequences by finding the best mapping position with the minimum distance between two points. The distance between two points can be depicted more formally as: $x = [x_1, x_2, \ldots x_n]$ and $y = [y_1, y_2, \ldots y_n]$ and then calculated using the Euclidean distance [140] as depicted at Equation 2.17.

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \tag{2.17}$$

However, an important factor for DTW in this method is that calculating the exact distance can only be achieved if the distances are the same in length [183]. Figure 2.10 depicts the difference of mapping the distances between Euclidean and DTW. It is interesting to note that Euclidean matches timestamps and DTW matches between the two time-series data through feature values.

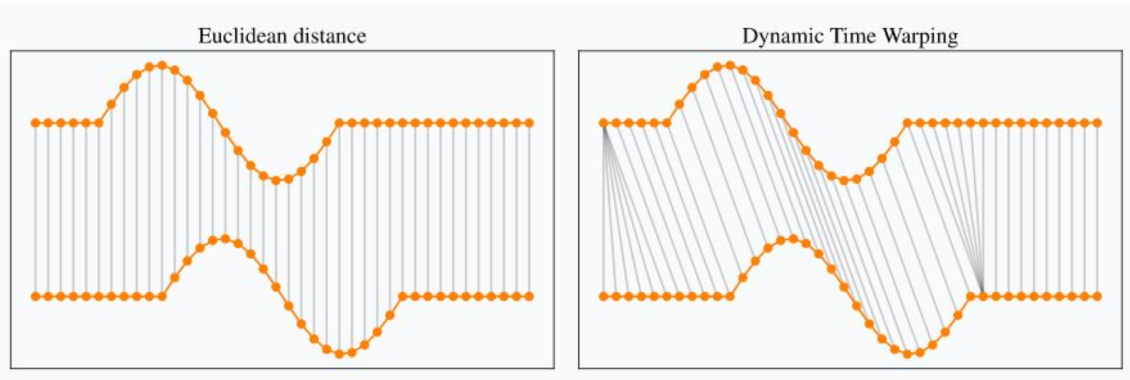


Figure 2.10: Differences between calculating the Euclidean Distance and Dynamic Time Warping of a trajectory [220].

DTW is a more appropriate approach as it finds the optimal mapping with the minimum distance between two data points [220]. The sequences for x and y are used to generate a grid whereby each point individually is labelled as (i, j) which results in the alignment between x[i] and y[j]. More formally using DTW, the distance between two paths can be represented as Equation 2.18.

$$\gamma(i,j) = d\left(x_i, y_j\right) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\} \tag{2.18}$$

Whereby, (x, y) is the sequence data, (i, j) are the local constraints for any given node. y is represented as the distance of warping path, d is the distance measured between each element of two sequences [120].

In previous work DTW has been shown to be highly successful in being able to classify large time-series based data sets [104]. In particular, the mapping concept of DTW enables a higher level signal processing technique used with human-to-human interaction [180], to reveal responses to the nervous system [81]. Additionally, this has been a proven method by being able to achieve high classification results through using a sensor fusion approach [90]. Furthermore, studies using DTW to classify accelerometer data between two participants found that the system has the capabilities to perform efficiently and effectively in real-time compared to alternative classifiers such as a Hidden Markov Model (HMM) based system [107].

2.6 Reflection and Research Challenges

2.6.1 Challenges

The design and development of pervasive computing technology systems pose numerous challenges. This section explores several key considerations related to these challenges and concludes by highlighting the identified gaps and research opportunities.

2.6.1.1 Privacy and Ethics

In recent years there has been significant attention towards privacy and ethical considerations particularly towards mobile sensing and pervasive computing technologies [88]. Both issues remain at the forefront of this area as the majority of individuals prefer for their mental health data to be kept secure and private [80].

There are many considerations when undertaking research such as processing data offline locally and avoiding cloud based storage which could encourage interception of the stored data. Additionally, before users are introduced to the study, care should be made to acknowledge with consent of the individual to understand what data will be obtained and how it is being processed. There has been considerable work to develop approaches to keep data secure and private, particularly through the General Data Protection Regulation (GDPR) in the EU which enables more control to users to access their stored data and understand how it is being processed.

2.6.1.2 Data Collection

A common problem with pre-existing work is that only a limited number of trials collect or analyse real-world data, since experiments that simulate situations within the environment may not demonstrate the same patterns as real-life situations. Additionally, the recruitment and engagement of participants to undertake real-world experiments is an on-going concern as many do not initially see the added benefits of supporting the work and only see what they can get out of it themselves. Similarly, considerations should be taken as how to provide a suitable incentive or reward for taking part in the study which equates to their time taken.

2.6.1.3 Portability

The portability of devices continues to pose a significant challenge in research focused on developing systems for monitoring the environment and mental wellbeing while on the move. A notable concern pertains to the size of the devices needed for environmental monitoring, which requires numerous sensors and a connection to a mains electricity source or suitable resource for long-life battery [68].

In recent years, there has been some early development in more portable approaches through rucksacks and hand-held devices but limitations remain in the verification of data collected. There are some new approaches to monitor the environment and mental wellbeing including the Urban Mind mobile application but this takes the form of a digital-questionnaire [22]. However, at present this remains static and subjective as to what the individual thinks about their environment and personal-wellbeing rather than based on objective real-world sensor data.

2.6.1.4 Battery Life

The proliferation of sensors and the ability to incorporate them into smart devices and mobile technology offer a significant opportunity to transform real-world monitoring of the environment and mental wellbeing. However, many occasions require the ability to unobtrusively obtain data which can result in little room for a battery to power the devices. Further work is needed to extend the life and capacities of batteries whilst also keeping them relatively small for use within devices. With this in mind it remains difficult to collect large amounts of wellbeing data.

2.6.2 Current Gaps and Research Opportunity

The following considerations from the literature studied in this chapter are expanded on below:

1. To date, the Exposome Concept has been an effective method in measuring the impact of the environment towards health outcomes for an individual but is lacking in terms of a focus towards a mental wellbeing centric investigation. A common limitation in the literature highlights the need for a large amount of data to be available to measure the health impact and gain a complete assessment for an individual, which has often involved using pre-existing, typically out-dated datasets. Further work to speed up the data collection and analytical approaches using technological state-of-the-art sensing equipment may offer greater opportunities to fully quantify the approach.

2. Only a few attempts at present in the literature have considered environmental noise and the impact this can have towards physiological responses and mental wellbeing [102]. Therefore, research conducted into other environmental factors through a multimodal approach pose a significant opportunity to quantify the relationship of the entire environment on physiological responses which may also improve real-time interventions.

3. Sensing systems pose a significant opportunity with the advancements of mobile sensors and sensing technologies which has enabled new approaches for capturing, diagnosis and assessment of the environmental impact, physiological responses and mental wellbeing. A considerable amount of previous research has been conducted on measuring physiological responses using sensors such as HR, HRV, EDA, BVP and Accelerometer with these studies aiming to highlight the differences in wellbeing among individual's based on the levels of the physiological modalities [157], [128], [102]. Although there are increasing efforts towards capturing a range of pollutants through environmental monitoring, most of which are directed towards use in a fixed location and require significant cost. Despite this, there is a start

in the literature with developing bespoke systems but these are only able to capture a limited range of environmental factors [57].

4. Previous studies have shown the potential of modelling time-series trajectory data can be an effective method of quantifying the impact of exposure at a specific location. Similarly, when trajectories are enriched with semantics they can offer new opportunities to gain a more wholesome understanding into human behaviour and movement patterns [246], although this is limited in terms of studies collecting and utilising real-world data. Additionally, research into the areas of 'episodes' or 'segmenting' in the literature has been very little. Therefore, research using real-time, real-world semantic trajectory data may unravel the impact at the point of exposure further.

5. The literature is limited in demonstrating how spatial visualisation concepts and techniques can show the direct impact upon the individual using the collected sensor data. Noisespy [100] effectively demonstrates the use of heat-mapping as a way to see how environmental noise changes across a city location. Although useful, there are concerns over the ability to allocate an individual sensor reading to each cell and to represent additional collected sensor data.

To address the gaps identified in the research above, the work in this thesis explores the relationship of the environment, physiology, human behaviour to quantify the impact to momentary mental wellbeing. Following the issues identified with the current *Exposome Concept*, the approach is furthered to explore the use of technology in the form of smartphone and miniaturised sensors. It is hoped that with this new framework will enable the quantification of the impact towards the environment on inference of wellbeing.

Chapter 3

DigitalExposome: Wearable and Mobile Sensing Technologies for Transforming Wellbeing

3.1 Introduction

Traditionally the collection of environmental, physiological and self-labelled wellbeing data required a vast amount of resources; mainly collected through questionnaires and multiple datasets over a long period of time, which can be challenging [41]. This was particularly evident in the previous chapter with the Exposome Concept [238], whereby each component must be obtained before analysis can take place. The general increase and availability in sensor-based technologies and mobile sensing, particularly in affordability and size, enables a more accurate analysis and understanding specific to location-based exposure [52], [47]. A technological approach using sensing devices to monitor an individual's affective state could be incredibly useful in improving assessment tools and techniques [18].

In this chapter, describes the exploration, design and framework for the 'DigitalExposome' Concept. This new approach presented aims to delve deeper into the relationship and quantification of environment, human physiology, behaviour characteristics and momentary mental wellbeing. Utilising the DigitalExposome concept to quantify impact to mental wellbeing, there are new opportunities in mobile sensing which offer exciting new possibilities in tracking a users' feelings as they journey between different urban environments. Finally, an overview of the data collection tools employed to gather information pertaining to the environment, physiological measurements, and momentary wellbeing assessment.

The following contributions in this chapter are as follows:

1. Introduction and exploration of DigitalExposome and the significant opportunities it can have in providing the step to quantify the relationship between environment, human physiology and mental wellbeing utilising digital technology solutions, mobile sensors, smartphones and data science through a novel conceptual framework.

2. Present the design and development of several custom-built experimental tools to (i) sample a range of environmental factors (such as Particulate Matter (PM1), (PM2.5), (PM10), Reducing and Oxidising gases, Ammonia (NH3) and Noise), (ii) Obtain physiological responses (such as Heart-Rate, Blood Volume Pulse, Heart-Rate Variability and ElectroDermal Activity), (iii) Smartphone technology to understand impact of environment and wellbeing at the point of exposure.

3. Three individual datasets with over 50,000 samples of data encompassing real-world environmental sensor, physiological and momentary mental wellbeing data.

3.2 Introduction to the DigitalExposome Conceptual Framework

DigitalExposome can be defined as the framework to quantify the relationship between the environment, physiology and mental wellbeing. This new field enhances and furthers the work undertaken as part of the previous *Exposome Concept* by utilising a range of digital technologies in the form of mobile sensing devices and data science concepts in order to collect the required data to assess and calculate the exposome more accurately and precisely. With the ultimate aim to measure multiple environmental factors using mobile technologies and then quantify them in real-life settings.

The combination of multiple data collection methods helps to support *DigitalExposome* and gain a better understanding into how exposures to the environment can impact mental wellbeing. Figure 3.1, depicts the range of technological data collection methods made up of fixed, wearable and smartphone sensing devices to support the concept. Through use of the concept many opportunities and possibilities of being able to investigate the relationship between the environment and wellbeing can be explored.

Figure 3.1: DigitalExposome Concept: The quantification step to unravel the relationship of environment and physiology on mental wellbeing

DigitalExposome is primarily made up of two parts: data collection and data analysis with both aspects using technological advances to accurately calculate the exposome and therefore close the gap in associations between the environment and human physiology. To aid in quantifying the process, the utilisation of data from sensors that show how an individual has been exposed to air pollutants is required. This will be a key part of DigitalExposome, where both terms are clearly connected through their vision of being able to capture the true exposure that an individual has been exposed to. As noted in the literature data that is generated through the use of technology, such as sensors are ideal to monitor various exposures and enable the possibility to link this to health [133], [138]. The DigitalExposome concept builds on previous work such as 'Digital Phenotyping' which involves quantifying the human individual-level of behaviour in a step using data collected from smartphones and wearables [177], [223]. This has been shown to be highly effective in developing awareness of the impact towards depressions and anxiety to support real-world prevention and treatment [153], [205].

3.2.1 Conceptual Framework

The conceptual framework outlines the several processes involved in the Digital-Exposome approach and aids to assess the correlations between urban exposures and self-report wellbeing as depicted at Figure 3.2. In this case, it is clearly evident as highlighted in the conceptual, sensing and computing layers that DigitalExposome creates a significant role in the amount of knowledge that can be used to quantify the relationship. The four key layers with further description and explanations are as follows:

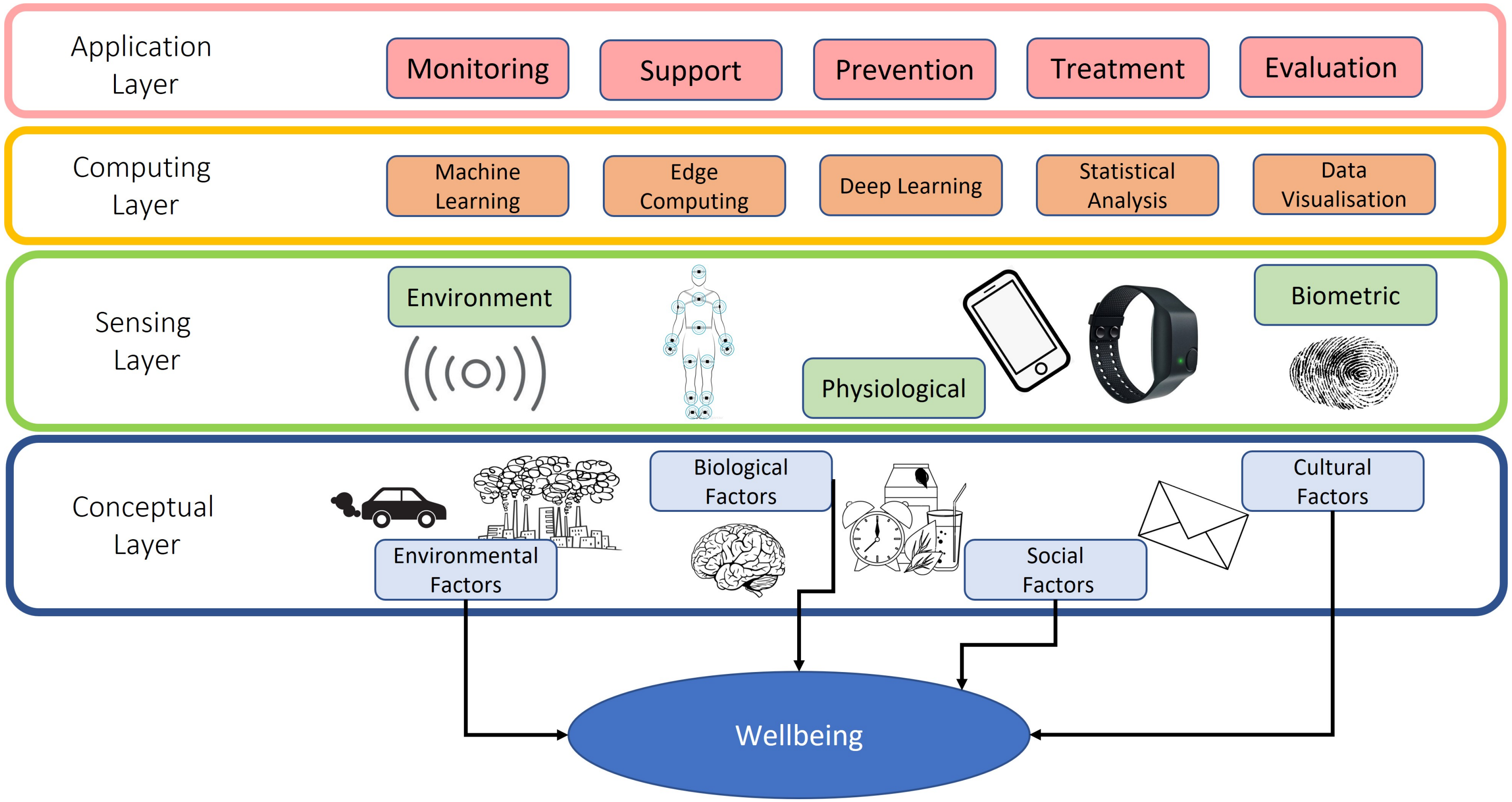


Figure 3.2: The Conceptual Framework for DigitalExposome on the Quantification Step for Mental Wellbeing.

- **Conceptual Layer:** The conceptual layer contains the four main areas in the literature that have been shown to have an impact towards mental wellbeing as identified within the research [125]. In this thesis, the work specifically targets the environmental aspects. The environmental factors could be made up of chemical, air pollution, climate change and access to poor water quality. Biological factors may include factors such as genetics, brain qualities and physical health. In addition social aspects which could include issues surrounding relationship with people in the vicinity, quality of sleep, a nutritional diet and regular exercise. Finally cultural factors may include aspects such as stigma around wellbeing and available resources for support.

- **Sensing Layer:** The sensing layer presents the three physical devices made up of several smart mobile sensors which include: physiological systems (such as a wearable) to monitor HR, HRV, EDA, BVP, Accelerometer and body temperature). Secondly, environmental monitoring systems made up of several sensors to capture and monitor in real-time pollutants such as noise, air pollution, particulates and gases. Finally, the layer involves the use of smartphones that can unobtrusively monitor. The work in this thesis explores the environmental, human physiological factors and smartphone mobile application to quantify the sensing approaches of DigitalExposome.

- **Computing Layer:** The computing layer lists the several key core data science and analytical techniques which will enable a greater understanding of the impact to mental wellbeing. This includes techniques such as machine learning and deep learning to classify wellbeing, edge computing to assess the potential impact on the device. In addition, statistical analysis to understand the variable importance of the collected sensor data. Finally data visualisation will support the work of being able to map using the GPS data points at where individual's wellbeing is impacted more. The work in this thesis has explored many areas of the Computing Layer with approaches such as statistical analysis, deep learning, and machine learning in Chapter 4. Furthermore, data visualisation has been explored through heatmaps and Voronoi in Chapter 5.

- **Application Layer:** Finally, the application layer of the concept presents the potential applications of DigitalExposome. This includes a greater perspective into active monitoring of wellbeing, environment, and physiological on-body variables. In addition, support, prevention and treatment can be better controlled by understanding the issue of the environment. In the context of this work, this thesis has supported the DigitalExposome approach in terms of monitoring mental wellbeing through the use of smartphone technologies as demonstrated in Section 3.3.3. Finally, approaches to support, prevent and evaluate DigitalExposome have been considered in Chapter 5, between the three applications of this research.

3.2.2 Opportunities with DigitalExposome

Development of the DigitalExposome concept offers a range of new and exciting opportunities to explore the impact of the environment and physiology to mental wellbeing. In this section, a range of techniques and tools are presented in a way to explore the possibilities of this concept with the addition of technology and mobile sensing.

3.2.2.1 Statistical Analysis

Various methods including descriptive statistics, linear regression and spatial analysis have previously been used to identify and gain a better understanding of data which can be used to monitor the overall impact and identify patterns, commonalities and correlations [27]. Previous research have used statistical analysis tools and techniques to measure the impact of environmental factors on the body [224].

The use of a statistical approach and the combination of body and environmental sensors have been shown to improve the general prediction to an individual's emotions [102]. Using data that is combined with location tracking (GPS) can offer the opportunities to carry out spatial analysis or correlation analysis [219] [101]. Results have shown a clearer understanding as to the triggers of behaviour and wellbeing within a specific location. Additionally, the use of Principal Component Analysis (PCA) provides an effective statistical mechanism to understand the impact relationship between two or more variables [99], particularly within large datasets [141]. To date little efforts have focused on statistical analysis to investigate the relationship of the environment to wellbeing, with limited work previously focused on noise [100].

3.2.2.2 Machine Learning

The use of machine learning can offer a statistical approach to predict and infer data collected by learning and building algorithms [33]. Previous work in the literature has shown on numerous occasions that using machine learning can enable a new perspective on the data as well as discover new patterns and correlations between variables. This has involved some but otherwise limited attempts to classify the impact of a singular environmental pollutant towards an individual namely Benzene and PM2.5 [17], [53]. In both cases the potential to utilise data science techniques such as machine learning are evident in results of high classification.

In the concept of DigitalExposome on the proposed framework at the 'Computing Layer', demonstrates that both deep learning architectures and machine learning classification would be advantageous and could be utilised to understand the data to monitor the direct impact towards individuals. To date there have been little research attempts using machine learning to monitor the impact of poor air quality to an individual's wellbeing, with previous work focusing on PM2.5 [196], noise/ audio [101], UV [102].

3.3 Data Collection

This section discusses a range of data collection approaches to further the DigitalExposome Concept; such as wearables, smartphones, self-report, fixed and portable sensors which can be utilised as a multimodal fusion approach to gather a large amount of data reliably. Mobile sensing has become a very popular approach of data collection due to their ability to collect large amounts of data efficiently and without much human interaction. This has involved developing a custom-build environmental monitoring system using a range off-the-shelf low-cost sensors, capable of sensing: Particulate Matter, Oxidising and Reducing Gases, Ammonia and Noise with Internet-of-Things technology on-board. Readily available systems with these sensors built in are often very large, in fixed sensing stations and are not practical for 'in-the-wild' experiments. In order to gain as much information as possible from an individual a range of collection types explored for DigitalExposome are given at Figure 3.3.

Figure 3.3: Several data collection types for DigitalExposome.

As a result, this includes a wellbeing application to self-report wellbeing states, custom-build environmental sensing system, physiological wrist wearable and a cross platform mobile application to self-report wellbeing states and to collect environmental data in situ at the point-of-exposure.

3.3.1 Wellbeing Smartphone Mobile Application

To encourage and support a more digital-approach to self-recorded wellbeing, an Android-based mobile application was developed, as depicted at Figure 3.4. On opening the mobile application, five well known emojis are presented to the participant, from very happy to very sad. Each emoji is placed on buttons from *1 = negative/ Very Sad* to *5 = positive/ Very Happy*.

Figure 3.4: Screenshot of smartphone wellbeing application made up of the five well-known emojis from crying to very happy.

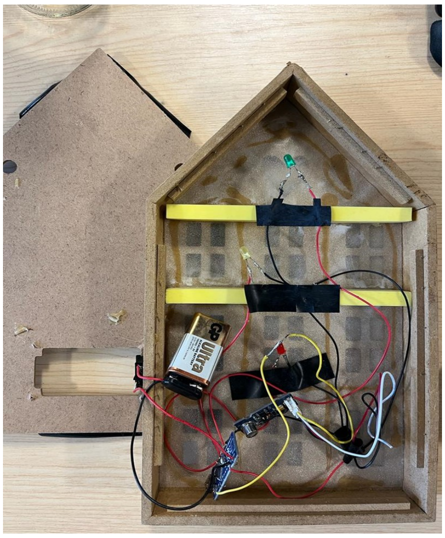
The general idea with this mobile application is that in a research scenario, participants would be constantly reminded to label their wellbeing through the emojis as to ascertain how they were feeling. Several previous studies have used this self-report approach [102] to quantify wellbeing states. Additionally, research shows that wellbeing labels can change quickly whilst moving through environments so a 'quick label' process is paramount in obtaining wellbeing at the correct time [5].

To collect wellbeing data from participants the 'Personal Wellbeing Index for adults' has been selected in this scenario which asks the user how they are feeling with their life as a whole [46]. This has been adapted further for this work, in the form of a five-point Likert SAM scale [31] with an emoji at each point. This approach has been shown to be a proven method in other literature studies for self-report wellbeing [93], [26], [242].

3.3.2 Environmental Monitoring System

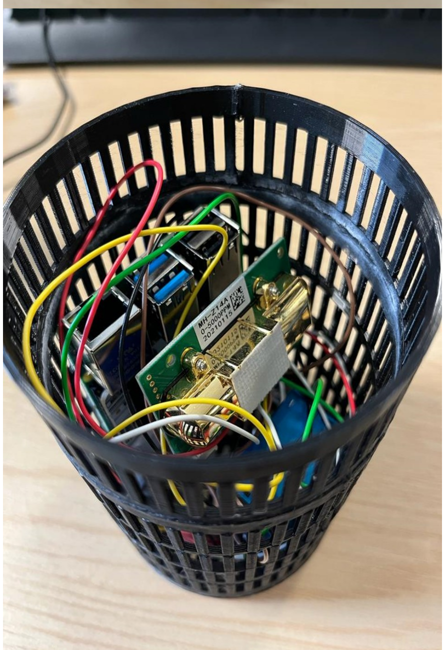
To observe changes and capture levels of air quality in the environment an adequate monitoring system must be sought. As identified in the literature, readily available environmental monitoring systems are typically unsuitable either for high cost, size and adaptability into real-world experiments. In this thesis, four prototypes have been developed with a range of environmental sensors with a particular focus on using low-cost alternatives. Table 3.1, depicts the prototype devices, including device name, images and short description.

Table 3.1: Four environmental sensing prototypes with embedded electronics and low-cost sensors, developed to support the DigitalExposome process.

Prototype Images	Device Description
 	**Enviro-house:** The first prototype is a fixed device either to be placed within a room or on a wall. Within the system is one environmental sensor that is capable of sensing harmful gases such as Carbon Dioxide and alcohol based products. The device is built using an Arduino Nano and has several LEDs to perform visual feedback based on the levels of gases sensed within the room. The power to the device involves a small battery. This is built on a traffic light scenario, where for example if the air quality is good the 'house' device will illuminate green and poor air quality will light red. The device works in real-time and will respond to actions such as opening the window which in the literature shows can improve indoor air quality [234]. An approximate costing of the device including low-cost sensors, embedded board and battery is around £50.

Continued on next page

Table 3.1: Four environmental sensing prototypes with embedded electronics and low-cost sensors, developed to support the DigitalExposome process. (Continued)

Prototype Images	Device Description
 	The second prototype is a 3D printed cylinder suitable for being held in the hand whilst walking around an environment due to the positioning of holes around the device. The size of each device is 10cm x 8cm. Powered from an external battery the system is connected to a Raspberry Pi 4 through several embedded sensors are incorporated including Particulate Matter (PM1), (PM2.5), (PM10), Reducing and Oxidising gases, Carbon Dioxide (CO2), Volatile Organic Compound (VOV), Ammonia (NH3) and Noise. The device including case, electronics and sensors cost around £125 to build.

Continued on next page

Table 3.1: Four environmental sensing prototypes with embedded electronics and low-cost sensors, developed to support the DigitalExposome process. (Continued)

Prototype Images	Device Description
	**Enviro-IoT:** The third prototype developed is a 3D printed device specifically for outdoor use. Each system incorporates a Raspberry Pi 4, several environmental sensors and built-in Internet-of-Things technology. Specifically the devices target Particulate Matter (PM1), (PM2.5), (PM10), Nitrogen Dioxide (NO2), Oxidising gases, Ammonia (NH3) and Noise. The air is sampled every 5-minutes with the on-board device averaging the air quality level to give an hourly rate. Data is sent to a secure online database for analysis which is a similar approach carried out by industry standard off the shelf devices The device including 3D printed case, electronics, sensors and router for internet cost around £300 with a maintenance for sim-replacement costing £50 every 6-months.

Continued on next page

Table 3.1: Four environmental sensing prototypes with embedded electronics and low-cost sensors, developed to support the DigitalExposome process. (Continued)

Prototype Images	Device Description
	**Enviro-Rucksack:** The fourth prototype builds on prototype 3 with a similar set up of a custom-built rucksack to unobtrusively collect environmental data whilst walking around an environment. Within each rucksack houses a Raspberry Pi 4+ along with several environmental sensors such as Particulate Matter (PM1), (PM2.5), (PM10), Reducing and Oxidising gases, Ammonia (NH3) and Noise. The cost for developing this including rucksack, electronics, sensors and battery power costs £250.

3.3.2.1 Enviro-IoT: A Low-Cost Alternative to Assessing Air Quality Levels

The Enviro-IoT is a custom-built environmental sensing device equipped with a Raspberry Pi 4 to continually sample the environment with a timestamp, obtained every 5 minutes and then averaged to give the mean concentration per hour. Additionally, within the 3D printed case a selection of small, lost-cost air quality sensors are placed to observe changes, which include the following variables:

- Particulate Matter (PM)
  - PM 1.0 ug/m3
  - PM 2.5 ug/m3
  - PM 10 ug/m3
- Nitrogen Dioxide (NO2)
- Oxidising gases
- Reducing gases
- Ammonia (NH3)
- Noise (dB)
- Timestamp (Format of DD/MM/YYYY, HH:MM:SS)

The circuit is encased within a 3D plastic container to ensure the electronic components are kept clean, secure and water-tight. As shown at Figure 3.5, depicts the electronic circuit made up of the key components required to obtain air quality levels. To place the device in a fixed location, the battery can be replaced with a connection to the mains power.

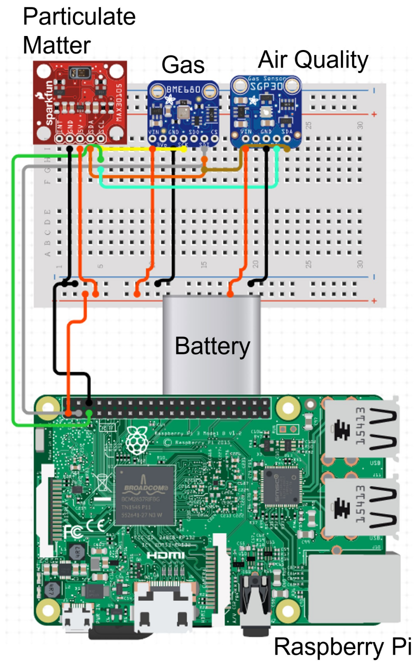


Figure 3.5: Electronic circuit of the Enviro-IoT and Enviro-Rucksack depicting a Raspberry pi, power through a battery and several environmental air quality low-cost sensors.

3.3.2.2 Deployment

In July 2022, two Enviro-IoT devices were positioned at an outdoor location alongside the Automatic Urban and Rural Network (AURN) stations in different urban environments across Nottingham City Centre as depicted at Figure 3.6. The two areas were selected as part of a pilot study working with Nottingham City Council (NCC) and the added interest of monitoring (NO2) and (PM) levels of air pollution. The two pollutants are of significant importance due to NCC's commitment to reduce the level of (NO2) and (PM) over the next 5 years as currently they remain in one of the UK Government's 'Air Quality Mangement Area' [44].

The Enviro-IoT were fitted on top of the AURN with cables thread through an inlet to access power and network connections. Due to the location of both sensors, the Enviro-IoTs were placed into a wire cage to protect them against vandalism. In both cases, the Enviro-IoTs have been placed as close to the other sensors that make up the AURN as possible.

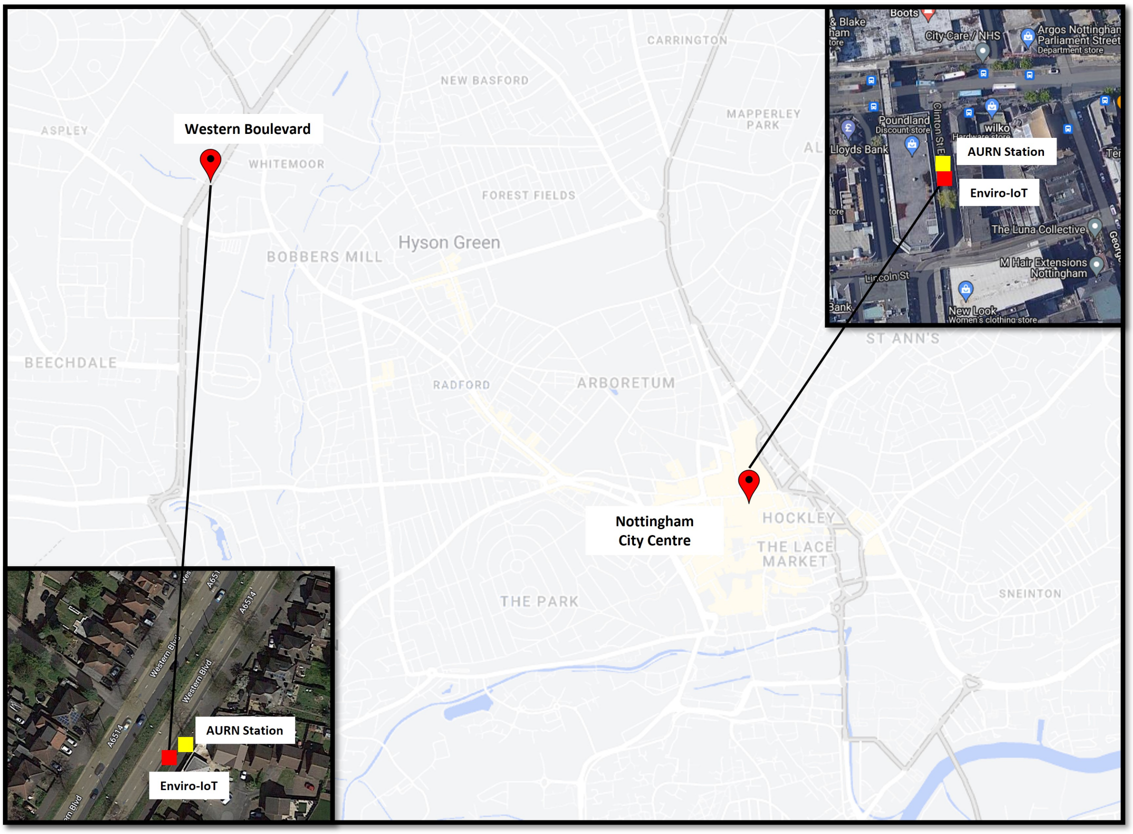


Figure 3.6: Two locations of the Air Quality Monitors (Enviro-IoT) deployed in Nottingham City Centre and Western Boulevard alongside the DEFRA AURN.

3.3.3 Urban Wellbeing: Design of an Ecological Momentary Assessment Tool

In this section describes the design and development of Urban Wellbeing as a cross-platform (iOS and Android) interactive ecological momentary assessment tool that aids in supporting the work of DigitalExposome by unravelling the relationship between the environment and mental wellbeing. Urban Wellbeing is made up of multiple on-board mobile sensors for noise detection and fixed external environmental sensors where data is provided by The Department of Environment, Food and Rural Affairs' (DEFRA) and Automatic Urban and Rural

Network (AURN) [67] to obtain location specific air quality data, as demonstrated at Figure 3.7.

Figure 3.7: Urban Wellbeing Mobile Application and the six sub-processes.

The Urban Wellbeing application builds on previous work such as NoiseSpy [100] which involves the collection of noise levels using smartphone technology to monitor and map the impact along a journey within the environment. In addition, previously paper-based surveys have been conducted to understand how people feel within the environment [82]. By incorporating more data from the environment using technology will enable a deeper understanding of the impact towards wellbeing and behaviour. The full system process of the mobile application as detailed at Figure 3.8, depicts the structure of the application and data obtained specifically to the device: smartphone and DEFRA AURN.

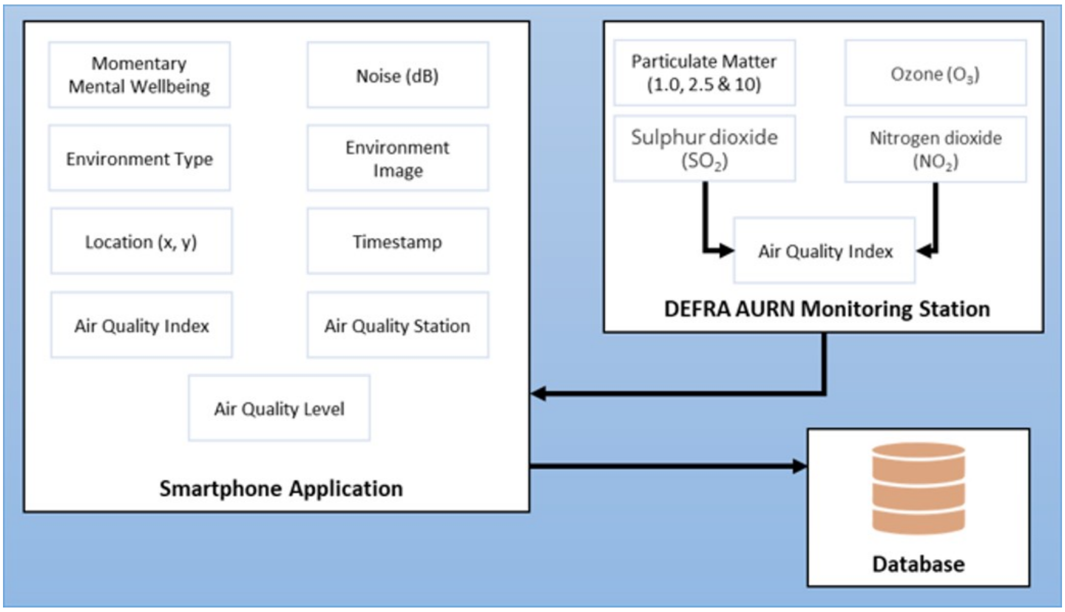
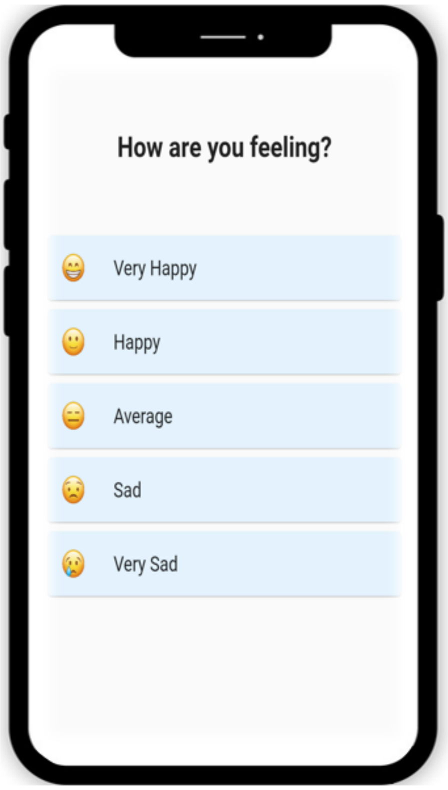


Figure 3.8: Urban Wellbeing system processes: The combined data collection and processing workflow of Urban Wellbeing involving Smartphone application, AURN Air Quality Monitoring Station and Database.



Emoji	Wellbeing Score
😁	5 – Very Positive
🙂	4 – Positive
😑	3 - Neutral
😟	2 - Negative
😢	1 – Very Negative

Figure 3.9: (Left) User view of Urban Wellbeing demonstrating the wellbeing assessment page, (Right) table equivalence of emoji vs scoring comparison.

As part of each assessment completed on the Urban Wellbeing application, the following data is obtained:

1. **Momentary Mental Wellbeing:** The application users are required to record their wellbeing using five well-known emojis and text-equivalent meanings displayed on buttons as depicted at Figure 3.9. In addition, the table at Figure 3.9 shows how wellbeing is calculated in terms of assigning an individual score to each emoji, from *1=negative/low* to *5=positive/high* . The 'Personal Index for Adults' self-assessment of measured satisfaction, as previous utilised has been adopted in this work to ask users how they are feeling with their life as a whole [46]. This has been adapted into the form of a five-point Likert SAM scale [31], to provide a proven method for self-reporting subjective wellbeing.

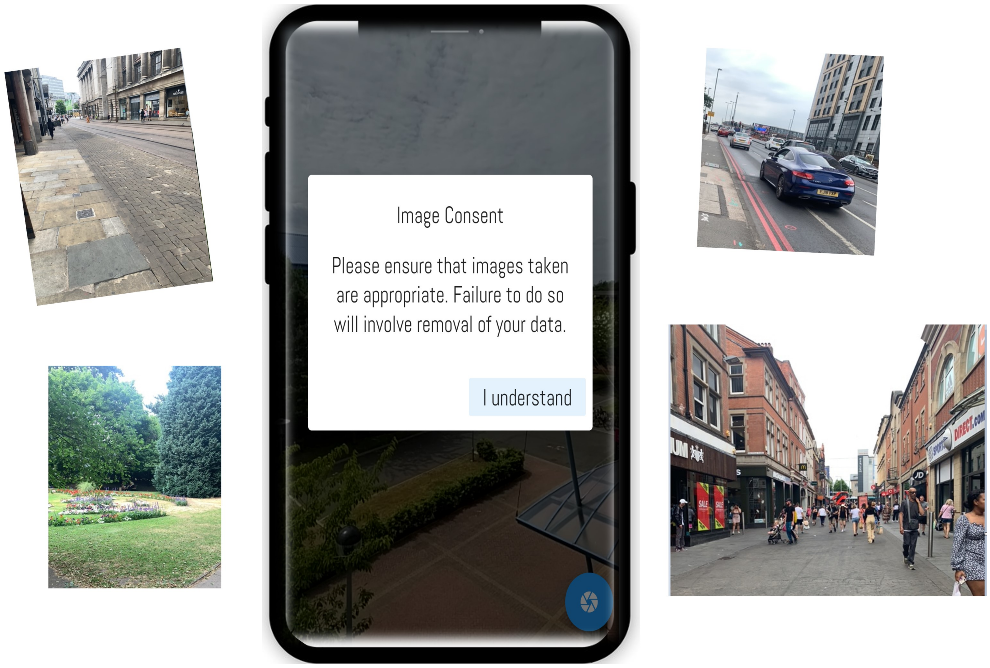


Figure 3.10: Image capture page demonstrating the option for users' to understand terms of use accompanies with four example images taken by participants.

2. **Environment Type:** To obtain where the user is located, a question is presented to the app user on whether they are currently within an inside or outside environment.

3. **Environmental Image Capture:** To gain an understanding into what the environment visually looks like when the assessment is taking place, participants are invited to capture an image using the in-build camera of the wider environment they are currently standing within. Depicted at Figure 3.10 demonstrates this page in action, along with the image consent which participants must 'agree to understand' prior to being able to take the image. Whilst testing the Urban Wellbeing application, four images at Figure 3.10 show examples of what was collected from several participants.

4. **Environment Air Quality and Noise:** The final process for the assessment is to collect the environmental levels of air quality and noise which is relative to the participants' specific location. On clicking 'Capture' as depicted at Figure 3.11, a loading bar will form around the noise icon which

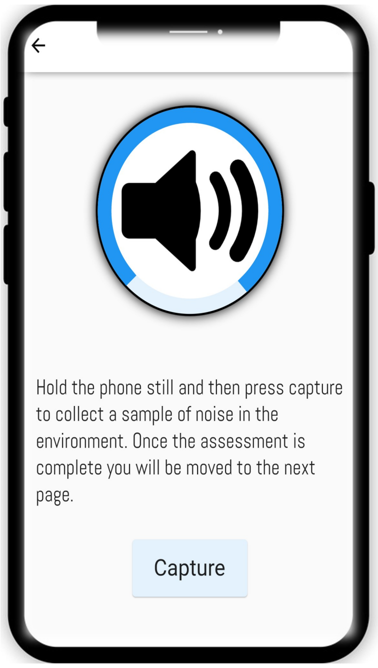


Figure 3.11: Screenshot of Urban Wellbeing assessment in process as air quality and noise sampling is obtained from the environment matching the location of the participant.

results in several processes of obtaining the participant's location, Air quality and noise level.

Air Quality readings are gathered by using the smartphone's specific location (longitude and Latitude) of the user and then data collected by the nearest DEFRA AURN station [67]. In particular, the data obtained includes: Real-time Air Quality Index, Air Quality Level and AURN station ID. The Air Quality Index is a process of combining all of the individual pollutants collected at each AURN station either taking the highest recorded value or averaging out the values across a period of time [73]. Secondly, noise is obtained from the assessment which is calculated by recording a series of noise clips in decibels which are collected over a period of 5 seconds. Finally, at this point all the data has been saved locally stored on the phone and with the timestamp (DD/MM/YYYY) added, the combined data is sent to a secure database ready for analysis. To protect users and in-line with ethics agreement, no data is stored in the database to identify a participant.

3.3.3.1 User Testing

A preliminary study has taken place to evaluate the design of the application and performance as a tool to capture live environment sensor data and wellbeing, to aid in quantifying the relationship between the two variables. In total, 5 participants were recruited for testing of Urban Wellbeing and interviewed following a full day of utilising the application in the wild. There was an equal download of Urban Wellbeing made up of iOS and Android platform between the participants.

When interviewed after using the Urban Wellbeing app for the day, it was found that the majority, 4 out of 5 participants stated that they enjoyed using the mobile application as a new approach to understand how the environment could play a part in either a positive or negative wellbeing. Overall, these participants were able to use Urban Wellbeing and complete the assessment process several times throughout their day. One participant struggled with the concept and process through the application and what they had to do. Some of the participants at times reported that the final screen (Figure 3.11) was a little slow at loading causing some issues with waiting around for the assessment to be completed.

A concern shared by one participant was that 'the application is not letting me past the first screen to capture an assessment.' One closer analysis, it was found that this participant had not clicked on the 'accept' permissions when prompted by the smartphone so therefore the application was not able to be used. All participants agreed that there should be some sort of inventive to carry out the experiment using the mobile application, with one stating 'perhaps a series of badges per environment or a step counter activity could be included'.

Following discussions with all participants after their day, it was mentioned that a loading screen before the assessment starts should be presented which briefly explains the main ideas and understanding of the work to be carried out. As such, a landing page has been developed into the application with three separate pages detailing the app itself, the walk outline and how the results of this study will be used, as depicted at Figure 3.12.

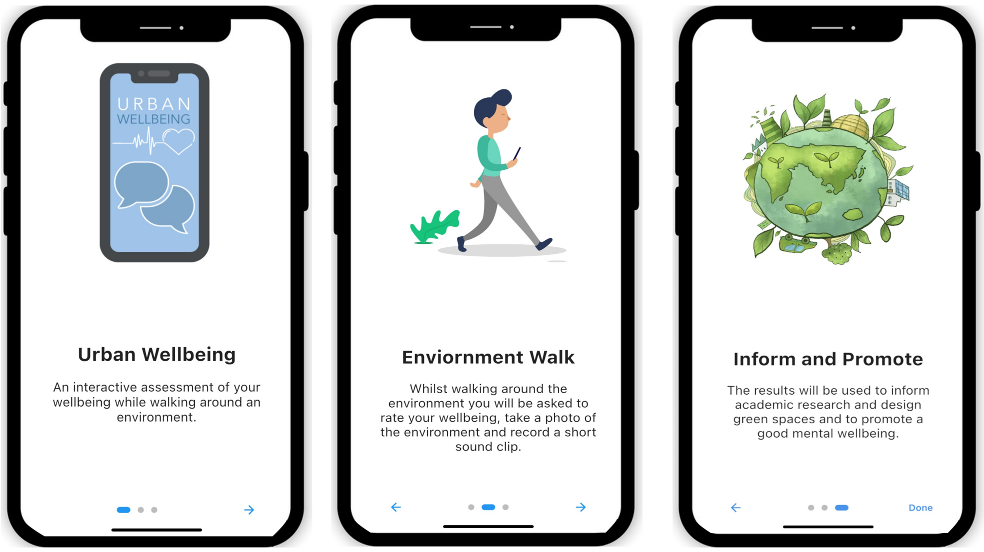


Figure 3.12: Urban Wellbeing mobile application three landing screens to give a general overview of the application before starting the individual assessment.

3.4 Datasets

To conduct this research, three different datasets are discussed in this section with a brief description of each provided, additionally to the collected variables obtained as part of the study. Section 3.4.1, presents the first dataset, that explores individual-level data in the form of trajectories with semantic enrichment and episodes to understand the impact of walking within an urban environment on mental wellbeing across an entire journey. A pictorial example of the experiment is provided in Figure 3.13 depicting the different data captured.

In Section 3.4.2, the second dataset is introduced which embodies a sensor fusion approach of aggregated data collected from a real-world, real-time study to quantify the relationship of the environment, physiology, behaviour and mental wellbeing. The experimental set up, similar to dataset 1 is demonstrated at Figure 3.13. Finally, Dataset 3 (Section 3.4.3) embodies several interactive assessments from the Urban Wellbeing smartphone application, to enable the development of computational solutions to detect changes to wellbeing in different environments in real-time. Further details and descriptions of each dataset are provided below:

Figure 3.13: Experimental setup and tasks associated with this research encompassing a (1) Enviro-IoT Custom-built environmental sensing kit, (2) E4 Empatica to measure physiological responses, (3) Custom-built smartphone mobile application.

3.4.1 Dataset 1: Semantic-Enriched Trajectories

The first dataset utilised in this thesis involved six participants (made up of 3 females and 3 males, aged between 18-50) who were all screened prior to the study to ensure they were fit and healthy. This study gained ethical approval from Nottingham Trent University's Human Invasive Ethics Committee. The six participants were instructed to walk through a range of different urban environments which should take roughly no longer than 40 minutes to complete the journey. The decision behind the length of time was due to previous user experience whereby a longer distance was found to be difficult. Additionally, the aim was to not exhaust participants as this could result in an impact to their bodies responses. Several studies with similar length of experiment time have since found it difficult to motivate participants to walk further [6], [5], [102]. The experiment was carried out around Nottingham Trent University, Clifton Campus.

Figure 3.14: Selected route taken by participants demonstrating the journey through a several green spaces (left picture) and a busy, polluted environment (right picture).

Figure 3.14, presents the journey route followed by all participants in the experiment. This is made up of the raw trajectory which involves the x, y and a timestamp. The selected route took the participants on a mixture of urban environments from several green to busy and polluted spaces which would help to demonstrate the impact of different levels of exposure to air pollutants. Each participant was given an E4 Empatica Wristband, Samsung Smartphone (pre-loaded with wellbeing application) and a custom built Enviro-IoT Monitoring Rucksack as depicted at Figure 3.13. Additionally, each is given an Enviro-IoT Rucksack as detailed in Section 3.3.2.1 to capture the current air quality every 20 seconds. Similarly, the Smartphone application samples a time stamp, longitudinal and latitude. While the E4 Empatica sensors' data is sampled at different rates with *HR at 1Hz and EDA, BVP, HRV* and body temp at*64Hz*. The data set resulted in 3,953 samples made up from the following factors as detailed in Table 3.2.

Table 3.2: The Semantic Trajectory dataset involving factors including: (1) raw trajectory, (2) semantic and (3) mental wellbeing data.

Raw Trajectory	**Semantics**	**Mental Wellbeing**
Longitudinal (x) Latitude (y) TimeStamp (DD:MM:YY, HH:MM:SS)	Heart Rate Heart-Rate Variability Electrodermal Activity	Self-Labelled emotions in the form of Emojis as outlined in Section 3.3.1.
	Accelerometer Blood Volume Pulse Body Temperature Particulate Matter 1.0 Particulate Matter 2.5 Particulate matter 10 Ammonia Reducing Gases Oxidising Gases Noise	

3.4.2 Dataset 2: Quantifying DigitalExposome

A total of 40 participants (made up from 25 Males and 15 females, aged between 18 and 50) who were all screened prior to the study to ensure those who participated in this research study were fit and healthy. This study gained ethical approval from Nottingham Trent University's Human Invasive Ethics Committee. Previous literature has used a similar number of participants to carry out studies in the same area [19], [106]. To obtain an adequate amount of data to evaluate the DigitalExposome framework, participants were given a specified route to walk around which would take no longer than 40 minutes to complete. As with Dataset 1, the choice in length of the walk was based on other studies and experiments with the aim of not exhausting participants which obvious could result in an impact to their body responses [6], [5], [102]. The route itself was selected within a range of different urban environments which involved several busy, polluted and open green spaces, as presented at Figure 3.15.

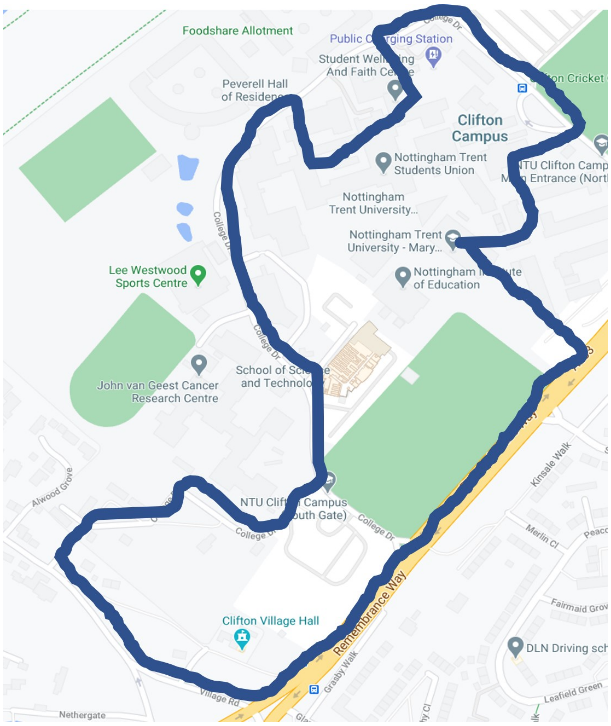


Figure 3.15: The pre-specified route made up of busy, polluted (right) and green spaces (top-left) taken by all participants in the experiment.

As with the previous dataset, in this experiment all participants were asked to wear an E4 Empatica, carry a custom-build environmental sensing station (as discussed in Section 3.3.2.1) in which both devices would be continually collecting data. To collect mental wellbeing data, this dataset adopted the use of the 'Personal Wellbeing Index for adults' which asks the user how they are feeling with their life as a whole [46]. In this work, this has been adapted in the form of a five-point Likert SAM scale [31] to provide a proven method for self-reporting subjective wellbeing. In our pre-installed mobile app the user is met with five well-know emojis, displayed on buttons from *1=negative/low* to *5=positive/high*. The idea is that the participant will be constantly prompted by the researcher to ascertain how they are feeling. Several studies such as [102] and NeuroPlace [5] have shown how momentary wellbeing labels can change quickly as moving through environments. The data set resulted in 41,037 samples and the following collected factors as depicted in Table 3.3.

Table 3.3: The DigitalExposome dataset involving factors including: (1) environmental, (2) physiological and (3) self-report wellbeing.

Environment	**Physiological**	**Wellbeing**
Particulate Matter 1.0 Particulate Matter 2.5 Particulate Matter 10 Noise Reducing Gases Oxidising Gases Ammonia Carbon Dioxide Volatile Organic Compound	Heart Rate Heart-Rate Variability Electrodermal Activity Accelerometer Blood Volume Pulse Body Temperature	Self-Labelled emotions in the form of Emojis as outlined in Section 3.3.1.

3.4.3 Dataset 3: Urban Wellbeing: Ecological Wellbeing Assessment

The third dataset involves the *Urban Wellbeing* mobile application as presented in Section 3.3.3, a dataset that contains real-world, real-time environmental sensor and self-report mental wellbeing data. In total there were over 50 downloads of Urban Wellbeing on the iOS and Android platforms resulting in 74 samples collected from across the United Kingdom as depicted at Figure 3.16. As part of the application, participants signed an in-app consent form that allows the collection of wellbeing data. The dataset was made up of the following variables:

1. Self-report mental wellbeing
2. Indoor/ Outdoor location
3. Location (Longitude and Latitude values)
4. Noise
5. Environment image taken by user
6. Timestamp
7. Real-time Air Quality Index stamp
8. Air Quality level
9. Air Quality nearest station for gathering data

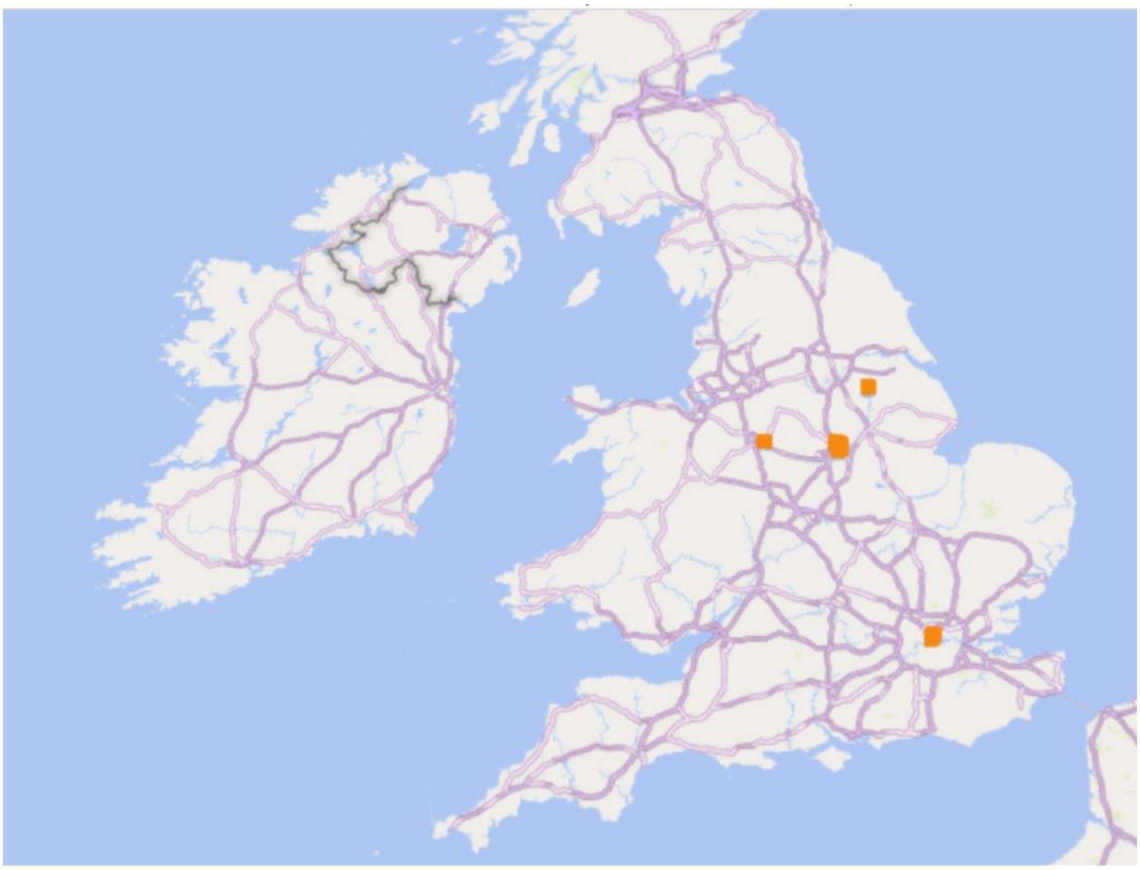


Figure 3.16: Map overlay from all participants that labelled. Each dot represents a location trace using the longitudinal and latitude values recorded by the application.

3.4.4 Data Storage Protection and Management

The collection of data requires careful considerations of various aspects, including the approaches employed to manage, protect and uphold the ethical standards throughout the research process.

Invasive ethics was granted by the Nottingham Trent University Ethical Committee (Application No. 068) to undertake the research activities that resulted in Dataset 1, 2 and 3. In line with the ethical agreement a data management plan was put in place for the collection of participants data to ensure that the analysis will be more efficient and avoid lots of irrelevant information [230]. Once the collected data from each participant device was downloaded, it was collated into a single .csv file combining the environmental, physiological factors and self-report wellbeing level. It was then transferred to a password protected storage database. Additionally, all non-digital data such as the consent forms were stored within a locked cabinet in the research lab. As a result of collecting personal information the data will be fully anonymous.

3.4.5 Data Pre-Processing

Following data collection, considerations must be made for pre-processing to make it suitable for modelling. As evident in Section 3.3, there is a range of sensors with varying sample rates which must be taken into account so as to either up or down sample data correctly to ensure that nothing is lost. Following each experiment discussed in future chapters', the data has been pre-processed to ensure that it is ready for analysis.

The important considerations with the data obtained firstly involves the Enviro-IoT device that records data once every 20 seconds. Secondly, for the E4 Empatica since there are varying rates of samples collected from HR at 1Hz and EDA, BVP, HRV and body temperature at 64Hz. Due to the varying sample rates, the physiological data collected (*EDA, BVP, HRV* and body temperature) were down-sampled to a rate of *1Hz* to match the sample rate of collected *HR* by the device. In addition, the collected environmental sensor data had to be up-sampled to match the sampled rate of the physiological data at *1Hz*. This was due to the low sample rate produced by the environmental device. Finally, the labelled data from the mobile smartphone was extracted and up-sampled to the same rate as the environmental and physiological data to 1Hz to remain consistent with the other data. To sample the data linear interpolation has been used on all datasets contained in this thesis [149]. If the two known points are given by the coordinates (x_1 , y_1) and (x_2 , y_2). The linear interpolant is the straight line between these points. For a value x in the interval (x_2 , x_1), the value y along the straight line is given from the equation of slopes as shown below:

$$y = y_1 + (x - x_1)\,\frac{(y_2 - y_1)}{(x_2 - x_1)} \tag{3.1}$$

Once completed, all signals from data sources were then normalised to bring all variables within the same range for both the data analysis and machine learning.

3.5 Conclusion

This chapter has presented the novel concept DigitalExposome and demonstrated the potential of employing a sensor fusion approach to further understand the relationship between the environment and the impact towards mental wellbeing. This led to the development of several environmental and mental wellbeing through the use of miniaturised low-cost sensors and smartphone technology. This involved the collection of 3 individual datasets made up of real-world, real-time environmental, physiological and mental wellbeing based on self-report data. Using DigitalExposome and the sensing technologies developed in this work could enable a new perspective on the environment and how people breath, feel and interact with different surroundings.

The next chapter, explores into the relationship using the wellbeing label to quantify further in using deep learning algorithms to understand the impact of environment on momentary mental wellbeing in urban environments.

Chapter 4

Multimodal Environmental Sensing for Mental Wellbeing Prediction

4.1 Introduction

As explored in Chapter 3, the DigitalExposome concept and framework was explored to provide a viable alternative step to quantifying the relationship to environment, physiology and mental wellbeing. In this chapter the work is detailed as a two-fold approach investigating the individual more directly through the use of semantic trajectories with episodes and aggregated data to quantify mental wellbeing. In this way, multimodal data is explored by using computational models to accurately classify mental wellbeing.

The experiments involve the use real-world multimodal collected sensor data from over 50 participants which showed that (i) the collection of environmental, physiological and momentary mental wellbeing enables accurate classification of mental wellbeing, (ii) time-series data has the potential to quantify the environment and mental wellbeing at the point of exposure.

As part of this chapter, the following contributions are made:

1. Conduct an extensive data exploration of collected environmental, physiological and momentary self-reported labelled mental wellbeing With the aim of being able to quantify the relationship between the environment, physiology, personal characteristics and wellbeing at an individual level and aggregated to observe the impact of each with over 50 users made up of more than 45,372 samples collected across two studies.

2. Utilised exploratory and statistical analysis techniques to begin to evaluate the link between environment, physiological variables and self-reported wellbeing. The results clearly show that the fluctuations in urban environments where pollution is high is impacted by individual's physiological and wellbeing states.

3. Explore several supervised learning methodologies to accurately classify and infer wellbeing based on collected environmental, physiological and behaviour classified separately and together to ascertain the impact throughout a range of urban environments.

4. Demonstrate for the first time the feasibility of acquiring time-series data in the form of trajectories by adding environmental, physiological and momentary mental self-labelled wellbeing as semantics while applying the concept of episodes to classify wellbeing states across different urban environments.

4.2 Semantic Trajectories to assess the Impact at the Point-of-Exposure

Trajectory modelling is becoming increasingly common to explore spatio-temporal patterns in mobility and the assessment movement of multiple objects in order to assess impact at the 'point-of-exposure'. Additionally, as identified in the literature semantic enrichment provides many benefits in being able to add greater detail to a trajectory and be able to understand the impact of mobility directly. The main focus in this study is to explore the use of a personalised individual trajectory with the addition of semantics to store internal characteristics (such as physiological HR, HRV, EDA) and external context (such as the environmental factors) to observe the impact to mental wellbeing, specifically at the point of exposure. This investigation will link together trajectories to the actual environment where exposure takes place which is an important step to quantify the impact of the environment on time-series data in moving objects.

The complexity and increased use of spatio-temporal data have helped in leading to achieve a better understanding into the processes of a movement. Several studies have previously shown the capabilities of this data for describing movement phenomena graphically [158], [32]. Many authors have proposed solutions for the task of gaining a better understanding on the impacts of movement. However, the gap remains in understanding of how other factors such as environmental and physiological could have a direct impact towards mental wellbeing and mobility.

4.2.1 Experimental Setup

The work in this section uses Dataset 1 at Section 3.4.1 which contains the multi-sensor fusion approach embodying the trajectory (x,y and timestamp), semantic as environmental and physiological data, in addition to the individual episodes as self-reported mental wellbeing label. The fused trajectory dataset is made up of 3,953 samples which include the following variables:

- Urban environmental attributes and air quality: *Particulate Matter (PM1.0), (PM2.5), (PM10), Oxidised, Nitrogen Dioxide (NO2), Reduced, Ammonia*

(NH3) and Noise

- Body physiological reactions including: Heart rate (HR), Electrodermal activity (*EDA*), Heart Rate variability (*HRV*)
- People count via wireless proximity detection
- Individuals' perceived responses: Self-reported valence (Refer to Section 3.3.1)

4.2.1.1 Identification of Trajectory Episodes

This study considers a different approach to identifying episodes from semantic trajectory data. The general idea and objective of this work is to explore using the self-labelled emotion to construct several individual episodes to divide up the human trajectory. At each episode point, along with the raw trajectory (x, y, timestamp) contains a self-labelled wellbeing-level based on an emoji and several semantic elements to enrich the episode, in the form of environmental (Air Quality) and physiological factors. In this case, Figure 4.1, demonstrates this visually, depicting an example trajectory of one participant's route with a red outline dot representing a single wellbeing change along the route.

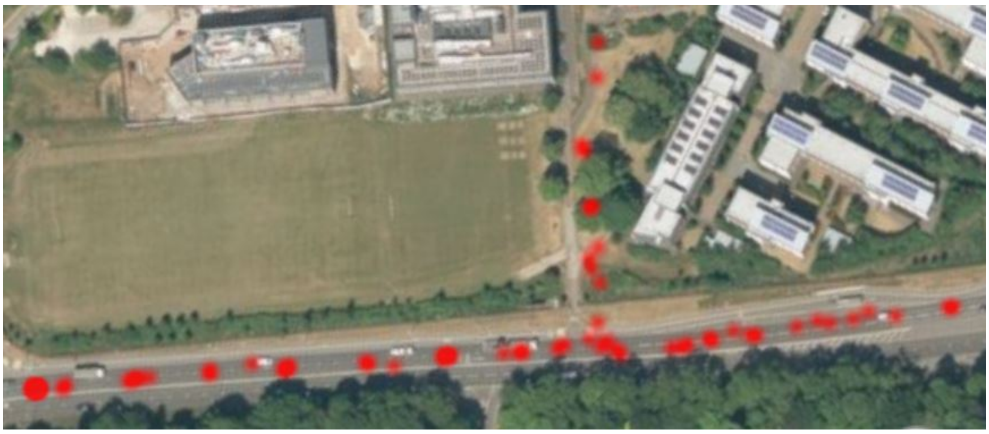


Figure 4.1: Wellbeing overlay from one participant's trajectory. Each dot represents a change in wellbeing, tagged with a trajectory (x, y) containing environmental and physiological factors.

To further envisage this approach Figure 4.2, demonstrates a single individual's trajectory with the addition of semantics made up of environmental, physiological and self-report labelled wellbeing as the y-axis. This is put in context of several episodes (e.g. 913 at Figure 4.2) which is made from an individual's wellbeing change at each point across a whole trajectory (x-axis). The arrangement of information in this way shows the usefulness of a representation that allows multimodal data to be combined.

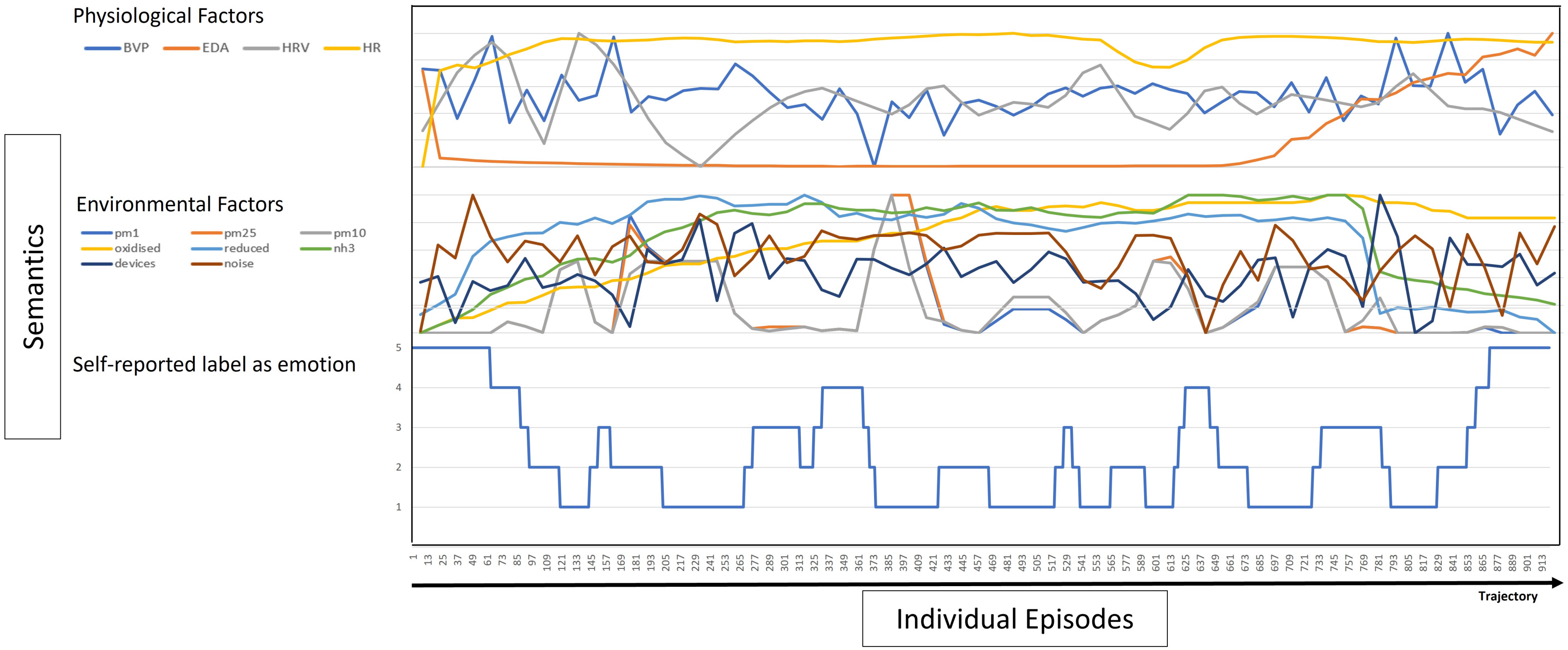


Figure 4.2: An individual trajectory combining semantic features; Environmental, physiological factors depicting each single episode made up of the changes in emotions across the entire trajectory.

4.2.2 Statistical Analysis

To begin to unravel the relationship between the different environmental and physiological variables and the impact each one has on another, Pearson's R Correlation Coefficient Matrix is employed. The correlation coefficient helps in understanding the strength of the linear relationship between two variables [144]. Figure 4.3 presents the Correlation Matrix depicting the environmental, physiological and self-report wellbeing variables.

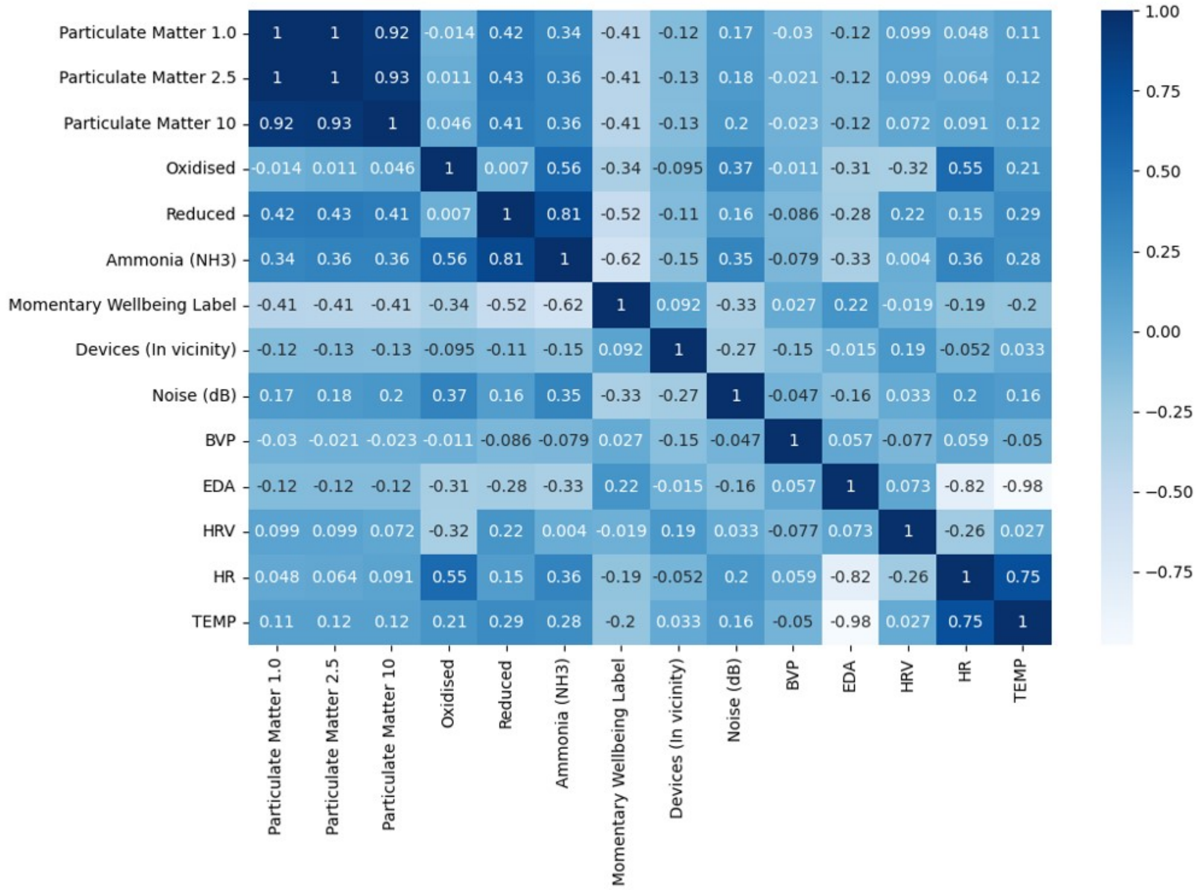


Figure 4.3: Pearson's Correlation Coefficient Matrix made up using the collected environmental, physiological and self-labelled wellbeing.

In particular, momentary self-report wellbeing label correlates negatively with Particulate Matter 1.0, 2.5, 10, Oxidised, Reduced and Ammonia demonstrating that higher levels is causing a negative wellbeing. Additionally, it can be noted that EDA correlates positively with the wellbeing label (0.22) and negatively with Particulate Matter 1.0, 2.5 and 10. Furthermore, HR suggests to correlate with Oxidising Gases (0.55), body temperature (0.75) ammonia (0.36) and noise (0.20).

4.2.3 Classification of Emotion Change using Semantic Trajectory Episodes

The use of machine learning has been explored to help classify the changes in wellbeing based on the five self-reported states of momentary wellbeing label using the environmental pollutants and physiological data. These labels segment the trajectories into individual episodes of the environment.

4.2.4 Dynamic Time Warping

Dynamic Time Warping (DTW) has been shown to be an effective algorithm in many areas, particularly in classification of time series data [192] in being able to measure the similarity between two temporal sequences. In this case, for classifying the episodes of wellbeing change, the similarity distance between the different episodes of the trajectories was first measured using DTW. The semantic trajectories containing the individual wellbeing episodes from 6 participants (3,953 samples) have been extracted using DTW, which were then used to train a number of machine learning classifiers to classify mental wellbeing. The input data used to train was made up of the semantic trajectories which includes: Particulate Matter (1.0, 2.5 & 10), Ammonia, Noise, Oxidised and Reducing Gases, HR, HRV, EDA, BVP, Body Temperature and labelled wellbeing in the form of self-report. The model was trained over 20 epochs with a batch size of 128 and tested using a 10-fold cross validation with a 20% test split. To segment the collected data an overlapping sliding window strategy was adopted with a window size of 8 and a step of 4 chosen experimentally to improve accuracy.

4.2.4.1 Classification Results

The F1-Scores for each of the classification models trained using the combined 12 features made up of environmental and physiological data are depicted at Figure 4.4. Using DTW with a KNN classifier can be seen to be the best performing model with accuracy as 0.88 (F1-Score), outperforming the other statistical models by 0.10 (DT) and 0.37 (RF) respectively.

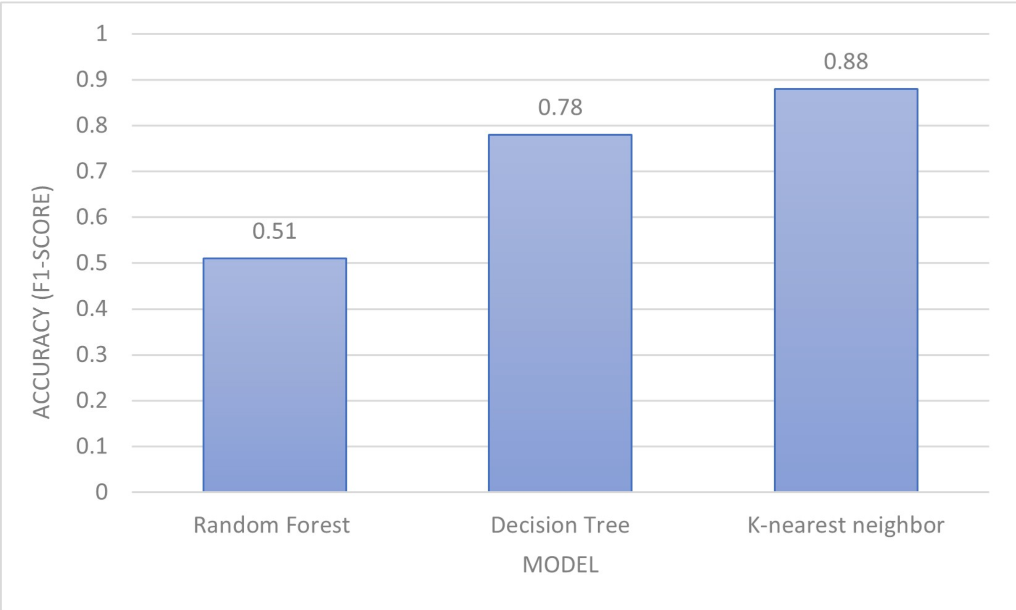


Figure 4.4: Comparison of classification models used with combined environmental and physiological data to classify wellbeing across episodes.

To delve deeper into the relationship between using environmental and physiological data and why it can be combined to predict wellbeing, the three classifiers (KNN, DT and RF) were used on the separate data modalities (environmental and physiological) in the dataset as depicted at Figure 4.5.

When the DTW model was trained using just the environmental data alone, a KNN classifier achieved the highest at F1-Score 0.84 indicating that wellbeing can be inferred using environmental data and physiological data can be used to infer with an F1-Score of 0.76. These results build on previous work as the use of physiological data has been shown in other studies to accurately classify wellbeing, due to its high correlation with the sympathetic nervous system [197]. From

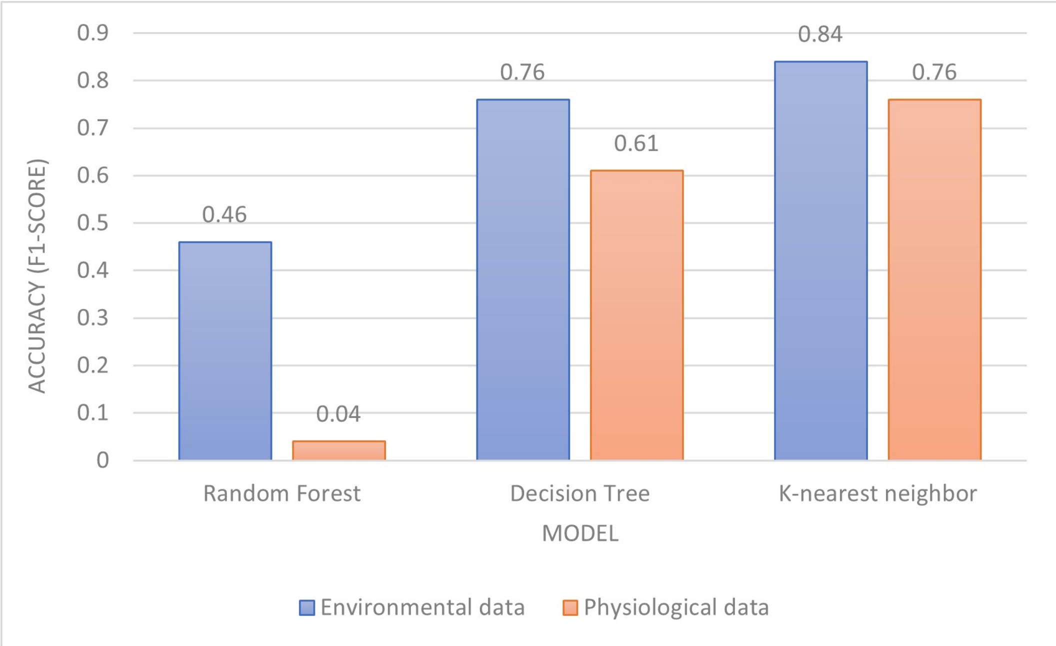


Figure 4.5: Comparison of F1-Scores for both environmental and physiological data separated.

these results, it is particularly interesting to note that environmental combined with physiological data outperformed the model trained using environmental data alone, demonstrating the benefits of the environmental air quality data collected. This suggests that environmental data is a robust source for wellbeing classification alone.

Moreover, the results also show that using a Decision Tree classifier with the separate environmental and physiological data achieved 0.76 and 0.61 respectively. Combining the training of environmental and physiological data together and trained using DTW and classified using Decision Trees, could be used to infer emotion achieving an F1-Score of 0.78. It is worth noting that in both classifying environmental data using DT and a KNN, wellbeing could be inferred at a higher level than physiological data. This further highlights the importance air quality data can have in wellbeing classification.

4.3 DigitalExposome Aggregated Study

This section delves deeper into the exploration and discussion of the proposed methodology that focuses on quantifying data related to environmental, physiological, and mental wellbeing. The approach employed here emphasises the collection and analysis of data in an aggregated manner, drawing insights from a group of users.

4.3.1 Experimental Setup

The work in this section uses Dataset 2 (see Section 3.4.2) which contains the multi-sensor fusion approach embodying environmental, physiological and self-report mental wellbeing data. Following sampling of the sensor data, the resulting dataset includes:

- Urban environmental attributes and air quality: *Particulate Matter (PM1.0), (PM2.5), (PM10), Oxidised, Reduced, Ammonia (NH3) and Environmental Noise*
- Body physiological reactions including: Heart rate (HR), Electrodermal activity (*EDA*), Heart Rate variability (*HRV*), Body temperature (*TEMP*)
- Body Movement via Accelerometer
- People count via wireless proximity detection
- Individuals' perceived responses: Self-report Wellbeing (Refer to Section 3.3.3)

4.4 Results

In this section, the experimental results are presented, explored and discussed. Section 4.4.1 employs mathematical and statistical approaches for the exploratory analysis stage including variable Correlations, *PCA* factor maps, variable importance and Pearson's R Correlation Coefficient to measure the association between two categorical variables. Section 4.4.2 explores the use of Multi-Variant Regression to understand the variable importance between dependent and independent factors. Finally, Section 4.5, presents the classification results of infer mental wellbeing on both environmental and physiological factors.

4.4.1 Statistical Factorial and Variable Importance Analysis

To explore the correlations between environmental, physiological and self-reported wellbeing mathematical and statistical approaches have been explored. These include descriptive statistics, variable correlations, *Principal component analysis (PCA)* factor maps, variable importance and Pearson's R Correlation Coefficient to measure the association between two categorical variables. Table 4.1, gives an overview of the descriptive statistics including the mean, mode, Min, 1st, 2nd, 3rd quartile, maximum value, skewness and kurtosis for the data collected during this study.

Table 4.1: Summary of the Descriptive Statistics of the collected Environment and Physiological Data.

Variables	Mean	Median	Min	1st Qu.	2nd Qu.	3rd Qu.	Max	Skewness	Kurtosis
BVP	-1.5	0	-1050	-36.02	0	34.714	1075	-0.019	11.022
EDA	0.35	0.175	0	0.12	0.18	0.26	4.54	3.92	15.84
HR	100.2	100.6	0.71	91.23	100.64	108.96	174	0.13	1.66
HRV	0.46	0.55	0	0.21	0.55	0.62	1.34	-0.52	-0.54
NH3	878.6	686	15	509	686	1064	3794	1.3	1.4
Noise	97.4	96.4	47.2	94.5	96.37	100.3	140.3	-1.73	19.95
Oxidised Gases	38.1	38	2	30	38	42.3	88	0.08	0.21
PM 1.0	4.36	3	0	0	3	7	65	3.21	18.6
PM 2.5	5.8	0	0	3	3	9	65	2	7.1
PM 10	7.3	3	0	0	4	12	65	1.88	4.4
Reduced Gases	453	509	47	341	509	548	1201	-1.4	0.5

A correlation matrix for the study data depicted at Figure 4.6 aims to demonstrate the relationship between variables. From this, it is clear to see that some variables are highly correlated together. Through analysing the individual cells *HRV* is shown to correlate with PM10 (0.118) and NH3 (0.211). In addition, *EDA* demonstrates a correlation with PM10(0.189), oxidised (0.213) and reduced gases (0.15), and NH3 (0.198).

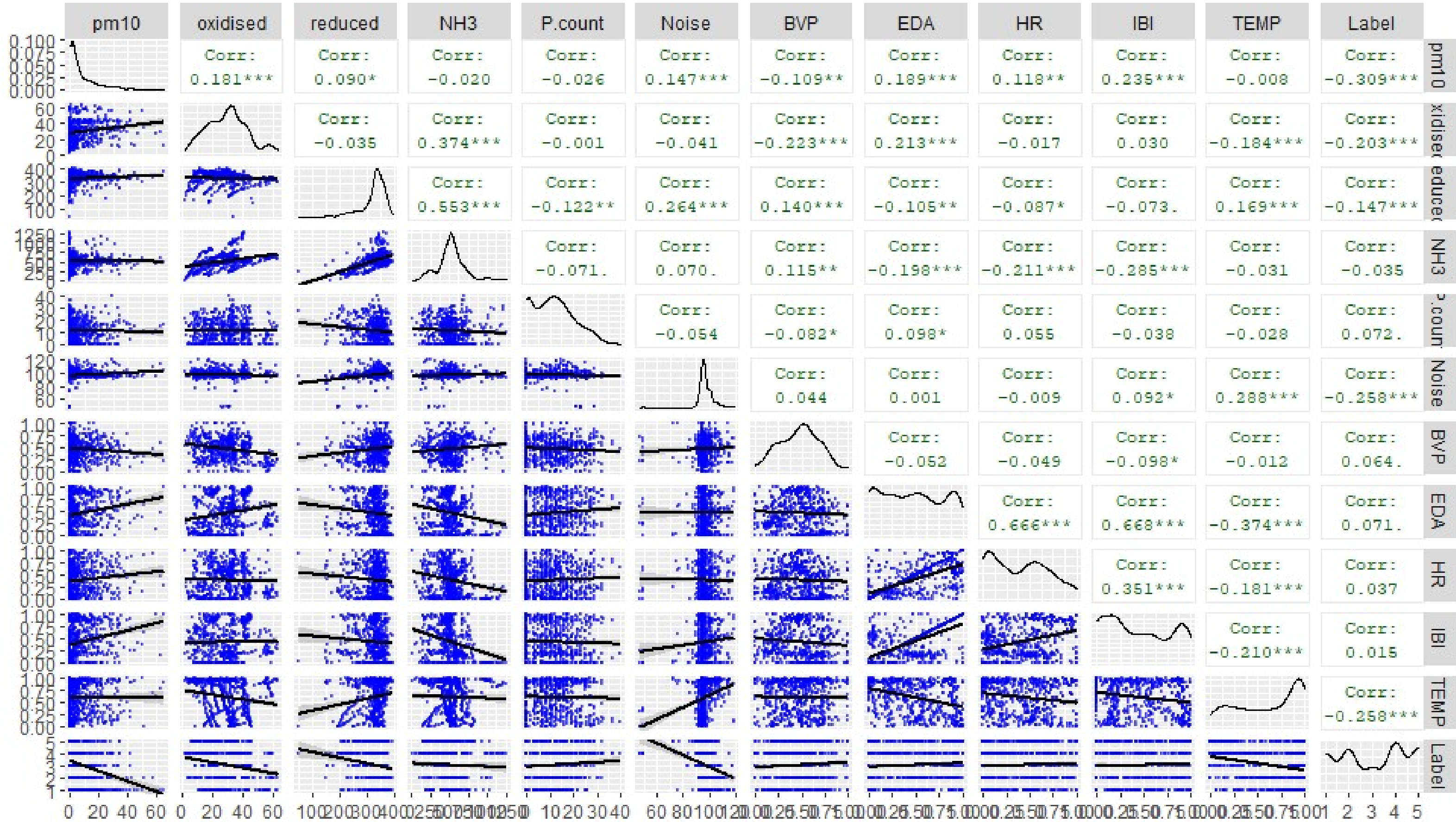


Figure 4.6: Correlation Matrix for Environmental, Physiological and self-reported label.

Principal component analysis factor maps are an effective method for large datasets, to help understand the relational impact between different variables, with reducing information loss [99]. The use of PCA maps offers a visual approach to presenting data and examining the relationships among variables [102]. This is achieved by giving a view of all the plotted variables projected on to a plane, spanned by the first two principle components. This method demonstrates the structural relationship between the different variables. Figure 4.7 (A & B), present the captured environmental and physiological variables depicted on a *PCA* map.

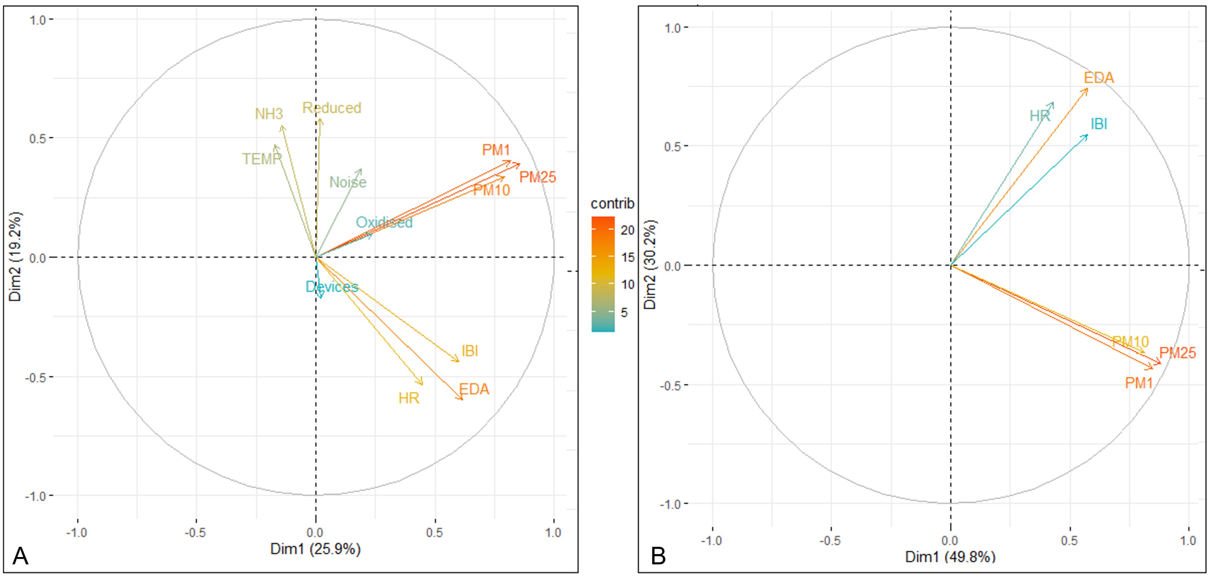


Figure 4.7: PCA Analysis - (A) Variance between the different variables, (B) Variance between the different variables without EDA.

It is worth noting, that most of the body attributes *EDA, HR and HRV* are all at the top of the figure, whilst, the environmental variables *PM1, PM2.5, PM10* and Reducing gases are located in the middle. From the diagram (A), there is Dim1 25.9% and Dim2 19.2%, resulting in 45.1% in total variance across the environmental and physiological variables. It is worth noting that the most important, (or, contributing) variables are highlighted using the colour gradient (i.e. darker colours indicate higher contributing factor).

At the second PCA diagram (B), depicts the most important variables as identified at diagram (A) including *PM1,PM10, PM2.5,EDA, HR and IBI* with the least contribution variables discounted. With this PCA diagram, it is noticeable that the total variance increases to 80%, comprising of 49.8% from Dim1 and 30.2% from Dim2. Higher increased of variance in other studies has shown the stronger association between variables [189], as evident in PCA diagram (B). The close grouping and proximity of the independent variables suggests that *HRV, HR and PM10* are correlated and that *HRV, HR, PM2.5* can also be correlated. Analysing these early findings indicates that lower the HRV and higher HR is correlated to a higher level of air pollution within the environment.

Figure 4.8 demonstrates the variable importance by depicting the impact of wellbeing against levels of PM2.5 within the environment. The bars on the chart are associated with how many times a particular user would label how they were feeling (self-report wellbeing) whilst walking around the environment.

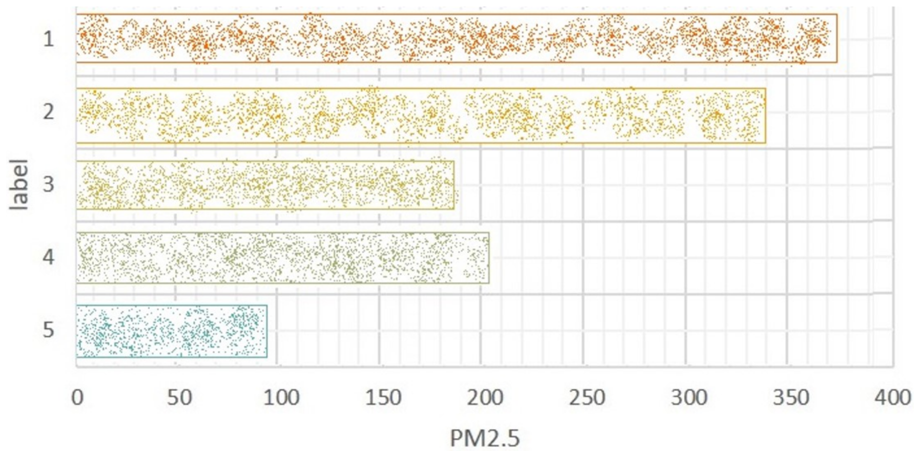


Figure 4.8: Depicts the relationship between the self-reported Participant's wellbeing (Label) and PM2.5.

The results of this indicate that high levels of PM2.5 are more commonly associated with a negative wellbeing, shown by participants choosing (1- Very negative wellbeing) on the device. Whereas less association is apparent where participants labelled '5' (very positive wellbeing), the levels of PM2.5 were much lower. This early analysis on the collected sensor data helps to further understand and demonstrates the impact of pollution on mental wellbeing.

4.4.2 Multi-Variant Regression Analysis

Multi-Variant Regression offers the opportunity to see the importance and impact each variable has on the other. In this work, the variable dependency on two different modalities using Multivariate Regression and Principle Component Analysis (PCA). For each of the dependent variables (physiological data), Multiple Linear Regression was utilised to compare them against the independent variables (environmental data). The aim of this is to see which dependent variable can be predicted from using the environmental data as independent variables.

4.4.2.1 Multiple Regression Model for EDA

Firstly, a multiple linear regression module for EDA has been used to understand the impact of this physiological on-body sensor to the other independent environmental variables including NH3, Noise, PM1, PM2.5, PM10 and Reduced. Table 4.2, shows the multiple regression results for EDA:

Table 4.2: Multiple Regression Analysis between EDA and Environmental variables.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-0.02381894	0.02225209	-1.07041	0.284452
nh3	0.000291595	$1.14608E-05$	25.44285	$1.5E-139$
noise	0.004050864	0.000221511	18.2874	$7.92E-74$
oxidised	-0.00590754	0.000143065	-41.2928	0
pm1	-0.00768185	0.00081832	-9.38735	$7.11E-21$
pm10	0.000939923	0.000285371	3.29369	0.000991
pm25	0.003698711	0.000800215	4.622149	$3.83E-06$
reduced	-0.00058528	$4.59985E-05$	-12.7239	$7.06E-37$

At Table 4.2, the coefficients demonstrate that environmental variables (NH3, Noise, PM10 and PM2.5) involve an increase in EDA. A negative coefficient shows that as EDA increases the remaining environmental variables decrease showing that there is a less of association between them. In addition, a negative (-) t-stat value for each environmental variable depicts a negative impact on the variable of EDA. Whereas a positive value indicates an association between the environmental variable and EDA. The data in Table 4.2 was then evaluated using a regression curve shown in Figure 4.9. This shows the relationship between the calculated residual values verses the fitted values shown at (A) and (B) respectively.

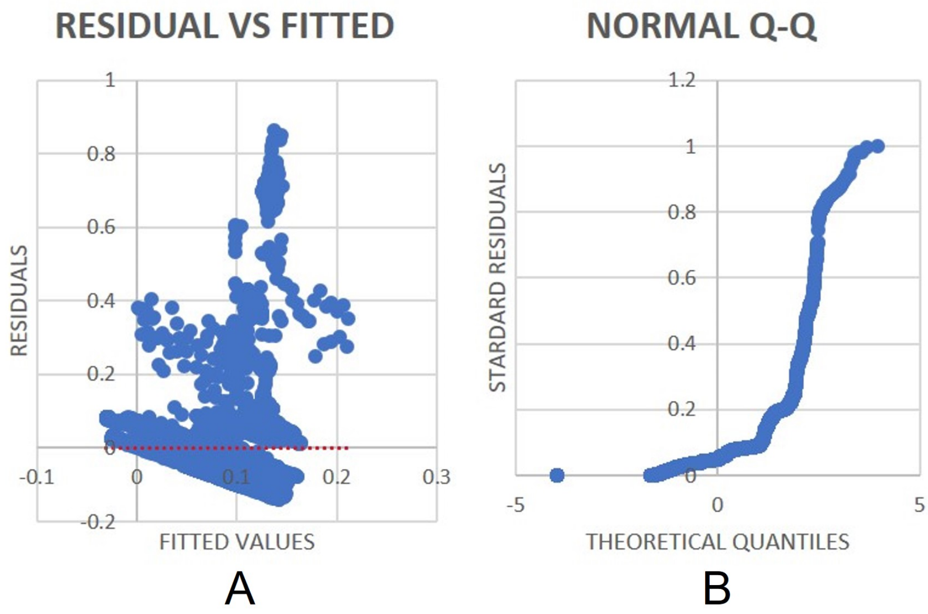


Figure 4.9: EDA Regression: Residuals Vs Fitted values curve (A) and Q-Q Plot (B).

Figure 4.9 depicts the graphs of Residuals Vs Fitted (A) and a normal Q-Q plot (B) for EDA by using bi-modal data. The aim of a Residual Vs Fitted graph is to ascertain whether linearity holds which is normally indicated by the mean of the residual values being close to 0. In the case of (A), this is shown by the red dotted line being close to 0. On the other hand, the Q-Q plot (B) is used to in order to fit a linear regression model. In many Q-Q plots, the data on the graph takes the shape of a twist like seen in this plot [102], [191]. This plot is presenting a symmetric distribution with 'fat-tails', otherwise known where the ends of the line curve. The lower part of the plot is almost linear, suggesting a normal distribution in relation to one mode of data distribution. In addition, the upper part of the Q-Q plot again suggests linear, showing an approximate distribution. The steep line between the upper and lower curve is steeper than the line y = x which suggests the distribution plotted on the vertical axis is more dispersed than the distribution plotted on the horizontal axis. The implication to this is that the data points are normally distributed.

4.4.2.2 Multiple Regression Model for HR

Below presents the multiple linear regression model for HR using the other independent variables (environmental). This includes; NH3, Noise, PM1, PM2.5, PM10 and Reduced. Table 4.3, shows the multiple regression results for HR:

Table 4.3: Multiple Regression Analysis between HR and Environmental variables.

	Coefficients	Standard Error	t Stat	P-value
Intercept	128.8420806	1.721454748	74.84488381	0
nh3	0.007322924	0.000886623	8.25933903	$1.59864\text{E}-16$
noise	-0.083286817	0.017136432	-4.860219147	$1.1856\text{E}-06$
Oxidised	-0.051833849	0.011067698	-4.683344891	$2.84951\text{E}-06$
pm1	0.118538171	0.063306454	1.872450026	0.061165731
pm10	0.112184632	0.022076708	5.081583292	$3.79248\text{E}-07$
pm25	-0.232804742	0.061905795	-3.760629225	0.000170194
reduced	-0.072042617	0.003558515	-20.24513451	$8.12597\text{E}-90$

These initial findings are in agreement with previous research that shows PM2.5 can directly impact HR [170]. In addition, research has shown how differing levels of irregular environmental noise can impact a regular heart-beat. In particular, recent studies exploring this find that noise levels between 55 and 75 Decibels (dB) are linked to a higher risk of developing heart related diseases [147].

Figure 4.10 depicts the Residual VS fitted values and normal Q-Q Plot as shown at A and B respectively. Similar to the EDA Q-Q plot, HR Q-Q plot shown in Figure 4.10 (right) demonstrates a twist at either end of the plot. In addition the data shows a clear bi-modal distribution. The lower part of the plot is almost linear suggesting an approximate normal distribution. The line in the middle of the upper and lower parts follows a more linear (y=x) line, meaning that the distribution is less dispersed. It is worth noting that there were three outliers for HR distribution due to erroneous sensor readings.

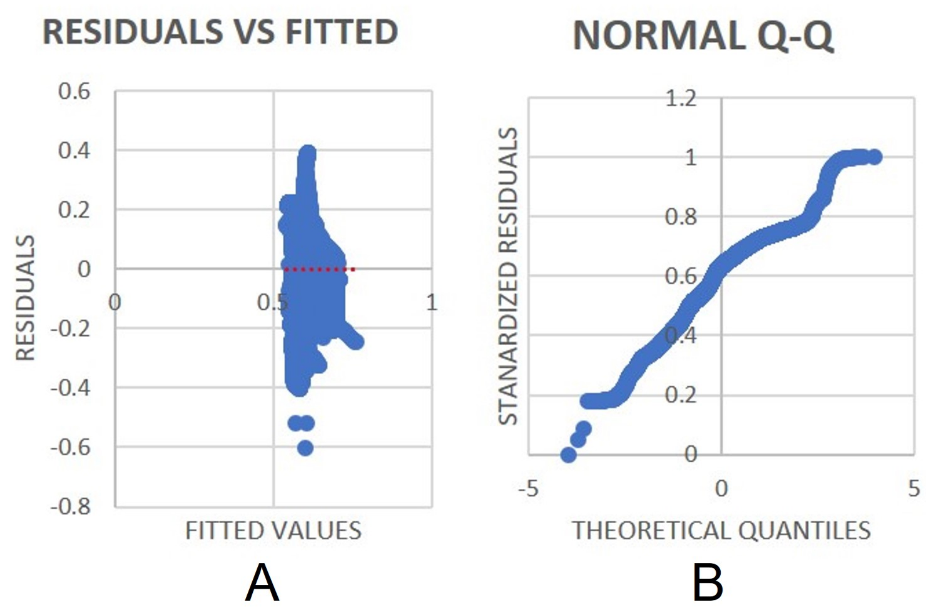


Figure 4.10: HR Regression: Residuals VS Fitted values curve (A) and Q-Q Plot (B).

4.5 Classification of Aggregated Mental Wellbeing

To explore the relationship deep learning networks and machine learning classification techniques have been incorporated to classify the five self-reported states of wellbeing using the environmental pollution and physiological data from the 40 participants who successfully labelled their wellbeing. There were 3 participants whose data was removed prior to the classification due to issues around the self-recorded label and sampling of the environmental variables not recording correctly.

4.5.1 Deep Belief Network (*DBN*) Analysis

Firstly, a preliminary study involving 11 participants took part in the experiment collecting environmental and physiological data that has been used to train an unsupervised one dimensional deep belief network to classify the obtained mental wellbeing data. The input data into the model includes: Particulate Matter (1.0, 2.5 & 10), Ammonia, Noise, Oxidised and Reducing Gases, HR, HRV, EDA, BVP, Body Temperature and labelled wellbeing in the form of self-report. The model was trained over 20 epochs with a batch size of 128 and tested using a 10-fold cross validation with a 30% test split. The raw input data was firstly divided into segments of fixed lengths. To segment the collected data an overlapping sliding window strategy was adopted with a window size of 8 and a step of 4 chosen experimentally to improve accuracy.

4.5.1.1 Classification Results using a DBN

The extracted features from the *DBN* were combined with Random Forest, Support Vector Machine (*SVM*), Decision Tree, Gaussian Naive Bayes, Logistic Regression and Gradient Boosted supervised machine learning models to classify the five self-reported states of wellbeing using the input data of environmental air quality pollution and physiological. These machine learning models were selected due to their high popularity and were also tested using only common statistical features: mean, median, max, min, max-min, standard deviation and quartiles [128]. Additionally, a Convolutional Neural Network (*CNN*) has been trained using the same raw data to enable comparison with the *DBN* models for the 11 participants' data.

Figure 4.11 shows the accuracy for each of the classification models trained using standard statistical features and features extracted using the *DBN*. The results demonstrate that the models trained using features extracted from the *DBN* outperformed the models trained with statistical features for three out of the six classifiers and achieved on average 3.2% higher accuracy. Random Forest combined with the *DBN* was the best performing model achieving 80.83% accuracy, outperforming all statistical models and the *CNN* which is frequently used for wellbeing classification by 5.54%. To explore the impact air quality pollution

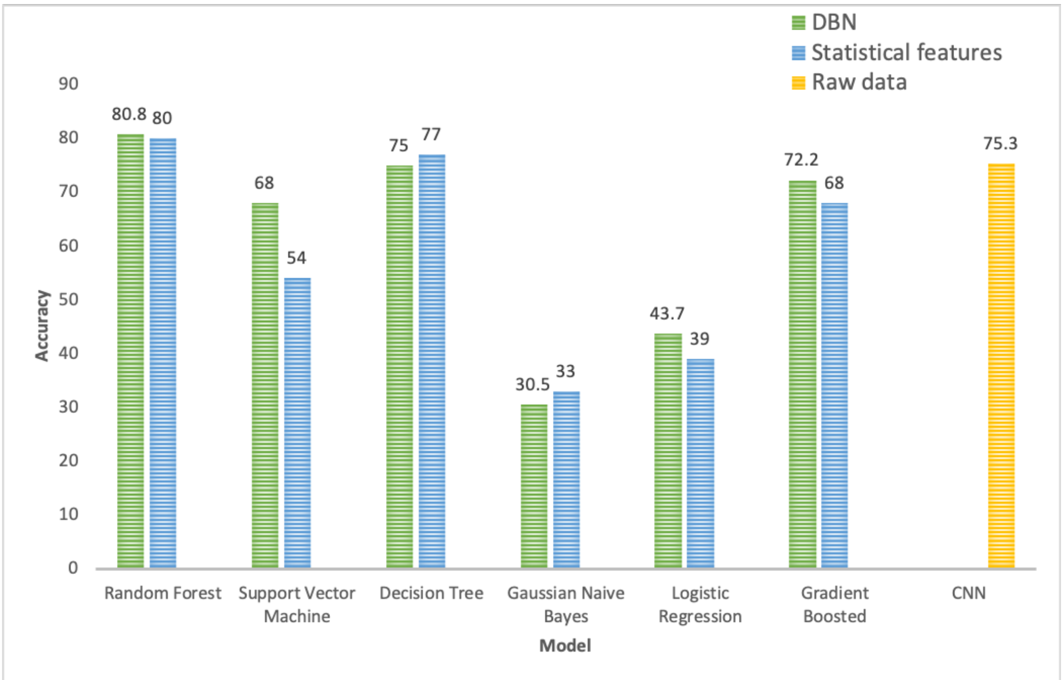


Figure 4.11: Comparison of classification models trained using statistical features and features extracted from DBN.

has on wellbeing the best performing model (Random Forest) in addition to the *CNN* were individually trained using only the pollution and physiological data as shown in Figure 4.12.

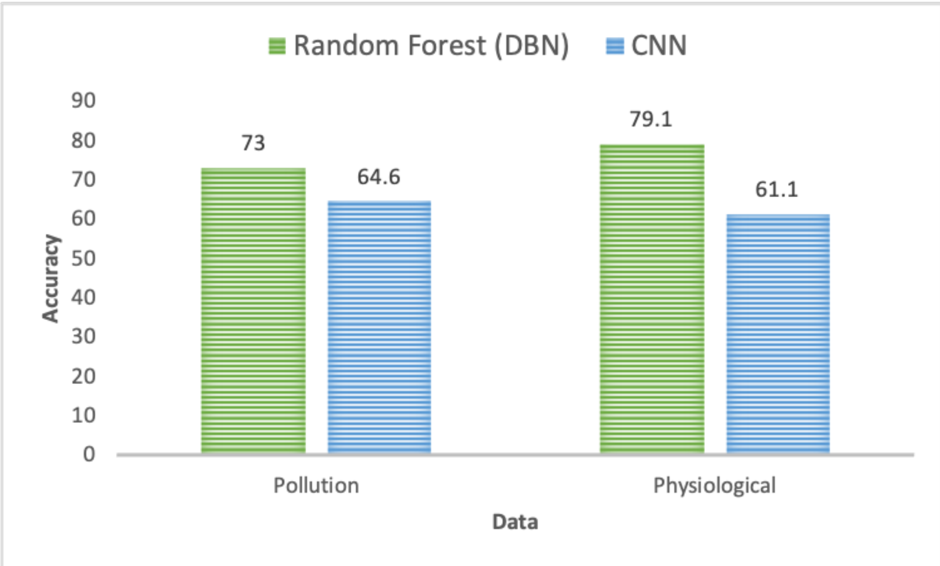


Figure 4.12: Comparison of Random Forest combined with DBN and CNN when trained using only pollution or physiological data.

The results show that wellbeing can be inferred using pollution data alone with 73% accuracy while wellbeing can be inferred from physiological data with 79.1% accuracy. It was expected that psychological would accurately classify wellbeing due to its high correlation with the sympathetic nervous system [197]. From the results in this smaller study of 11 participants it is surprising that the environmental air quality pollution data when combined with physiological data outperforms the model trained using environmental data alone. This situation implies that the environmental air quality data can play a crucial role towards impacting mental wellbeing in urban environments.

4.5.2 Convolutional Neural Network (*CNN*) Analysis

Secondly, to compare model performance, environmental and physiological data from 40 participants have been used to train a 1D-CNN to classify the five mental wellbeing states. The input data variables for the model was the same as stated in Section 4.5.1. The model was trained over 20 epochs with a batch size of 128 and tested using a 10-fold cross validation with a 30% test split. The network architecture of consists of 2 1-dimensional convolutional layers (64 and 32 neurons respectively) followed by a dropout layer with a rate of 0.5 to prevent over fitting before the 'softmax' activation function. Batch normalisation has been utilised within the network to normalise the inputs of each layer followed finally by a fully connection layer. The learning rate has been set at 0.001. The raw input data was firstly divided into segments of fixed lengths. To segment the collected data an overlapping sliding window strategy was adopted with a window size of 8 and a step of 4 chosen experimentally to improve accuracy. Adam was selected as the optimiser, although others were trialled in order to try and improve the accuracy.

4.5.2.1 Classification Results using a 1D-CNN

The extracted features from the *CNN* were combined with Random Forest, Support Vector Machine (*SVM*), Decision Tree, Gaussian Naive Bayes, Logistic Regression and Gradient Boosted supervised machine learning models to classify the five self-reported states of wellbeing using the environmental air quality pollution (*PM1, PM2.5, PM10, Oxidised, Reduced, NH3 and Noise*) and physiological (*BVP, EDA, HR, HRV* and body temperature) data. These machine learning models were selected due to their high popularity [128].

Figure 4.13 presents the F1-scores for each of the classification models trained using standard statistical features. The classifier Random Forest was the best performing model achieving an F1-score of 0.76, outperforming the other statistical models by 0.09 and the CNN which is frequently used for wellbeing classification by 0.13. To explore the environmental air quality impact on mental wellbeing, the best performing classifier (Random Forest) was trained using the environmental and physiological data independently, as shown in Figure 4.14. The results indicate that wellbeing can be inferred using environmental data alone, achieving

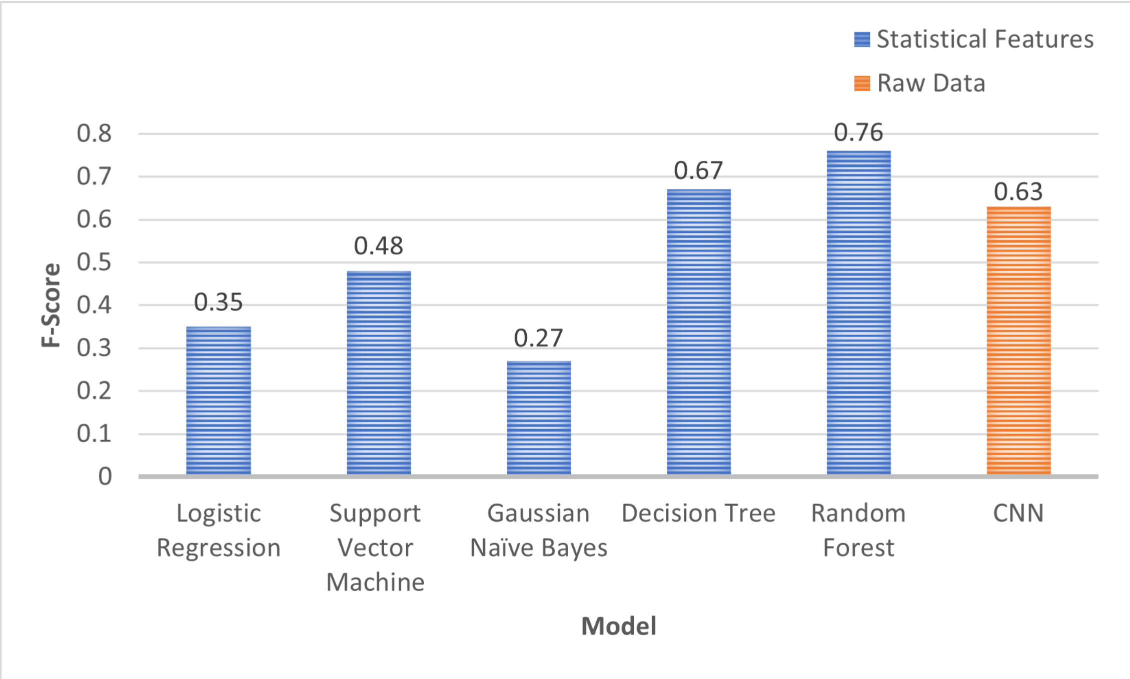


Figure 4.13: Comparison of classification models trained using statistical features and raw data from a CNN.

an F1-score of 0.67 while well-being can be inferred from physiological data with 0.61. As previous studies have shown, physiological data should be expected to accurately classify well-being due to its high correlation with the sympathetic nervous system [197]. However, similar to the first study (Section 4.5.1.1), it is interesting to note that air quality pollution data when combined with physiological data outperformed the model trained using the environmental data alone, demonstrating the benefits of this data in classifying wellbeing.

Analysing the results further in terms of precision and recall across the six different models for classification the scores were very similar in values. In particular, Logistic Regression and Support Vector Machine scored very low on precision and recall (0.44 and 0.33 respectively), struggling at predicting the middle of labels (3 and 4). At the two highest achieving models (Decision Trees and Random Forest) both were similar resulting in higher precision and recall values (0.70 and 0.74 respectively), with slightly lower scores of 0.33 when predicting label 2.

Sometimes there can be some variations in a dataset, which could be the result

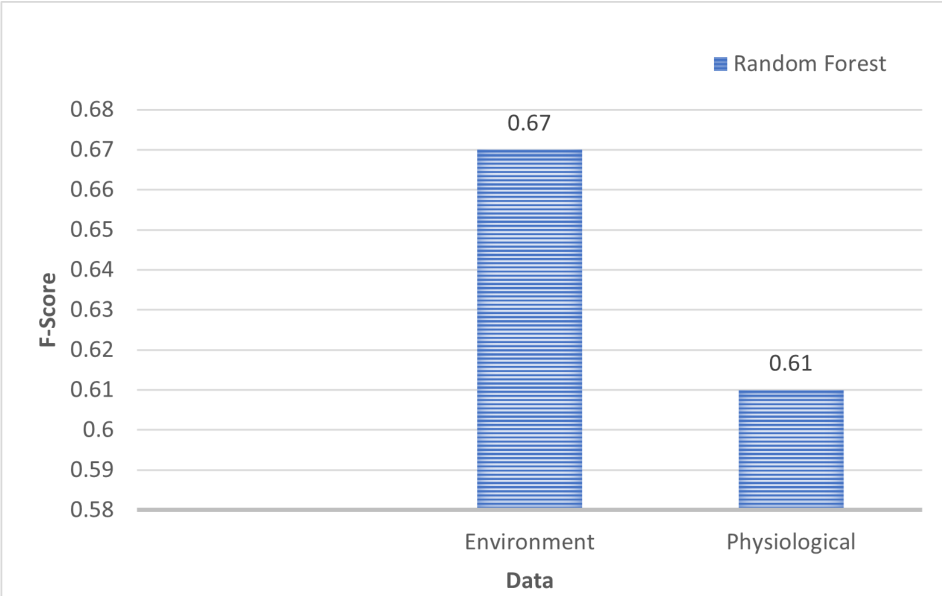


Figure 4.14: Comparison of Random Forest combined with CNN when trained using only the environmental or physiological data.

of one user labelling poorly across the experiment that can potentially lead to the training dataset being significantly impacted. In order to evaluate the training dataset in this case, two radar charts are generated from 10 users' models (CNN and CNN with a Random Forest Classifier) as depicted at Figure 4.15. A CNN alone resulted in the highest variation between users with environmental data achieving 0.45 to 0.70, physiological 0.20 to 0.50 and combined environmental and physiological achieving between 0.30 to 0.90 (F1-Scores respectively). The combination of a CNN and using a random forest classifier depicted a significantly lower variation in accuracy between users, apart from user 4. Out of the 10 users' selected, user 4 achieved a lower accuracy from environment (0.5), physiological (0.3), combined environment and physiological (0.6). However, the other 9 users were consistent in the achieved accuracy of environment (0.9), physiological 0.7 and both combined environment and physiological data (0.9). This approach has helped to understand how sometimes users' data can impact the overall training of the model dataset. Furthermore, the combination of environmental and physiological data to train with has highlighted the association between the variables to help infer wellbeing states.

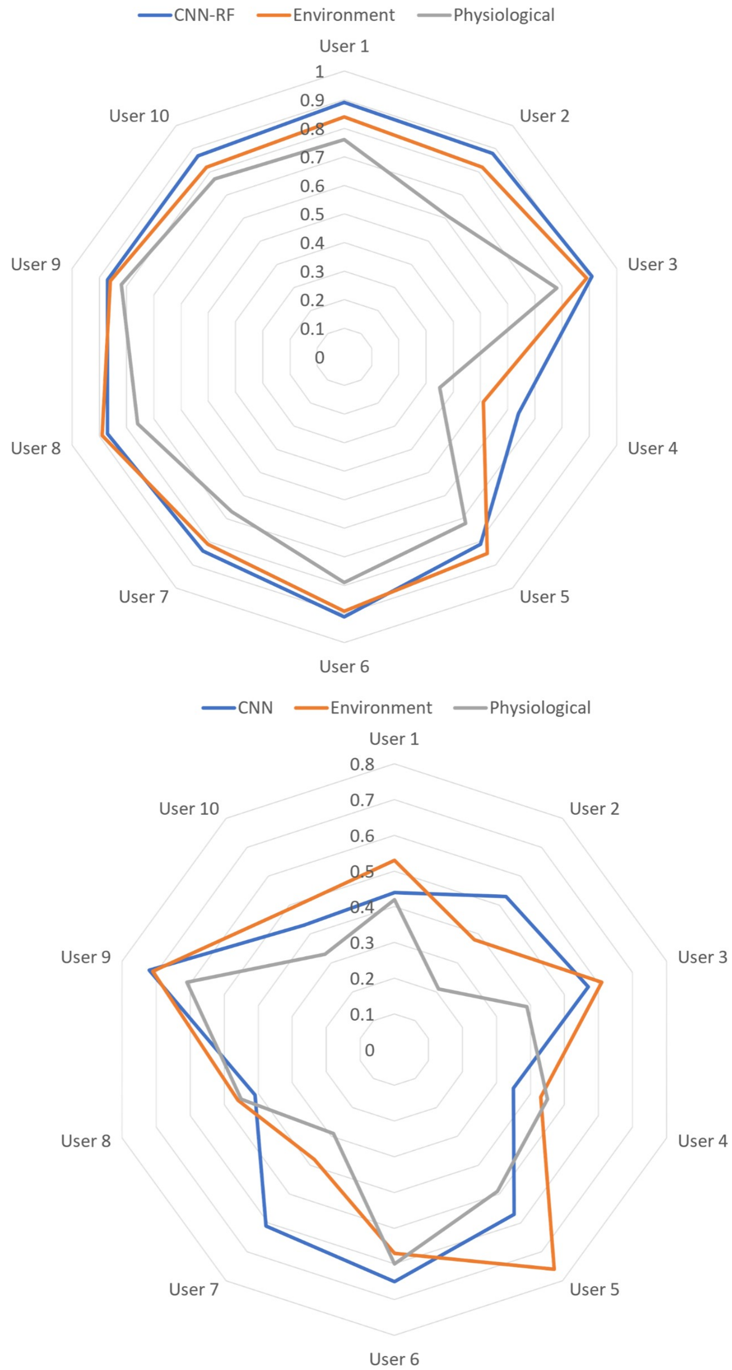


Figure 4.15: Radar charts showing the accuracy levels of three models (top) CNN, (bottom) CNN-RF, based on ten users data in ad-hoc and fused modes.

4.6 Discussion

The findings from the semantic trajectory and DigitalExposome studies reveal that the utilisation of objective sensor data enables a continuous collection and integration of real-world environmental and physiological sensor data. This approach proves beneficial in understanding the impact of our surroundings on mental wellbeing and provides insights into our interactions and behaviors across various environmental conditions. By combining these data sources, valuable knowledge can be gained about the way to perceive and engage with our environment.

The impact of various variables on mental wellbeing has been established through their interrelationship. Previous research in the field has predominantly concentrated on examining the effects of noise and ultra-violet radiation on emotions [100], as well as exploring methods for assessing individual exposure to the environment [57], particularly through mobile and environmental sensing. In contrast, the present study takes a more comprehensive approach by considering a broader range of environmental factors and their influence on physiological responses and overall wellbeing. By continually collecting and fusing real-world environmental and physiological sensor data has enabled an understanding into how as individuals we interact and behave within different environmental conditions.

The results of the principal component analysis (PCA) indicate that when all the collected variables are merged, they can effectively capture and describe the overall variability of the data. In particular, on the PCA map, the physiological sensors (*EDA, HR and HRV*) point towards a different location to the environmental variables. From the analysis conclusions can be made that a range of environmental factors (*PM1.0, PM2.5, PM10*) impact physiological changes (*HRV, HR*). Placing this work in relation to related work shows that similar work find that HR and HRV are correlated to PM2.5 [50] but little research has been explored around EDA, which this work demonstrates there is significant correlation, demonstrating the significance of this work.

The multi-variant regression models applied to heart rate (HR) data reveal a distinct bimodal distribution, with a relatively weaker positive correlation observed towards NH3 and Particulate Matter 1.0 and 2.5. Additionally, the exploratory data analysis (EDA) also exhibits a bimodal distribution, with a relatively weaker positive correlation observed towards NH3, environmental noise, and Particulate Matter 1.0 and 10.

The ability to classify the collected data presents many possibilities for the real-world inference of wellbeing using environmental air quality data. The results show that using features extracted from a *CNN* successfully improved the accuracy in which wellbeing can be inferred. Combining physiological with environmental pollution data achieved an F1-score of 0.76 compared with an F1-score of 0.61 when trained using only physiological and 0.67 when trained using only environmental pollution data. In both studies, classification of mental wellbeing using both environmental and physiological features using an aggregated approach achieved 80.8 (DBN + KNN), 0.76 (CNN + KNN).

In addition, observations can highlight the impact of focusing more directly to the individual using semantic trajectories of self-report wellbeing to create episodes, which results in a better accuracy performance at 0.88 using DTW to extract and classify using a KNN to infer wellbeing. Whilst classifying the data obtained it is interesting to note that when utilising an LSTM classifier, on average the accuracy scored was around 0.11 F1-Score. This was an unusual result in that LSTMs used in previous work has been shown to achieve a very high classification accuracy [173], [98].

By putting the results obtained in context of related works, it further highlights the impact of this work. Similar datasets have a slightly lower accuracy/ F1-Score (around 0.80) for the different task of predicting impact of environmental noise to emotion [100], [101]. In this work, the ability for pollution data to increase overall F1-score demonstrates its impact on wellbeing and shows pollution should continue to be considered as a factor that influences changes in wellbeing.

There were some limitations that were encountered during both studies, particularly with the E4 Empatica in that it was not accurately collecting participants *EDA*. While the EDA sensor worked successfully for some, for other participants no variation in *EDA* was recorded throughout the experiment. At the point of fusing the collected sensor data, both CO2 and VOC were found to have collected data for some participants but not all, resulting in its dismissal from the analysis.

4.7 Conclusion

In this chapter, the proposed concept *DigitalExposome* has demonstrated the potential of employing a multimodal mobile sensing approach and utilising data science techniques can be employed to delve deeper into the intricate connections among environmental factors such as air quality, human physiology, and their effects on mental wellbeing. To achieve this, two real-world experiments were conducted in which participants' walked around a specified route, reporting their wellbeing responses using self-report and collecting environmental, behavioural and physiological on-body sensor data.

The recent developments in mobile sensing and wearable technologies have shown in this work how trajectory studies can be enriched through the use of semantics. These can offer a greater understanding into how other factors such as physiological and the environmental factors can have an impact on mental wellbeing. Sensing technologies can help in shedding light on how people breath, feel and interact with their environment in different surroundings. This can help in offering a better security for city dwellers and creating a bond with their environments.

In the work carried out, the studies explored in this chapter have demonstrated that physiological (on-body) sensor data is directly correlated with high levels of pollution (particulate matter in particular) within the environment. The results clearly demonstrates the advantage of combining environmental and physiological data to improve the performance of the models used and as a result increase in accuracy to infer mental wellbeing. The addition of these factors led to improve classification from environmental 0.67 and physiological 0.61 to combined of 0.76 when aggregating data. On the other hand, analysing impact to an individual participant saw classification of wellbeing improve further through environment 0.84 and physiological 0.76 to a combination of 0.88.

In the future, the hope is to consider additional sensors to observe greater changes that may improve our sense of places and characterise the relationship between people and spatial settings, which in turns might influence the future design of urban spaces. Also, although the trajectory distance walked by participants was sufficient, selecting a longer route will help to further understand the changes that impacts mental wellbeing states on a trajectory throughout different episodes.

Chapter 5

Exploration of Research Applications and Real-World Interventions

5.1 Introduction

This chapter explores the many applications this research has within the real-world which aims to monitor, evaluate and provide methods to improve the interplay between the environment, physiology, and mental well-being.

Currently, there is a noticeable lack of emphasis on low-cost environmental monitoring [250], on-the-go labelling, and practical approaches to leverage smartphone technology [84] for collecting real-world data to assess wellbeing states [51]. This limited focus restricts the ability to fully understand the intricate relationship between the environment, physiology, and mental well-being in a real-world context. Existing technologies such as Urban Mind [22], Noisespy [100], ExpoApp [57] to provide a basis, although these system are limited to monitoring the environment either through using few sensors or relying on questionnaires. The potential of smartphone technology and cost-effective monitoring methods provide a valuable opportunity to gather more comprehensive data to gain insights into how the surroundings can impact physiological states. This in turn can help document a person's emotions [102], stress levels [247], mood [91], travel/

journey [84] and other physiological factors as they go about their daily lives.

The real-world interventions explored serve as a bridge between research and the application of its findings in tangible and practical ways. In particular, the following contributions have been made through carrying out the work in this thesis with each described and explained. These are as follows:

1. Visualising Space and Time: Exploring spatial and temporal patterns through visualisations.
2. Cross-platform multimodal sensing mobile application enabling comprehensive data dollection.
3. Enviro-IoT: A cost-effective solution for environmental sensing in the urban environment.

5.2 Applications

This section presents three case studies that demonstrate the practical application of this research in real-world scenarios. These provide clear examples of how the findings and methodologies from the study can be implemented to gain insights and make informed decisions on wellbeing interventions.

5.2.1 Visualising Space and Time: Exploring Spatial and Temporal Patterns

Spatial statistical visualisation is the ability to represent a geographical space to explore the characteristics, usually in terms of geo-referenced data in space [150] and offers significant opportunities to visualise space and time in new approaches. These techniques help to understand the events happening within a specific location as to where the data was taken from [154]. Some previous studies have used spatial analysis to observe the impact relationship between two variables [39], [100]. Although, at present little research has been explored in this area with the impact of noise at the forefront [100]. The current literature is lacking in terms of plotting a wider range of environmental factors and exploring their impact. Therefore, the first application of this research aims to explore ways of observing the relationship between environmental and physiological factors through visualisation techniques on the data collected in Section 3.4.2.

5.2.1.1 Heat Maps

An approach to explore patterns is dividing geographical study areas into smaller sections with the use of layers, otherwise known as heat maps. These were first introduced to visualise values in individual cells using a colour gradient [141]. Previous work of mapping sensor data in this way has shown to visualise dynamic data effectively [136], [100]. To explore the impact of air quality and human physiology data, each were extracted from the experiment at Section 4.3 and investigated further by plotting the data on the top of a map using the exact location of a user as they completed the experiment. As a result, nine heat-maps with intensity of sensors data are depicted at Figure 5.1.

Each heat-map at Figure 5.1 depict several hot-spots scattered along the busy road (right of map) and not confined to one particular area. Results from this visualisation demonstrate further that when participants are met with abnormal levels of poor air quality, particularly PM2.5, noise, reducing gases and EDA, directly impacts HRV and EDA. While heat-maps show the intensity of each variable based on GPS coordinates, the maps also indicate the real distribution of all sensor data. This can be demonstrated by the temperature colour scheme used whereby red signifies a more concentrated level.

Although the use of heat maps have significant advantages to visualise spatio-temporal data, sometimes this approach may not contain enough information or give readers the ability to draw conclusions, or to gain a clear perspective and understanding into the associated impact [181]. A different option is to further divide the geographical study area into a series of grid cells [100], however in other works it is difficult to allocate an individual cell to a single sensor reading. Moreover, it is not possible to decide on a cell size since the density of the sensor mobility traces can be of a different density distribution.

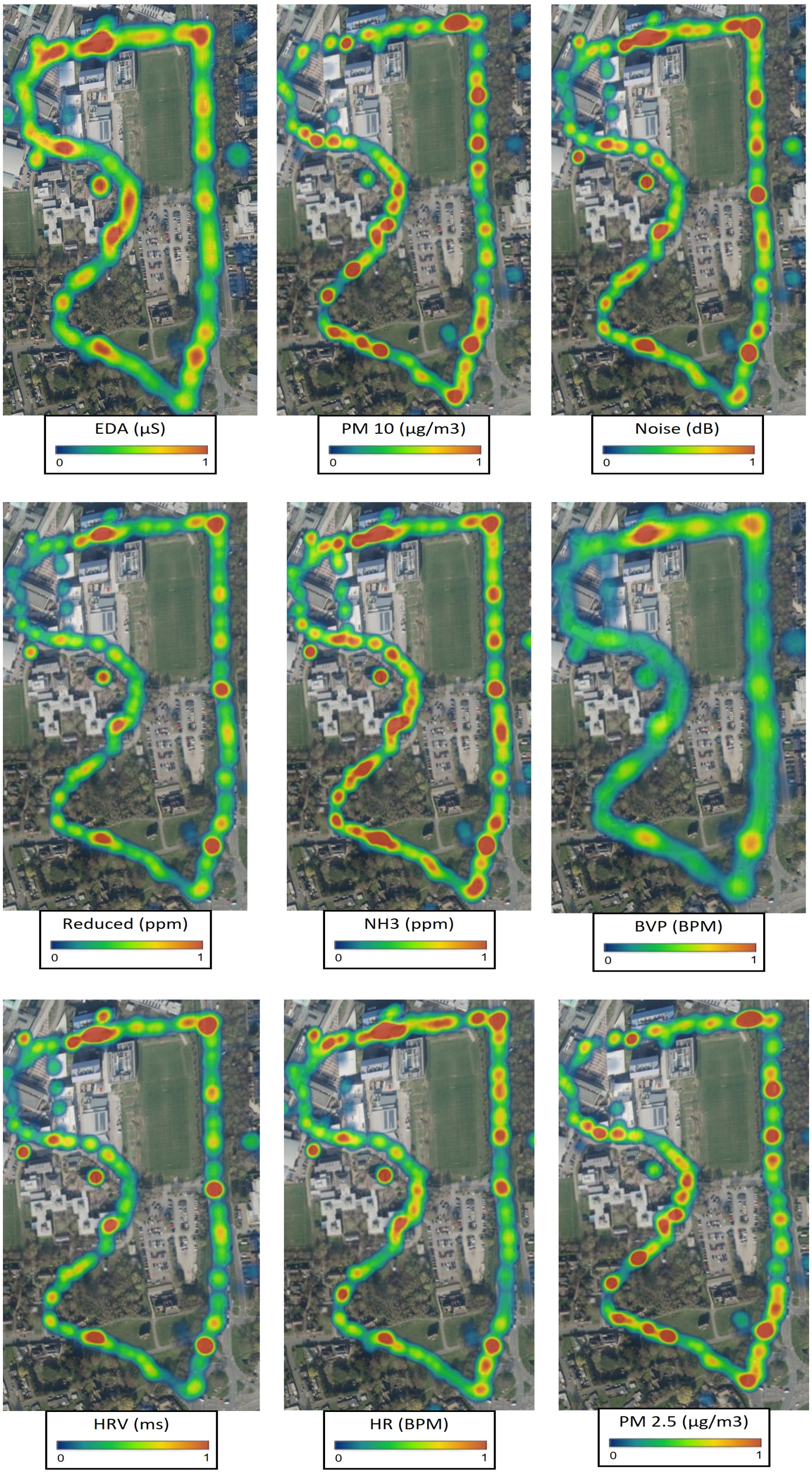


Figure 5.1: Nine heat maps using the collected environment and physiological data applied from the DigitalExposome Concept pre-specified route.

5.2.1.2 Dirichlet tessellation: Application of Voronoi

The concept of Voronoi visualisations is a computational geometry algorithm which allows the visualisation of large data sets [55]. The concept works by defining a set of polygon regions called cells, whereby the cells give an indication of the overall density of an object area of the size of the object itself [176].

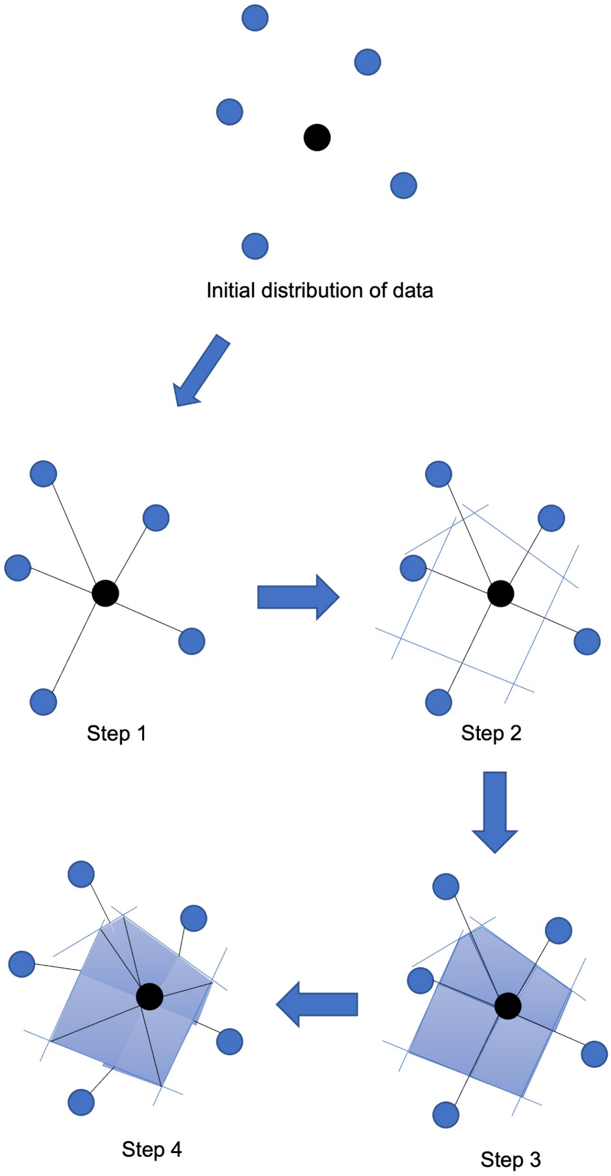


Figure 5.2: The four step process in constructing a Voronoi diagram using subsections as segments.

Voronoi is split into four stages as demonstrated at Figure 5.2 [175]. Step 1 involves drawing segments from the selected data (labelled in black) to each of the respected neighbours. Step 2, depicts the addition of bisectors for each segment. Step 3 shows the bisectors connecting from one to another to construct a space around the individual sensor data location. Finally, step 4 divides the space into joined triangles. The concept divides the space into a set of regions called Voronoi cells, including the space that is closest to the object (route location, in this case). The size of these cells gives an indication of the density of the area a certain object is in or the size of an object [176]. The cell structure also shows the Delaunay triangulation, which easily allows calculating an object's immediate set of neighbours.

The definition of a Voronoi cell is given by the Equation 5.1, where x is a planar metric space; p is the set of generator points in the metric space; and d is the distance between all points in x and a specific generator point (where the distance can be defined using any distance definition such as Euclidean, Manhattan, or road-network distance):

$$Vor_i = \{x \mid d(x, p_i) \leq d(x, p_j), j \neq i\} \tag{5.1}$$

Thus, the Voronoi diagram is composed of a collection of tessellations (i.e. polygons) defined as Vor, where:

$$Vor_i = \{Vor_1, Vor_2 ... Vor_n\} \tag{5.2}$$

The creation of a Voronoi tessellations is a dynamic procedure till all the points are represented in adjacent polygons. If sufficient number of particles did not satisfy Equation 5.2 then Voronoi gets partially filled. In this case, the data is then redistributed. By giving each polygon a class value Ci that corresponds to the sensor value collected in a particular GPS coordinate, it is then possible to divide the space into adjacent polygons with different sensor reading which are represented in colours. To provide an alternative approach to heat mapping of sensor data, Voronoi has been explored which incorporates the self-report wellbeing data depicted at Figure 5.3.

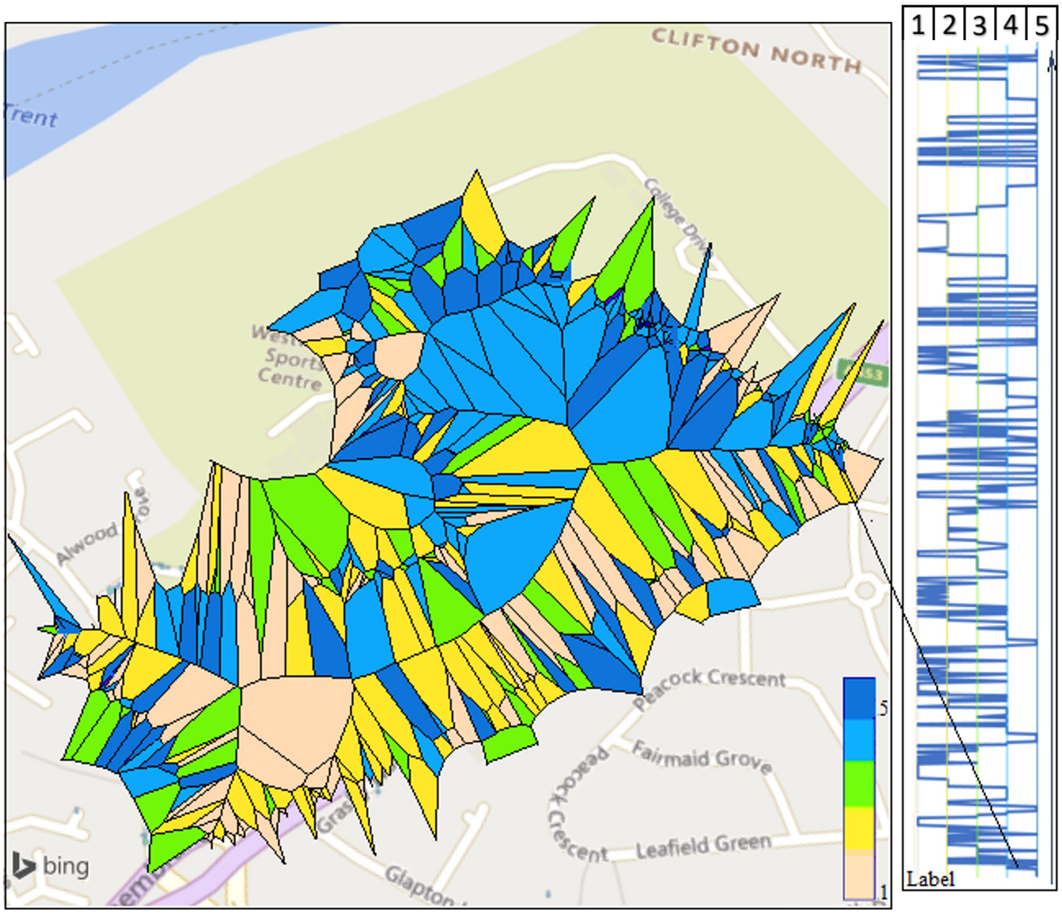


Figure 5.3: (left) Voronoi overlay from one participant data. Each polygon represents one location trace tagged with a wellbeing label while collecting the data in a specified route (the map layer from Microsoft Bing), (right collected label data from start to end).

Voronoi visualisations have demonstrated how changes within the environment specifically air quality levels can help to provide an immersive and illustrative approach to investigate the influence of surroundings, human physiology, and transient emotional states. The colour of the polygons represent the self-report wellbeing data collected from participants from very low/ negative (1) to very high/ positive (5). Poor wellbeing is indicated by lighter colours (i.e. cream and yellow) which was most reported along the main road where high levels of air quality were also experienced as evidenced at Section 5.2.1.1.

On the other hand a positive state of wellbeing was recorded from participants in less polluted areas such as fields and open spaces which are indicated by dark colours (i.e. blue). In comparison to the heat maps at Section 5.2.1.1 it can be identified that at these specific points where positive wellbeing was labelled there were lower levels of PM 2.5, PM 10, noise and gases recorded. At the right of the Voronoi diagram depicts the labelled data from a participant whilst walking along the route. As an example, the arrow shows that when participants are met with a change in the environment such as an increase in poorer air quality levels, participants label as being unhappy.

By utilising spatial and temporal approaches, the intricate connections between the physical surroundings, bodily responses, and momentary well-being can be explored to enhance our insights in a visually captivating manner. Recent advances particularly in mobile sensing, GPS and wellbeing labelling at the point-of-exposure enables easy collection of spatio-temporal data to a new level. As a direct result large datasets raise issues on how researchers can derive useful knowledge through patterns and as a whole the data collected.

5.2.2 Urban Wellbeing

The second research application builds upon the DigitalExposome concept and explores the potential of acquiring and integrating multimodal environmental air quality and wellbeing data at the point of exposure. By fusing these different data modalities, a more comprehensive picture of the impact of air quality on wellbeing can be obtained, leading to more targeted interventions and improved public health outcomes [236]. As a result, *Urban Wellbeing* as previously discussed in Section 3.3.3 is an alternative to existing traditional approaches of collecting mental wellbeing through questionnaires [22], [112] and environmental data from pre-existing datasets [201]. As it stands little research has focused to date, around the use of smartphone technology to support and understand the impact of air quality and mental health [114]. The mobile application uses fixed environmental sensing systems to digitise and enable real-time monitoring of impact towards self-labelled mental wellbeing using a smartphone.

Following the process of designing and developing *Urban Wellbeing*, this section focuses on a real-world experiment of 52 devices who downloaded the application resulting in over 100 assessments completed. *Urban Wellbeing* was tested 'in the wild' across a range of different urban environments from places across the UK to evaluate the performance of understanding the relationship between urban environments and momentary wellbeing.

To analyse the collected environmental variables and self-labelled wellbeing data, firstly, a Pearson's R Correlation Coefficient Matrix was utilised to measure associations between the variables as depicted at Figure 5.4. The results highlight environmental sensor data (labelled Air Index) positively correlated (0.2) with noise. In addition, the self-report label correlated negatively (-0.4) with noise indicating that when noise is increased wellbeing is reduced. This is a similar occurrence in the work at Section 4.3 and across other studies [102].

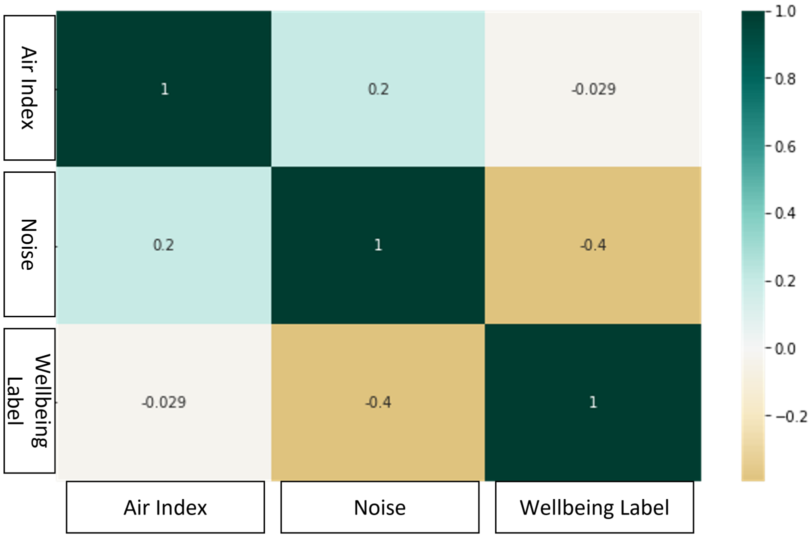


Figure 5.4: A Correlation Matrix depicting the relationship between the collected environment air quality and physiological factors.

The collected sensor data from the application can be visualised as seen in Figure 5.5 demonstrating three assessments completed by independent users across Nottingham, from at each point, Environmental Air Quality Index level; self-labelled wellbeing; noise level and an image of the environment was collected from the sensors. An Air Quality Index is deemed acceptable when the score is between 0-50 with values of 50-100 becoming more serious for those with sensitive needs to air pollution [3]. Across the individual assessments collected over 65% recorded above the level of 50.

The visualisation results highlight that in green spaces wellbeing was labelled positively or very positively, and also showed noise levels were relatively low and air quality levels were very good between AQI 19 - 28. The findings of green spaces having lower levels of air quality is in line with previous research studies [59], [187]. In comparison to a very negative wellbeing indicated when air quality was poor at AQI 70 and noise high. In other studies, noise has correlated with wellbeing showing that the level can impact how someone is feeling [102]. Additionally, previous research on the impact of noise has shown that anything above 80 dB can be dangerous in prolonged scenarios [139]. At each point analysed, it demonstrates that by using environmental sensor data, mental wellbeing

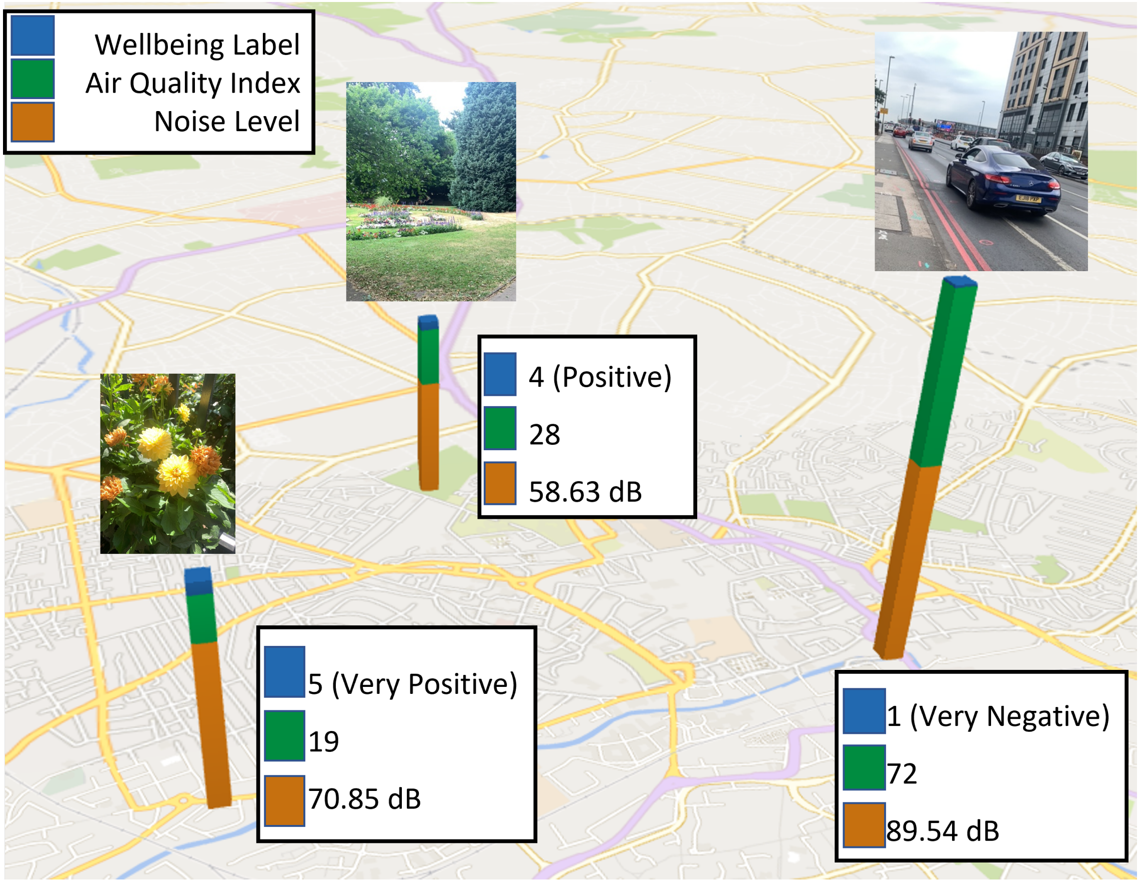


Figure 5.5: Graph overlay from three participant's data. Each column represents a location trace tagged with the level of noise, Air Quality Index, self-reported wellbeing label and an image of the current environment demonstrating a range of impacting factors within the environment.

is impacted by factors within the environment such as Air Quality, Noise and the surrounding image of an environment.

Figure 5.6 depicts several images taken from each time a participant completed an assessment where they self-labelled their wellbeing as either very negative or negative. The purpose of obtaining an image in the process was to increase knowledge into the relationship impact between the environment; namely air pollution factors, mental wellbeing and noise.

The results and subsequent analysis of the photographs obtained from the mobile application in the scenario of a negative wellbeing show that walking near or close to busy traffic and a congested high street could impact mental wellbeing

Figure 5.6: Several images obtained from photographs taken during an individual assessment from Urban Wellbeing where self-report wellbeing was labelled as 1 (very negative impact) or 2 (negative impact) as indicated at each image.

significantly. In addition, it is worth noting that in these 6 scenes the collected sensing data shows that noise levels were averaged at 85 dB in these locations. On the other hand, further analysis into the photographs taken where wellbeing was either positive or very positive shows individual's in green spaces, free from traffic and noise levels typically lower than 65 dB, depicted at Figure 5.7. The work on Urban Wellbeing as identified in Chapter 3 goes beyond previous work which shows how alternative factors within the environment can be quantified rather than just noise [100], [139], [159], [82], [22].

Figure 5.7: Several images obtained from photographs taken during individual assessment from Urban Wellbeing where self-report wellbeing was labelled as 5 (very positive impact) or 4 (positive impact) as indicated at each image.

5.2.3 Enviro-IoT

The second research application is the development of the Enviro-IoT as a low-cost alternative to environmental air quality monitoring. This work formed part of a 'co-location' project working alongside Nottingham City Council Environmental Health, Department for Environment, Food and Rural Affairs (Air Quality) and

Ricardo Air Quality Specialists.

The *Enviro-IoT* as discussed in Section 3.3.2.1 is capable of sensing levels of Particulate Matter (1.0, 2.5 and 10), Ammonia, Carbon Monoxide, Nitrogen Dioxide and Oxidising gases in real-time. The obtained sensor data is sent directly to a database for storage. All of the electronics have been encased into a small 3D printed plastic box as depicted at Figure 5.8, which can be easily installed within any environment. In this case, the three photographs depict the Enviro-IoT placed at the Nottingham City Centre beside the DEFRA AURN sensors for air quality monitoring.

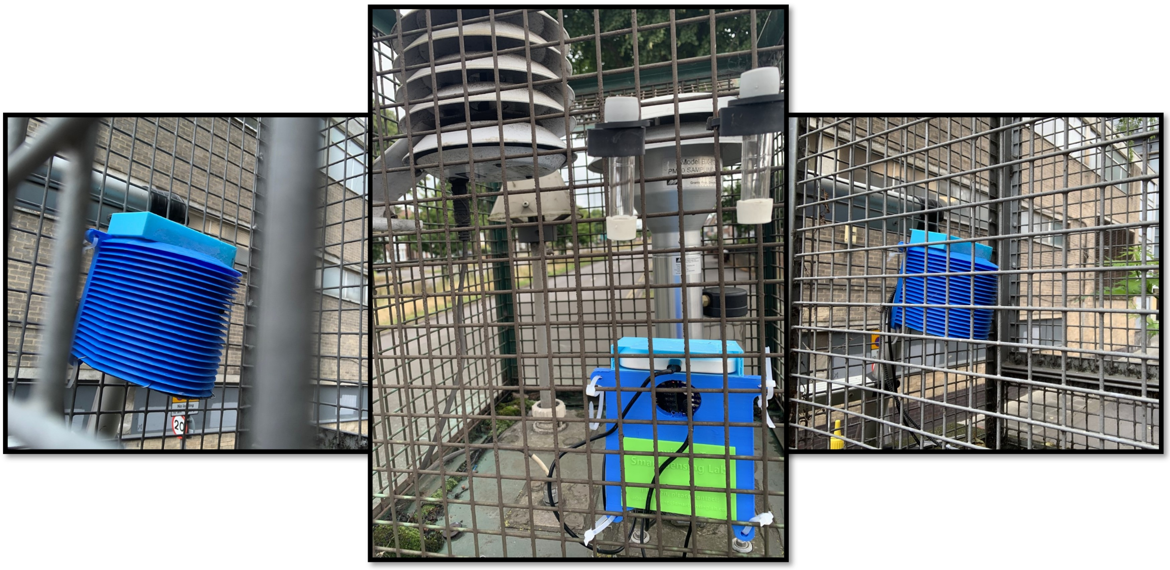


Figure 5.8: The Enviro-IoT devices in-situ at the DEFRA owned AURN Site in Nottingham City Centre.

Two Enviro-IoT devices were placed at both the Nottingham City and Western Boulevard DEFRA AURN Station sites. Real-world trials across a 9-month period (from August 2022 to April 2023) using the Enviro-IoT alongside the AURN device showed high accuracy levels indicating that incorporating low-cost sensors into an environmental monitoring system highlighted considerable reliability.

In this evaluation, three air quality factors were selected which includes: Particulate Matter 2.5, 10 and Nitrogen Dioxide. The sensor data was sampled at a rate of 5-minute intervals and then these values were averaged to give the mean concentration per hour to match the frequency reported by the AURN station.

Statistical analysis has been applied to the data collected from both devices in the form of Pearson's R and Spearman's to understand the correlations between collected values as depicted at Table 5.1.

Table 5.1: Statistical Analysis of Environmental Pollution Variables Comparison between AURN and Enviro-IoT.

Pollution Type	Pearson's R	Spearman's
PM2.5	0.983	0.964
PM10	0.98	0.92
NO2	0.96	0.94

Table 5.1 indicates that the Enviro-IoT system is highly reliable and producing accurate results for the three pollutants specifically: PM 2.5, PM 10 and NO2 which is in line with the industrial standard equipment with results above the 90% accuracy. In addition, Figure 5.9, 5.10 and 5.11, visualises the readings for both the Enviro-IoT and DEFRA AURN across a 9-month period of testing. Analysis of Figures 5.9 and 5.10, show both devices record very similar values with both lines following the same path. January values have been omitted from the graphs as the Enviro-IoT encountered WiFi connection issues whereby the sim-card ran out of data so the air quality data for this month was not recorded by the system.

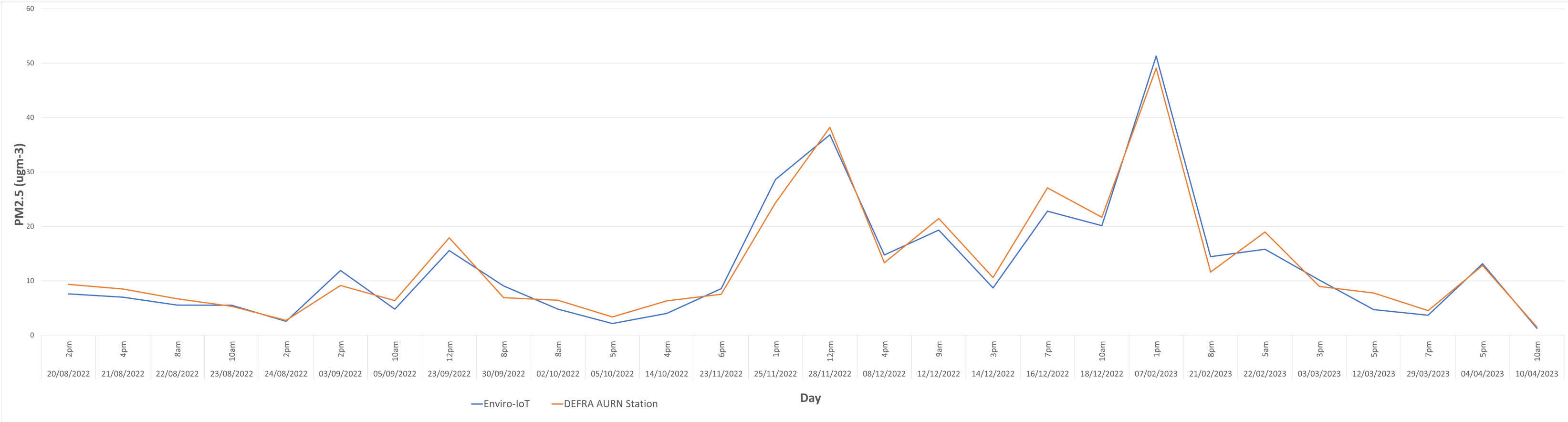


Figure 5.9: The impact of PM 2.5 on air quality of readings from low-cost sensors (Enviro-IoT) and Industry standard equipment (AURN).

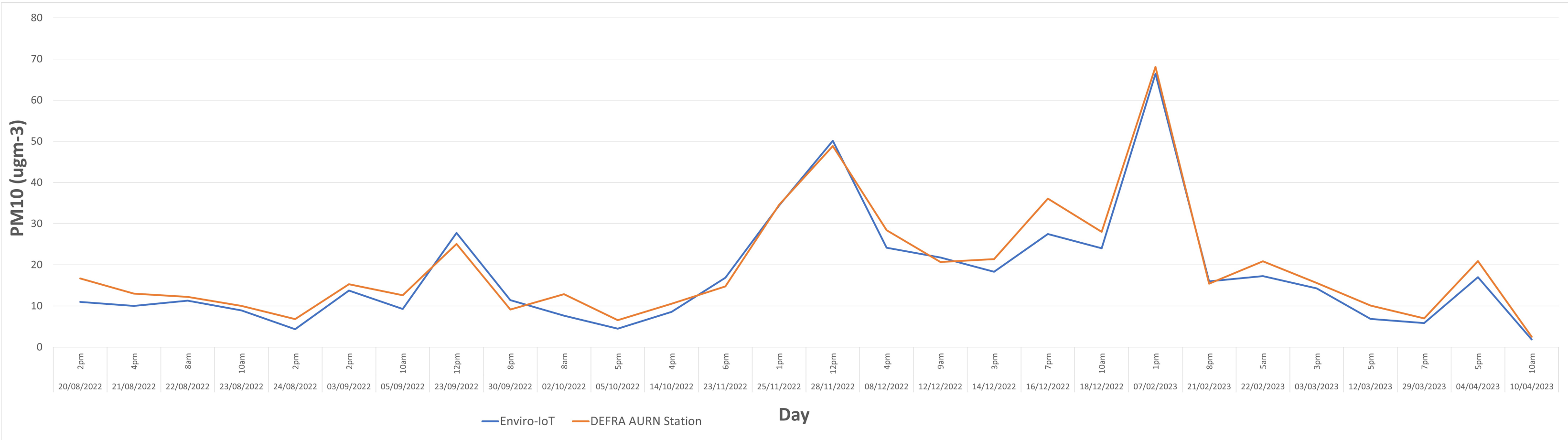


Figure 5.10: The impact of PM 10 on air quality of readings from low-cost sensors (Enviro-IoT) and Industry standard equipment (AURN).

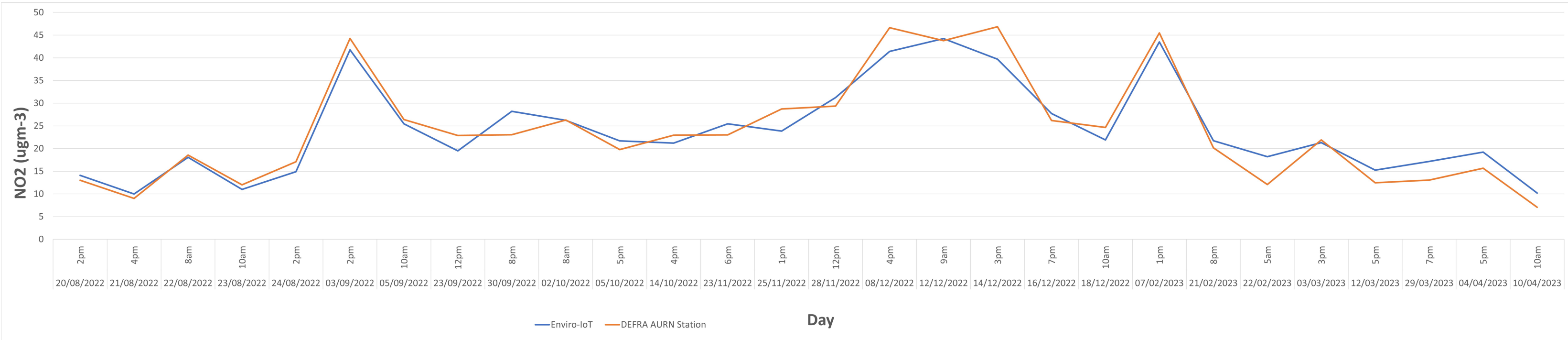


Figure 5.11: The impact of NO2 on air quality of readings from low-cost sensors (Enviro-IoT) and Industry standard equipment (AURN).

Figure 5.11 shows the analysis of NO2 levels recorded by two different devices. It is observed that the NO2 values recorded by both devices exhibit similar patterns over time. However, there seems to be slightly more variability and deviation from the expected trend between the two devices, especially noticeable during the months of November and December.

It is indeed important to note that colder months, such as November and December, are generally associated with higher air pollution levels [239]. This pattern is evident in Figures 5.9, 5.10, and 5.11, where both the Enviro-IoT and AURN devices show an increase in pollutant levels during these months. This aligns with the understanding that certain weather conditions, such as temperature inversions and reduced dispersion, can lead to higher pollution concentrations in colder months. Regarding the correlation of collected PM2.5 and PM10, the figures demonstrate that both devices show very similar levels throughout the 9-month study. This suggests that the small low-cost sensors used in the Enviro-IoT device are capable of accurately reporting hourly pollution levels. The strong correlation between the measurements obtained from the two devices reinforces the reliability and suitability of the Enviro-IoT system for environmental monitoring in real-world settings.

Overall, the comparison of the Enviro-IoT and AURN devices supports the conclusion that the small low-cost sensors used in the Enviro-IoT system are a reliable solution for environmental monitoring.

5.3 Conclusion

In this chapter, three distinct real-world applications have been presented and extensively discussed. These applications serve as effective means to quantify and evaluate the relationship between the environment and wellbeing, with the ultimate goal of monitoring and enhancing wellbeing within different urban environmental contexts. They also support the DigitalExposome Concept and Framework. The insights gained from these studies can be used to develop strategies and interventions aimed at promoting better health and quality of life in different environmental settings.

The Enviro-IoT has demonstrated the potential to revolutionise environmental monitoring through the utilisation of affordable sensors and on-board IoT. This innovative approach enables the real-time capture of air quality levels, thereby offering valuable insights in areas where monitoring air quality is challenging. Leveraging low-cost sensors and IoT capabilities, the project addresses the need for accurate and up-to-date information on air quality, which can be crucial for identifying regions with poor air quality and facilitating targeted interventions to mitigate potential health risks. This advancement has the potential to significantly improve public health and enhance overall air quality management.

Secondly, spatial and temporal visualisations demonstrate a robust approach in quantifying environmental sensor and self-labelled wellbeing by unravelling the impact on movement and understanding towards patterns. In this work heat maps and Voronoi have been used to model environmental, physiological and mental wellbeing data and shown clear relationships between the modalities. The focus of this work is valuable and meaningful, particularly to urban planners which can utilise this data more widely by designing environments that contain calming areas (green spaces) in those areas where air quality is poor and as a result brings on a negative mental wellbeing.

Finally, Urban Wellbeing presents a new significant opportunity to monitor an individual's mental wellbeing while obtaining real-world environmental sensor data to observe and understand the impact directly at the point of exposure. The cross-platform technology behind Urban Wellbeing provides a non-obtrusive mean for individual's to understand how their wellbeing is impacted in urban environments.

Chapter 6

Conclusion and Future Work

This thesis has introduced a novel framework aimed at achieving the objective of quantifying the intricate relationship between the environment, physiology, and mental wellbeing. By leveraging data science techniques has provided significant strides in advancing the understanding and measurement of affective states in a real-world urban environment. In this final chapter a general overview of the undertaken work, summary of contributions, areas of improvement and future work are all considered.

6.1 Conclusion

The work undertaken in this thesis presents the design, implementation and analysis of the framework DigitalExposome which quantifies the relationship between environment with a focus on air quality towards human physiology and mental wellbeing. Using a sensor-based fusion approach and data science aims to explore this perspective. The research questions originally posed in Chapter 1 have been explored and addressed through the studies in this thesis. There were summarised as the following:

1. How can we quantify the person-environment interaction to help explore an urban environment that promotes a positive wellbeing?

2. How we monitor, fuse and model the relationship of a multimodal approach

to understand the impact between urban environment, human physiology and mental wellbeing?

3. Can real-world sensor data be quantified and leveraged to infer wellbeing or mental states through personal and community approaches?

4. What are the best approaches for validation and application of DigitalExposome for real-world use cases?

To investigate and evaluate the effectiveness of DigitalExposome in quantifying mental wellbeing, a series of real-world experiments were conducted involving the development of smartphone technology incorporating cross-platform mobile applications to obtain wellbeing at the point of exposure. Hence, drawing on these attempts the following sections of this chapter summarise the findings and major conclusions that have been achieved.

Chapter 3 focuses on the design of the conceptual framework known as *DigitalExposome* which aims to quantify the relationship between the environment, physiology, and mental wellbeing. This chapter specifically explores the utilisation of digital technologies, including mobile sensors and data science, to enable a more comprehensive assessment of environmental exposures and inference of mental wellbeing. To facilitate this process, the development of various environmental low-cost, affordable sensing stations was undertaken, including both fixed and portable contained within 3D-printed devices. Among these devices, a fixed sensing system was selected and tested within a real-world setting, alongside industry standard equipment. The objective was to ensure that the measurements obtained from the low-cost sensing system were highly reliable and accurate in comparison to established standards. The validation of the sensing system's reliability was a significant milestone in the research process. The step ensured that the air quality data collected throughout the subsequent Chapter 4 and Chapter 5 could be relied upon for assessing the impact of environmental exposures on mental wellbeing.

As identified in the literature at Section 2.2.3, the CDC [70] explored the use of the Exposome concept in its current form and acknowledged two areas that require further work in that the use of digital technologies to measure external and internal factors and the way in which the new techniques can be validated. The development of DigitalExposome and the sensing technologies validated in Chapters 3, 4 and 5 demonstrate that this approach can be a viable step forwards to continue the work in quantifying the relationship of the environment to wellbeing.

Smartphone technology has been explored in the form of a cross-platform custom-built mobile applications (Android & iOS) that collects real-world, real-time current environmental air quality levels, type of environment including a respective image and in situ noise levels and self-labelled mental wellbeing. This application helps to provide a better understanding into affective states with the ability to label sensor data at the point of collection. Furthermore, three datasets of this research are explained and discussed in terms of experimental setup, data obtained and route selected in addition to the steps taken to pre-process the data as many of the data types were collected at different sampling rates.

Quantifying the relationship through classification of real-world mental wellbeing states of affect, the developed technological devices identified in Chapter 3 were utilised for real-world data collection of environmental, human physiology and self-labelled mental wellbeing. Therefore, Chapter 4 explores from a multi-modal perspective of classifying mental wellbeing in a two-fold approach. Firstly, an initial analysis at an individual level studied the focus on examining the instances of stops and moves across a single trajectory towards understanding the direct personalised impact.

Although noise has previously been considered, this is some of the first work to show the incorporation of environmental air quality factors and physiological data using real-world sensing technology as semantics to enrich the trajectory. As such, a DTW model was trained using 12 features made up of the collected environmental and physiological sensor data and then classified using three classifiers (KNN, DT and RF). The results showed that by combining trained environmental and physiological data together to classify wellbeing had significant benefit that resulted in a higher accuracy of 0.88 (F1-score) whilst individually trained and classified scored 0.76 (physiological) and 0.84 (environmental). Leveraging environmental and physiological data together helped in improving the overall accuracy, highlighting the importance of environmental data in real-world inference of wellbeing. Furthermore, the high classification rate of environmental data alone shows the importance these variables when calculating wellbeing in urban environments.

Secondly, a multimodal aggregated approach in Section 4.5 was explored initially using a DBN network and later a 1D-CNN both trained using environmental and physiological data. The results show that adopting a multimodal approach in classification of environmental and physiological data combined increased model performance achieving an average accuracy of 0.76 (F1-Score) compared to individually trained data environmental and physiological achieving 0.67 and 0.61 (F1-Score) respectively. Typically, it would be expected due to the correlation to the sympathetic nervous system that physiological should classify higher but in both studies environmental highlighted increased scores demonstrates the impact sensor data particularly air quality from the environment can have when classifying mental wellbeing states.

As a result of this research several real-world applications have been discussed and explored in Chapter 5. The growing level of the environment is now more important that ever before with an increasing number of people experiencing health, behavioural, wellbeing related issues and death. Therefore, data visualisation concepts have been explored in a spatio-temporal approach through heat mapping and Voronoi as a mechanism of understanding the relationship between environment, physiology and self-report wellbeing depicting a negative wellbeing in areas of a high exposure to poor air quality. The approach extracts the individual collected sensor modalities and plot across a map to understand the intensity of factors on wellbeing.

Furthermore, the *Enviro-IoT* has been developed to demonstrate the impact of using miniaturised low-cost, affordable sensors and Internet of Things (IoT) technologies to monitor the environment in a real-world context. Although the system is primarily for a fixed position, studies conducted in this thesis have shown the potential of using this within a rucksack for unobtrusive monitoring of the environment, as seen in Section 3.4.1 and 3.4.2. The work of Enviro-IoT has many benefits particularly to its capacity to monitor real-world air quality levels in real-time and having the ability to be positioned in multiple places without the need for many resources. Finally, Urban Wellbeing, a smartphone mobile application has been developed that combines real-world environmental and momentary self-labelled wellbeing, obtained at the point of exposure. By developing a mobile application and connecting it to a UK network wide air quality monitoring system enables the possibility of continually and unobtrusively monitoring wellbeing through urban environments with the potential to inform of areas where wellbeing could be impacted.

Overall, the work explored in this thesis has consistently demonstrated the relationship between environment, human physiology and momentary self-labelled wellbeing that it can be accurately quantified through the incorporation of DigitalExposome and the use of custom-built sensing technologies to monitor and capture exposures encountered by a human in addition to data science. The DigitalExposome work has the potential to greatly improve understanding and assessment tools that assess real-world wellbeing.

6.2 Summary of Major Thesis Novel Contributions

This section highlights the significant contributions that have been made in this thesis with the specifics including:

6.2.1 A Novel Framework for Quantifying the relationship between Environment and Mental Wellbeing

The DigitalExposome framework encompasses the ability to unravel the relationship between environment, human physiology, behaviour and momentary mental wellbeing for the quantification inference of wellbeing for urban environmental settings. The developed framework consists of five stages with each defined as:

1. The conceptual layer identifies the four main areas that have been shown in previous research to have an impact towards mental wellbeing which involves (1) environmental factors, (2) biological factors, (3) social aspects and (4) cultural factors.

2. Sensing layer to detail the smart technology devices and low-cost sensors that help in monitoring wellbeing and the environment.
3. Computing layer identifies the core data science and analytical techniques that will unravel the impact between variables and help to infer wellbeing states.
4. Application layer details the potential of DigitalExposome within the real-world for a greater understanding to active monitoring of wellbeing, environment and physiological health.

6.2.2 Real-World Data Collection

- Environmental monitoring systems have been explored and a new low-cost, affordable resource has been custom-built and shown to be extremely reliable and accurate in a range of urban environments whilst co-located to industrial standard equipment. The approaches have resulted up to around 95% for Particulate Matter 1.0, 2.5 & 10 and to 97% for Nitrogen Dioxide in terms of reliability.
- Methods to label data in real-time have been explored through the use of smartphone technology. The results demonstrate the benefits of using emojis to label how participants were feeling within environments across a five-point scale which showed a reliable and accurate concept to label data in both scenarios at the point of exposure. Additionally, the aim of obtaining environmental air quality data and self-labelled wellbeing has been trialled showing the ability to understand the impact at the point of exposure.

6.2.3 Deep Learning Classification on MultiModal Sensor Fusion

- This work developed a personalised-individual impact approach through semantic trajectories and episodes based on classifying five-states of real-world wellbeing achieving up to 0.88 (F1-score) accuracy, highlighting the impact air quality data can have in classifying wellbeing states.
- Developed a second approach using aggregated data to classify the five-states of real-world mental wellbeing resulting in up to accuracy levels of 0.76 (F1-score). This has highlighted the importance environmental data can have on improving the classification of wellbeing.

6.2.4 Real-Time Research Application Interventions

- Three individual approaches developed from the research conducted: Enviro-IoT - low-cost environmental sensing system, Temporal and Spatial Visualisation - Use of Heat Maps and Voronoi as a way to visualise real-world data to understand movement and behaviour patterns and Urban Wellbeing - Cross-platform smartphone mobile application to obtain real-world environmental data and self-labelled wellbeing collected at the point-of-exposure.

6.3 Future Directions, Challenges and Recommendations

Following the work explored in this thesis, this section identifies the main directions for future research and recommendations for improvements:

1. DigitalExposome has demonstrated an achievable approach of monitoring mental wellbeing states through the use of low-cost sensing and smartphone technology, coupled with data science. Previously, calculating the Exposome concept was a very cumbersome process, relying on existing datasets to understand the impact to health which required many years of gathering prior to analysis. The addition of capturing more factors within the environment in the future would be advantageous to gain a more clearer understanding into how human wellbeing changes. With this, the additional new data science concepts to extend the analysis would be advantageous. Further analysis work behind the framework, particularly around areas not considered in this work such as the conceptual layer to review alternative factors such as biological, social and cultural would be advantageous to take this work further. Additionally, in the application layer at how we can provide more prevention in the form of treatment.

2. The area of environmental monitoring using low-cost sensors continues to receive increasing attention. However, as identified in Section 2.2.5, most of these efforts at the moment are directed towards heavy, expensive and fixed monitoring systems. An effort to improve the viability of low-cost sensors as explored in this thesis is promising and has showed clear potential, particularly for areas where it is difficult to monitor due to size and overall cost. Although, there are still areas for continued exploration through long-term monitoring and the impact, in addition to a wider selection of sensors selected and tested to gather a large amount of air quality data.

3. Investigating beyond the urban environment presents a world of opportunities when it comes to people, exploring social aspects, fostering connections and embracing the concept of social prescribing in a real-world setting.

Expanding the focus beyond the confines of cities opens up a deeper understanding of diverse populations and the need for a more positive change in our environment to improve mental wellbeing. Consideration towards social aspects can help in developing targeted interventions and policies.

4. Although not explored in this research, but stated in the DigitalExposome Conceptual Framework, on-device processing in edge-computing provides many future opportunities to develop this research further, particularly for real-time inference. In recent years there has been significant advances in edge computing which has reduced the need to have large computing devices for processing and smartphone technology for labelling. In this sense, there is the potential to develop a multimodal sensor-fusion on-edge approach incorporating environmental, human physiological and momentary self-labelled mental wellbeing to infer wellbeing states whilst moving between a range of urban environments, while using low-cost sensing technologies.

5. Approaches in this work have explored a range of deep learning models including a CNN, DBN and DTW in supporting the inference of affective states. With further advances in model architectures and developed new networks enable explorations of increasing affective state modelling accuracy performance.

6. Although the datasets that have been devised from this thesis are encouraging in monitoring of mental wellbeing states within urban environments, some are small. Future research in this area is to combine a more diverse dataset from increased participation from individuals and exposure to urban environments for longer periods of time to understand the impact on wellbeing better. Due to the global Covid-19 pandemic and the experimental resources required to obtain data it has been difficult to recruit further participants for experiments.

References

[1] David B. Abrams, J. Rick Turner, Linda C. Baumann, Alyssa Karel, Susan E. Collins, Katie Witkiewitz, Terry Fulmer, Molly L. Tanenbaum, Persis Commissariat, Elyse Kupperman, Rachel N. Baek, Jeffrey S. Gonzalez, Nicole Brandt, Rachel Flurie, Jennifer Heaney, Christopher Kline, Linda Carroll, Jane Upton, Patrícia Cardoso Buchain, Adriana Dias Barbosa Vizzotto, Alexandra Martini de Oliveira, Tania C. T. Ferraz Alves, Quirino Cordeiro, Lorenzo Cohen, M. Kay Garcia, Amy Jo Marcano-Reik, Siqin Ye, Yori Gidron, Marc D. Gellman, M. Bryant Howren, Manjunath Harlapur, Daichi Shimbo, Keisuke Ohta, Naoya Yahagi, Elizabeth Franzmann, Abanish Singh, Linda C. Baumann, Alyssa Karel, Debra Johnson, Benjamin L. Clarke, Debra Johnson, Rachel Millstein, Karen Niven, Karen Niven, Eleanor Miles, J. Rick Turner, Barbara Resnick, Yori Gidron, Carter A. Lennon, Kelly S. DeMartini, Kristin L. MacGregor, Susan E. Collins, Megan Kirouac, J. Rick Turner, Abanish Singh, Yori Gidron, Yoshiharu Yamamoto, Urs M. Nater, Nicole Nisly, Debra Johnson, Derek Johnston, Ydwine Zanstra, Derek Johnston, Youngmee Kim, Della Matheson, Brooke McInroy, Christopher France, Shin Fukudo, Emiko Tsuchiya, Yoko Katayori, Martin Deschner, Norman B. Anderson, Chad Barrett, Mark A. Lumley, Lindsay Oberleitner, Stephan Bongard, Siqin Ye, Amy Jo Marcano-Reik, Seth Hurley, Seth Hurley, Anna Maria Patino-Fernandez, Anna C. Phillips, Tatsuo Akechi, Anna C. Phillips, Amy Jo Marcano-Reik, Nicole Brandt, Rachel Flurie, Sarah Aldred, Kim Lavoie, Manjunath Harlapur, Daichi Shimbo, Kate L. Jansen, Katherine T. Fortenberry,

Molly S. Clark, Rachel Millstein, Toru Okuyama, William Whang, Mustafa Al'Absi, Bingshuo Li, Yori Gidron, J. Rick Turner, Elizabeth R. Pulgaron, Diana Wile, Linda C. Baumann, Alyssa Karel, Beth Schroeder, Mary C. Davis, Alex Zautra, Shannon L. Stark, William Whang, Ana Victoria Soto, Yori Gidron, Anthony J. Wheeler, Scott DeBerard, Josh Allen, Akihisa Mitani, Akihisa Mitani, Elizabeth R. Pulgaron, Akihisa Mitani, Jennifer Carter, William Whang, Beth Schroeder, Angela M. Hicks, Carolyn Korbel, Austin S. Baldwin, Kevin S. Spink, Darren Nickel, Michael Richter, Rex A. Wright, Julian F. Thayer, Michael Richter, Rex A. Wright, and Deborah J. Wiebe. Affect. *Encyclopedia of Behavioral Medicine*, pages 49–50, 2013. URL: https://link.springer.com/referenceworkentry/10.1007/978-1-4419-1005-9_1088, doi:10.1007/978-1-4419-1005-9_1088. 33

[2] Aeroqual. Aqm - ambient air quality monitoring system, 2019. URL: https://www.aeroqual.com/products/aqm-stationshttps://www.aeroqual.com/outdoor-air-quality-monitors/aqm-stations. 2, 28

[3] U.S. Environmental Protection Agency. Technical assistance document for the reporting of daily air quality - the air quality index (aqi), 2022. URL: https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf. 131

[4] Clean Air and Strategy Plan. Clean air strategy plan, 2011. 17

[5] Lulwah Al-barrak, Eiman Kanjo, and Eman M. G. Younis. Neuroplace: Categorizing urban places according to mental states. *PLOS ONE*, 12:e0183890, 9 2017. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0183890, doi:10.1371/JOURNAL.PONE.0183890. 65, 81, 84, 85

[6] Nouf Alajmi, Eiman Kanjo, Alan Chamberlain, and Nour El Mawass. Shopmobia: An emotion-based shop rating system. *2013 Humaine Associa-*

tion Conference on Affective Computing and Intelligent Interaction, 2013. doi:10.1109/ACII.2013.138. 81, 84

[7] Basma H. Albanna, Ibrahim F. Moawad, Sherin M. Moussa, and Mahmoud A. Sakr. Semantic trajectories: A survey from modeling to application. *Lecture Notes in Geoinformation and Cartography*, 216:59–76, 2015. doi:10.1007/978-3-319-16667-4_4. 24

[8] Hasah Alheneidi and Andrew P Smith. Perceptions of noise exposure, information overload, wellbeing and academic attainment. *13th ICBEN Congress on Noise as a Public Health Problem*, 2021. 2

[9] Navid Amini, Majid Sarrafzadeh, Alireza Vahdatpour, and Wenyao Xu. Accelerometer-based on-body sensor localization for health and medical monitoring applications. volume 7, pages 746–760. Elsevier B.V., 2011. doi:10.1016/j.pmcj.2011.09.002. 41

[10] Dominic Abrams and G N Martin.. M A Hogg. *Psychology: Social cognition and attitudes.* Pearson Education, 2010. URL: https://kar.kent.ac.uk/id/eprint/23659. 33

[11] Jonathan O. Anderson, Josef G. Thundiyil, and Andrew Stolbach. Clearing the air: A review of the effects of particulate matter air pollution on human health. *Journal of Medical Toxicology*, 8:166, 6 2012. doi:10.1007/S13181-011-0203-1. 30

[12] Xanthi D. Andrianou and Konstantinos C. Makris. The framework of urban exposome: Application of the exposome concept in urban health studies. *Science of the Total Environment*, 636:963–967, 2018. doi:10.1016/j.scitotenv.2018.04.329. 19

[13] Joshua S. Apte, Kyle P. Messier, Shahzad Gani, Michael Brauer, Thomas W. Kirchstetter, Melissa M. Lunden, Julian D. Marshall, Christopher J. Portier, Roel C.H. Vermeulen, and Steven P. Hamburg. High-resolution air pollution mapping with google street view cars: Exploiting

big data. *Environmental Science and Technology*, 51:6999–7008, 6 2017. doi:10.1021/acs.est.7b00891. 18

[14] AQMesh. Aqmesh | the best small sensor air quality monitoring system. URL: https://www.aqmesh.com/products/aqmesh/https://www.aqmesh.com. 27

[15] AQMesh. Aqmesh | the best small sensor air quality monitoring system, 2022. URL: https://www.aqmesh.com/case-studies/newcastle-s-urban-observatory/https://www.aqmesh.com. 27

[16] AQMesh. Newcastle's urban observatory | aqmesh | the best small sensor air quality monitoring system, 2023. URL: https://www.aqmesh.com/case-studies/newcastle-s-urban-observatory/. 5, 27

[17] Noel J. Aquilina, Juana Maria Delgado-Saborit, Stefano Bugelli, Jason Padovani Ginies, and Roy M. Harrison. Comparison of machine learning approaches with a general linear model to predict personal exposure to benzene. *Environmental Science and Technology*, 52:11215–11222, 2018. URL: https://pubs.acs.org/sharingguidelines, doi:10.1021/acs.est.8b03328. 37, 63

[18] Patricia A. Areán, Kien Hoa Ly, and Gerhard Andersson. Mobile technology for mental health assessment. *Dialogues in Clinical Neuroscience*, 18:163–169, 2016. doi:10.31887/dcns.2016.18.2/parean. 54

[19] Peter Aspinall, Panagiotis Mavros, Richard Coyne, and Jenny Roe. The urban brain: Analysing outdoor physical activity with mobile eeg. *British Journal of Sports Medicine*, 49:272–276, 2015. doi:10.1136/bjsports-2012-091877. 18, 26, 84

[20] Robert Avram, Geoffrey H. Tison, Kirstin Aschbacher, Peter Kuhar, Eric Vittinghoff, Michael Butzner, Ryan Runge, Nancy Wu, Mark J. Pletcher, Gregory M. Marcus, and Jeffrey Olgin. Real-world heart rate norms in the health eheart study. *NPJ Digital Medicine*, 2, 12 2019. URL:

/pmc/articles/PMC6592896//pmc/articles/PMC6592896/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6592896/, doi:10.1038/S41746-019-0134-9. 40

[21] Kenichi Azuma, Naoki Kagi, U. Yanagi, and Haruki Osawa. Effects of low-level inhalation exposure to carbon dioxide in indoor environments: A short review on human health and psychomotor performance. *Environment International*, 121:51–56, 12 2018. doi:10.1016/j.envint.2018.08.059. 31

[22] Ioannis Bakolis, Ryan Hammoud, Michael Smythe, Johanna Gibbons, Neil Davidson, Stefania Tognin, and Andrea Mechelli. Urban mind: Using smartphone technologies to investigate the impact of nature on mental well-being in real time. *BioScience*, 68:134–145, 2 2018. URL: https://academic.oup.com/bioscience/article/68/2/134/4791430, doi:10.1093/biosci/bix149. 5, 36, 51, 121, 130, 133

[23] Lisa Feldman Barrett and Eliza Bliss-Moreau. Affect as a psychological primitive. *Adv Exp Soc Psychol*, 41:167–218, 2009. doi:10.1016/S0065-2601(08)00404-8. 33

[24] Sarah Louise Bell, Suzanne Audrey, David Gunnell, Ashley Cooper, and Rona Campbell. The relationship between physical activity, mental well-being and symptoms of mental health disorder in adolescents: A cohort study. *International Journal of Behavioral Nutrition and Physical Activity*, 16:138, 12 2019. doi:10.1186/s12966-019-0901-7. 17, 41

[25] Maximus Berger and Zoltán Sarnyai. "more than skin deep": Stress neurobiology and mental health consequences of racial discrimination, 1 2015. doi:10.3109/10253890.2014.989204. 2

[26] Alberto Betella and Paul F.M.J. Verschure. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PLoS ONE*, 11, 2 2016. doi:10.1371/journal.pone.0148037. 33, 66

[27] Viv Bewick, Liz Cheek, and Jonathan Ball. Statistics review 7: Correlation and regression. *Critical Care*, 7:451, 12 2003. URL: /pmc/articles/PMC374386//pmc/articles/PMC374386/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC374386/, doi:10.1186/CC2401. 62

[28] Rizwan Hassan Bhat. Environmental stressors and its impact on human being. *International Journal of Humanities and Social Sciences.*, 5:37–40, 2017. 17

[29] Vania Bogorny, Chiara Renso, Artur Ribeiro de Aquino, Fernando de Lucca Siqueira, and Luis Otavio Alvares. Constant-a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, 18:66–88, 2 2014. doi:10.1111/TGIS.12011. 25

[30] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25:49–59, 1994. doi:10.1016/0005-7916(94)90063-9. 34

[31] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25:49–59, 3 1994. doi:10.1016/0005-7916(94)90063-9. 66, 76, 85

[32] Nieves R. Brisaboa, Miguel R. Luaces, Cristina Martínez Pérez, and Ángeles S. Places. Semantic trajectories in mobile workforce management applications. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10181 LNCS:100–115, 2017. doi:10.1007/978-3-319-55998-8_7. 92

[33] Jaime G. Carbonell, Ryszard S. Michalski, and Tom M. Mitchell. an overview of machine learning. *Machine Learning*, pages 3–23, 1983. doi:10.1016/b978-0-08-051054-5.50005-4. 63

[34] Delphine Caruelle, Anders Gustafsson, Poja Shams, and Line Lervik-Olsen. The use of electrodermal activity (eda) measurement to understand consumer emotions – a literature review and a call for action. *Journal of Business Research*, 104:146–160, 11 2019. doi:10.1016/j.jbusres.2019.06.041. 40

[35] George Casella and Edward I George. *Explaining the Gibbs Sampler*, volume 46. 1992. doi:https://doi.org/10.2307/2685208. 47

[36] CDC. Exposome and exposomincs, 2015. URL: http://www.cdc.gov/niosh/topics/exposome/. 2

[37] John A. Chalmers, Daniel S. Quintana, Maree J.Anne Abbott, and Andrew H. Kemp. Anxiety disorders are associated with reduced heart rate variability: A meta-analysis. *Frontiers in Psychiatry*, 5, 2014. URL: /pmc/articles/PMC4092363//pmc/articles/PMC4092363/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4092363/, doi:10.3389/fpsyt.2014.00080. 40

[38] Vikhyat Chaudhry. Arduair: Air quality monitoring, 2013. URL: http://www.ripublication.com/ijeem.htm. 2

[39] Yanguang Chen. A new methodology of spatial cross-correlation analysis. *PLoS ONE*, 10, 5 2015. doi:10.1371/journal.pone.0126158. 122

[40] Seungeun Chung, Jiyoun Lim, Kyoung Ju Noh, Gague Kim, and Hyuntae Jeong. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors (Basel, Switzerland)*, 19, 4 2019. URL: /pmc/articles/PMC6479605//pmc/articles/PMC6479605/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6479605/, doi:10.3390/S19071716. 3

[41] S. Clement, O. Schauman, T. Graham, F. Maggioni, S. Evans-Lacko, N. Bezborodovs, C. Morgan, N. Rüsch, J. S.L. Brown, and G. Thornicroft. What is the impact of mental health-related stigma on help-seeking?

a systematic review of quantitative and qualitative studies. *Psychological medicine*, 45:11–27, 1 2015. URL: https://pubmed.ncbi.nlm.nih.gov/24569086/, doi:10.1017/S0033291714000129. 54

[42] Pau Climent-Pérez, Ángela M. Muñoz-Antón, Angelica Poli, Susanna Spinsante, and Francisco Florez-Revuelta. Dataset of acceleration signals recorded while performing activities of daily living. *Data in Brief*, 41:107896, 4 2022. doi:10.1016/J.DIB.2022.107896. 38

[43] João Rala Cordeiro, António Raimundo, Octavian Postolache, and Pedro Sebastião. Neural architecture search for 1d cnns. different approaches tests and measurements. *Sensors*, 21, 12 2021. doi:10.3390/s21237990. 42

[44] Nottingham City Council. Air pollution and air quality - nottingham city council, 2023. URL: https://www.nottinghamcity.gov.uk/information-for-business/environmental-health-and-safer-housing/environmental-health-and-safer-places/air-pollution-and-air-quality/. 72

[45] Daniel T.C. Cox, Danielle F. Shanahan, Hannah L. Hudson, Richard A. Fuller, and Kevin J. Gaston. The impact of urbanisation on nature dose and the implications for human health. *Landscape and Urban Planning*, 179:72–80, 11 2018. doi:10.1016/J.LANDURBPLAN.2018.07.013. 1, 15

[46] Robert A Cummins and F A S Ps. *Personal Wellbeing Index-Adult (PWI-A) (English) 5 th Edition The International Wellbeing Group MANUAL 2013 Personal Wellbeing Index-Adult*. 2013. 27, 66, 76, 85

[47] Dongjuan Dai, Aaron J. Prussin, Linsey C. Marr, Peter J. Vikesland, Marc A. Edwards, and Amy Pruden. Factors shaping the human exposome in the built environment: Opportunities for engineering control. *Environmental Science and Technology*, 51:7759–7774, 2017. URL: https://pubs.acs.org/sharingguidelines, doi:10.1021/acs.est.7b01097. 30, 37, 54

[48] Antonio R. Damasio. Emotion in the perspective of an integrated nervous system. *Brain Research Reviews*, 26:83–86, 1998. doi:10.1016/S0165-0173(97)00064-7. 33

[49] Damião Ribeiro de Almeida, Cláudio de Souza Baptista, Fabio Gomes de Andrade, and Amilcar Soares. A survey on big data for trajectory analytics. *ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 88*, 9:88, 2 2020. URL: https://www.mdpi.com/2220-9964/9/2/88/htmhttps://www.mdpi.com/2220-9964/9/2/88, doi:10.3390/IJGI9020088. 23, 25

[50] Jeroen J. de Hartog, Timo Lanki, Kirsi L. Timonen, Gerard Hoek, Nicole A.H. Janssen, Angela Ibald-Mulli, Annette Peters, Joachim Heinrich, Tuula H. Tarkiainen, Rene van Grieken, Joop H. van Wijnen, Bert Brunekreef, and Juha Pekkanen. Associations between pm2.5 and heart rate variability are modified by particle composition and beta-blocker use in patients with coronary heart disease. *Environmental Health Perspectives*, 117:105–111, 2009. URL: https://pubmed.ncbi.nlm.nih.gov/19165395/, doi:10.1289/ehp.11062. 117

[51] Lianne P. de Vries, Bart M.L. Baselmans, and Meike Bartels. Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies, 6 2021. doi:10.1007/s10902-020-00324-7. 36, 121

[52] D. Gayle DeBord, Tania Carreón, Thomas J. Lentz, Paul J. Middendorf, Mark D. Hoover, and Paul A. Schulte. Use of the ”exposome” in the practice of epidemiology: A primer on -omic technologies. *American Journal of Epidemiology*, 184:302–314, 8 2016. doi:10.1093/aje/kwv325. 18, 20, 21, 37, 54

[53] Jan Kleine Deters, Rasa Zalakeviciute, Mario Gonzalez, and Yves Rybarczyk. Modeling pm2.5 urban pollution using machine learning and selected

meteorological parameters. *Journal of Electrical and Computer Engineering*, 2017, 2017. doi:10.1155/2017/5106045. 63

[54] Shivani Dhall, B. R. Mehta, A. K. Tyagi, and Kapil Sood. A review on environmental gas sensors: Materials and technologies, 1 2021. doi:10.1016/j.sintl.2021.100116. 31

[55] Adam Dobrin. A review of properties and variations of voronoi diagrams, 2005. URL: http://www.whitman.edu/mathematics/SeniorProjectArchive/2005/dobrinat.pdf. 125

[56] Patrycja Dolibog, Barbara Pietrzyk, Klaudia Kierszniok, and Krzysztof Pawlicki. Comparative analysis of human body temperatures measured with noncontact and contact thermometers. *Healthcare*, 10, 2 2022. URL: /pmc/articles/PMC8871951//pmc/articles/PMC8871951/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8871951/, doi:10.3390/HEALTHCARE10020331. 41

[57] David Donaire-Gonzalez, Antònia Valentín, Erik van Nunen, Ariadna Curto, Albert Rodriguez, Mario Fernandez-Nieto, Alessio Naccarati, Sonia Tarallo, Ming Yi Tsai, Nicole Probst-Hensch, Roel Vermeulen, Gerard Hoek, Paolo Vineis, John Gulliver, and Mark J. Nieuwenhuijsen. Expoapp: An integrated system to assess multiple personal environmental exposures. *Environment International*, 126:494–503, 2019. doi:10.1016/j.envint.2019.02.054. 5, 17, 53, 117, 121

[58] TSI DustTrak. Tsi dusttrak environmental mcerts dust monitor - rental, hire .... URL: https://www.ashtead-technology.com/product/tsi-dusttrak-environmental-8543-mcerts-outdoor-dust-aerosol-monitor. 28

[59] Angel M. Dzhambov, Iana Markevych, Boris Tilov, Zlatoslav Arabadzhiev, Drozdstoj Stoyanov, Penka Gatseva, and Donka D. Dimitrova. Lower noise annoyance associated with gis-derived greenspace: Pathways

through perceived greenspace and residential noise. *International Journal of Environmental Research and Public Health*, 15, 7 2018. URL: /pmc/articles/PMC6068578//pmc/articles/PMC6068578/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6068578/, doi:10.3390/IJERPH15071533. 131

[60] Paul Ekman and Wallace V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17:124–129, 2 1971. URL: /record/1971-07999-001, doi:10.1037/H0030377. 34

[61] Empatica. E4 wristband | real-time physiological signals | wearable ppg, eda, temperature, motion sensors, 2020. URL: https://www.empatica.com/research/e4. 38

[62] Public Health England. Air pollution: applying all our health - gov.uk, 2 2022. URL: https://www.gov.uk/government/publications/air-pollution-applying-all-our-health/air-pollution-applying-all-our-health. 1, 2

[63] Ying Xing Feng, Nur Syahirah Roslan, Lila Iznita Izhar, Muhammad Abdul Rahman, Ibrahima Faye, and Eric Tatt Wei Ho. Conversational task increases heart rate variability of individuals susceptible to perceived social isolation. *International Journal of Environmental Research and Public Health*, 18:9858, 9 2021. doi:10.3390/IJERPH18189858/S1. 2

[64] Giuliana Ferrante, Salvatore Fasola, Giovanna Cilluffo, Giorgio Piacentini, Giovanni Viegi, and Stefania La Grutta. Addressing exposome: An innovative approach to environmental determinants in pediatric respiratory health. *Frontiers in Public Health*, 10:871140, 6 2022. URL: /pmc/articles/PMC9237327//pmc/articles/PMC9237327/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9237327/, doi:10.3389/FPUBH.2022.871140. 2, 22

[65] A. Fino. Air quality legislation. *Encyclopedia of Environmental Health*, pages 61–70, 1 2019. doi:10.1016/B978-0-12-409548-9.11045-0. 31

[66] Food and Rural Affairs Department for Environment. Report a guide for local authorities purchasing air quality monitoring equipment. 2006. 1

[67] Food and Rural Affairs Department for Environment. Automatic urban and rural network (aurn)- defra, uk, 2018. URL: https://uk-air.defra.gov.uk/networks/network-info?view=aurn. 29, 74, 78

[68] Food and Rural Affairs (Defra) webmaster@defra.gsi.gov.uk Department for Environment. Automatic urban and rural network (aurn)- defra, uk. 2022. 5, 51

[69] Office for National Statistics. Road transport and air emissions, 2018. URL: https://www.ons.gov.uk/economy/environmentalaccounts/articles/roadtransportandairemissions/2019-09-16. 1

[70] National Institute for Occupational Safety and Health. Exposome and exposomics | niosh | cdc, 8 2022. URL: https://www.cdc.gov/niosh/topics/exposome/default.html. 22, 142

[71] Mental Health Foundation. Economic and social costs: statistics, 2023. URL: https://www.mentalhealth.org.uk/explore-mental-health/statistics/economic-social-costs-statistics. 2

[72] Barbara L Fredrickson. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, 56:218–226, 2001. doi:10.1037/0003-066X.56.3.218. 33

[73] Appasani Geetha, Pasumarthi Sai Ramya, Chenikala Sravani, and M. Ramesh. Real time air quality index from various locations. *International Journal of Recent Technology and Engineering (IJRTE)*, 9:368–372, 2020. doi:10.35940/ijrte.b3493.079220. 78

[74] Gregor Geršak and Janko Drnovšek. Electrodermal activity patient simulator. *PLoS ONE*, 15, 2 2020. URL: /pmc/articles/PMC7001969//pmc/articles/PMC7001969/?report=abstracthttps://www.ncbi.nlm.nih.

gov/pmc/articles/PMC7001969/, doi:10.1371/journal.pone.0228949. 40

[75] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. *Nature*, 29:1–73, 2016. URL: https://www.deeplearningbook.org/http://deeplearning.net/. 43

[76] Jack M. Gorman and Richard P. Sloan. Heart rate variability in depressive and anxiety disorders. *American Heart Journal*, 140, 2000. doi:10.1067/mhj.2000.109981. 40

[77] GOV.UK. Ammonia: general information - gov.uk, 2022. 31

[78] Grove. Grove - air quality sensor v1.3 - sensors - components. URL: https://store.arduino.cc/grove-air-quality-sensor-v1-3. 31

[79] Grove. Grove - pir motion sensor - seeed wiki. URL: http://wiki.seeedstudio.com/Grove-PIR_Motion_Sensor/. 41

[80] Sandeep Grover, Siddharth Sarkar, and Rahul Gupta. Data handling for e-mental health professionals. *Indian Journal of Psychological Medicine*, 42:85S, 10 2020. URL: /pmc/articles/PMC7736735//pmc/articles/PMC7736735/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7736735/, doi:10.1177/0253717620956732. 50

[81] Andrea Guidi, Antonio Lanata, Paolo Baragli, Gaetano Valenza, and Enzo Pasquale Scilingo. A wearable system for the evaluation of the human-horse interaction: A preliminary study. *Electronics (Switzerland)*, 5, 12 2016. doi:10.3390/ELECTRONICS5040063. 49

[82] H. F. Guite, C. Clark, and G. Ackrill. The impact of the physical and urban environment on mental well-being. *Public Health*, 120:1117–1126, 2006. doi:10.1016/j.puhe.2006.10.005. 15, 17, 18, 31, 75, 133

[83] Susan Guthrie, Sarah Giles, Fay Dunkerley, Hadeel Tabaqchali, Amelia Harshfield, Becky Ioppolo, and Catriona Manville. *The impact of ammonia*

emissions from agriculture on biodiversity An evidence synthesis. RAND Europe, 2018. 31

[84] Hebba Haddad and Audrey de Nazelle. The role of personal air pollution sensors and smartphone technology in changing travel behaviour. *Journal of Transport and Health*, 11:230–243, 12 2018. doi:10.1016/j.jth.2018.08.001. 121, 122

[85] Nadine Haddad, Xanthi D. Andrianou, and Konstantinos C. Makris. A scoping review on the characteristics of human exposome studies. *Current Pollution Reports*, 5:378–393, 2019. doi:10.1007/s40726-019-00130-7. 19, 20

[86] Md Junayed Hasan, Muhammad Sohaib, and Jong Myon Kim. 1d cnn-based transfer learning model for bearing fault diagnosis under variable working conditions. *Advances in Intelligent Systems and Computing*, 888:13–23, 2019. URL: https://link.springer.com/chapter/10.1007/978-3-030-03302-6_2, doi:10.1007/978-3-030-03302-6_2/FIGURES/8. 43

[87] Oura Health. Accurate health information accessible to everyone, 2022. URL: https://ouraring.com/. 39

[88] Lorenz M Hilty. Ethical issues in ubiquitous computing-three technology assessment studies revisited some of the authors of this publication are also working on these related projects: Rfid implications for recycling and waste treatment view project assessing the environmental impact of electronic vs. print media view project. 2015. URL: http://link.springer.com/bookseries/11156, doi:10.5167/uzh-109998. 50

[89] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 11 1997. doi:10.1162/NECO.1997.9.8.1735. 43

[90] Ko Ming Hsiao, Geoff West, Svetha Venkatesh, and Mohan Kumar. Temporal data fusion in multisensor systems using dynamic time warping temporal data fusion in multisensor systems using dynamic time warping department of computer science and engineering , the university of texas at arlington box. 2012. 49

[91] Nadine Humbel, Nadine Messerli-Bürgy, Kathrin Schuck, Andrea Wyssen, David Garcia-Burgos, Esther Biedert, Julia Lennertz, Andrea H. Meyer, Katherina Whinyates, Bettina Isenschmid, Gabriella Milos, Stephan Trier, Dirk Adolph, Jan Cwik, Jürgen Margraf, Hans Jörg Assion, Tobias Teismann, Bianca Ueberberg, Georg Juckel, Judith Müller, Benedikt Klauke, Silvia Schneider, and Simone Munsch. Self-reported emotion regulation difficulties are associated with mood but not with the biological stress response to thin ideal exposure. *PLoS ONE*, 13:e0199769, 6 2018. doi:10.1371/journal.pone.0199769. 39, 121

[92] Ahmed Ibrahim, Heng Zhang, Sarah Clinch, and Simon Harper. From gps to semantic data: how and why—a framework for enriching smartphone trajectories. *Computing 2021*, pages 1–25, 8 2021. URL: https://link.springer.com/article/10.1007/s00607-021-00993-z, doi:10.1007/S00607-021-00993-Z. 25

[93] Kamil K. Imbir. Affective norms for 718 polish short texts (anpst): Dataset with affective ratings for valence, arousal, dominance, origin, subjective significance and source dimensions. *Frontiers in Psychology*, 7, 7 2016. doi:10.3389/fpsyg.2016.01030. 66

[94] Fitri Indra Indikawati, Sri Winiarti, Ahmad Dahlan, Ringroad Selatan, and Di Yogyakarta. Stress detection from multimodal wearable sensor data. 2020. doi:10.1088/1757-899X/771/1/012028. 3, 6

[95] Sharifah Noor Masidayu Sayed Ismail, Nor Azlina Nor, and Siti Zainab Ibrahim. A comparison of emotion recognition system using electrocardiogram (ecg) and photoplethysmogram (ppg). *Journal of King Saud*

University - Computer and Information Sciences, 34:3539–3558, 6 2022. doi:10.1016/J.JKSUCI.2022.04.012. 41

[96] Viniece Jennings and Omoshalewa Bamkole. The relationship between social cohesion and urban green space: An avenue for health promotion. *International Journal of Environmental Research and Public Health*, 16, 1 2019. doi:10.3390/ijerph16030452. 18

[97] H I Ji. Ambient air quality and health, 2018. URL: http://www.who.int/mediacentre/factsheets/fs313/en/. 17

[98] Congfeng Jiang, Yeliang Qiu, Honghao Gao, Tiantian Fan, Kangkang Li, and Jian Wan. An edge computing platform for intelligent operational monitoring in internet data centers. *IEEE Access*, 7:133375–133387, 2019. doi:10.1109/ACCESS.2019.2939614. 118

[99] Ian T. Jollife and Jorge Cadima. Principal component analysis: A review and recent developments, 4 2016. doi:10.1098/rsta.2015.0202. 62, 104

[100] Eiman Kanjo. Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications*, 15:562–574, 8 2010. doi:10.1007/S11036-009-0217-Y. 3, 5, 18, 24, 53, 62, 75, 117, 118, 121, 122, 123, 133

[101] Eiman Kanjo, Eman M.G. Younis, and Chee Siang Ang. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, 49:46–56, 2019. doi:10.1016/j.inffus.2018.09.001. 37, 62, 63, 118

[102] Eiman Kanjo, Eman M.G. Younis, and Nasser Sherkat. Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. *Information Fusion*, 40:18–31, 2018. doi:10.1016/j.inffus.2017.05.005. 2, 3, 5, 6, 15, 17, 18, 31, 32, 37, 40, 41, 52, 62, 63, 65, 81, 84, 85, 104, 107, 121, 130, 131

[103] Yoshiyuki Kasahara, Chihiro Yoshida, Masatoshi Saito, and Yoshitaka Kimura. Assessments of heart rate and sympathetic and parasympathetic nervous activities of normal mouse fetuses at different stages of fetal development using fetal electrocardiography. *Frontiers in Physiology*, 12:482, 4 2021. doi:10.3389/fphys.2021.652828. 40

[104] Marie Kiermeier and Martin Werner. Similarity search for spatial trajectories using online lower bounding dtw and presorting strategies 1 introduction. 2017. doi:10.4230/LIPIcs.TIME.2017.18. 49

[105] Hye Geum Kim, Eun Jin Cheon, Dai Seg Bai, Young Hwan Lee, and Bon Hoon Koo. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, 15:235, 3 2018. URL: /pmc/articles/PMC5900369//pmc/articles/PMC5900369/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900369/, doi:10.30773/PI.2017.08.17. 40

[106] Hannu Kinnunen, Aleksi Rantanen, Tuomas Kentt, and Heli Koskim ki. Feasible assessment of recovery and cardiovascular health: Accuracy of nocturnal hr and hrv assessed via ring ppg in comparison to medical grade ecg. *Physiological Measurement*, 41:04NT01, 5 2020. URL: https://iopscience.iop.org/article/10.1088/1361-6579/ab840ahttps://iopscience.iop.org/article/10.1088/1361-6579/ab840a/meta, doi:10.1088/1361-6579/ab840a. 39, 84

[107] Ming Hsiao Ko, Geoff West, Svetha Venkatesh, and Mohan Kumar. Using dynamic time warping for online temporal fusion in multisensor systems. *Information Fusion*, 9:370–388, 7 2008. doi:10.1016/J.INFFUS.2006.08.002. 49

[108] Willem J. Kop, Stephen J. Synowski, Miranda E. Newell, Louis A. Schmidt, Shari R. Waldstein, and Nathan A. Fox. Autonomic nervous system reactivity to positive and negative mood induction: The role of acute psychological

responses and frontal electrocortical activity. *Biological psychology*, 86:230, 3 2011. doi:10.1016/J.BIOPSYCHO.2010.12.003. 40

[109] Md Abdul Kuddus, Elizabeth Tynan, and Emma McBryde. Urbanization: A problem for the rich and the poor? *Public Health Reviews*, 41, 1 2020. doi:10.1186/s40985-019-0116-0. 1

[110] Azadeh Kushki, Jillian Fairley, Satyam Merja, Gillian King, and Tom Chau. Comparison of blood volume pulse and skin conductance responses to mental and affective stimuli at different anatomical sites. *Physiological Measurement*, 32:1529–1539, 2011. doi:10.1088/0967-3334/32/10/002. 40

[111] Azadeh Kushki, Jillian Fairley, Satyam Merja, Gillian King, and Tom Chau. Comparison of blood volume pulse and skin conductance responses to mental and affective stimuli at different anatomical sites. *Physiological measurement*, 32:1529, 2011. URL: /pmc/articles/PMC5028198//pmc/articles/PMC5028198/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5028198/, doi:10.1088/0967-3334/32/10/002. 41

[112] Dominika Kwasnicka, Dimitra Kale, Verena Schneider, Jan Keller, Bernard Yeboah-Asiamah Asare, Daniel Powell, Felix Naughton, Gill A ten Hoor, Peter Verboon, and Olga Perski. Systematic review of ecological momentary assessment (ema) studies of five public health-related behaviours: review protocol. *BMJ Open*, 11:46435, 2021. URL: http://dx.doi.org/10.1136/bmjopen-2020-046435, doi:10.1136/bmjopen-2020-046435. 35, 130

[113] Shinjae Kwon, Hojoong Kim, and Woon Hong Yeo. Recent advances in wearable sensors and portable electronics for sleep monitoring. *iScience*, 24:102461, 5 2021. doi:10.1016/J.ISCI.2021.102461. 38

[114] A. Larkin and P. Hystad. Towards personal exposures: How technology is changing air pollution and health research, 12 2017. doi:10.1007/s40572-017-0163-y. 5, 30, 37, 130

[115] Ute Latza, Silke Gerdes, and Xaver Baur. Effects of nitrogen dioxide on human health: Systematic review of experimental and epidemiological studies conducted between 2002 and 2006. *International Journal of Hygiene and Environmental Health*, 212:271–287, 5 2009. doi:10.1016/j.ijheh.2008.06.003. 31

[116] Patrick Laube. Computational movement analysis. 2014. URL: http://link.springer.com/10.1007/978-3-319-10268-9, doi:10.1007/978-3-319-10268-9. 23

[117] Sandra Laville. Air pollution a cause in girl's death, coroner rules in landmark case | london | the guardian. *The Guardian*, 2020. URL: https://www.theguardian.com/environment/2020/dec/16/girls-death-contributed-to-by-air-pollution-coroner-rules-in-landmark-case. 15, 30

[118] Sandra Laville. Air pollution a cause in girl's death, coroner rules in landmark case | london | the guardian. *The Guardian*, 2020. 17

[119] Byeong-Jae Lee, Bumseok Kim, and Kyuhong Lee. Air pollution exposure and cardiovascular disease. *Toxicological Research*, 30:71–75, 2014. doi:10.5487/TR.2014.30.2.71. 17, 30

[120] Hyun Seob Lee. Application of dynamic time warping algorithm for pattern similarity of gait. *Journal of Exercise Rehabilitation*, 15:526–530, 2019. doi:10.12965/jer.1938384.192. 49

[121] Seunggyu Lee, Hyewon Kim, Mi Jin Park, and Hong Jin Jeon. Current advances in wearable devices and their sensors in patients with depression. *Frontiers in Psychiatry*, 12:962, 6 2021. doi:10.3389/FPSYT.2021.672347/BIBTEX. 41

[122] Florian Lehmann and Daniel Buschek. Heartbeats in the wild: A field study exploring ecg biometrics in everyday life. 2020. URL: https://dx.doi.org/10.1145/3313831.3376536, doi:10.1145/3313831.3376536. 39

[123] Jos Lelieveld, Andrea Pozzer, Ulrich Pöschl, Mohammed Fnais, Andy Haines, and Thomas Münzel. Loss of life expectancy from air pollution compared to other risk factors: A worldwide perspective. *Cardiovascular Research*, 116:1910–1917, 9 2020. doi:10.1093/cvr/cvaa025. 17, 30

[124] Xianzhi Li, Qiao Yu, Bander Alzahrani, Ahmed Barnawi, Ahmed Alhindi, Daniyal Alghazzawi, and Yiming Miao. Data fusion for intelligent crowd monitoring and management systems: A survey, 2021. doi:10.1109/ACCESS.2021.3060631. 6

[125] Yunji Liang, Xiaolong Zheng, and Daniel D. Zeng. A survey on big data-driven digital phenotyping of mental health. *Information Fusion*, 52:290–307, 12 2019. doi:10.1016/j.inffus.2019.04.001. 60

[126] I. Mei Lin, Sheng Yu Fan, Ye Hsu Lu, Chee Siong Lee, Kuan Ta Wu, and Hui Jing Ji. Exploring the blood volume amplitude and pulse transit time during anger recall in patients with coronary artery disease. *Journal of Cardiology*, 65:50–56, 1 2015. URL: http://dx.doi.org/10.1016/j.jjcc.2014.03.012, doi:10.1016/j.jjcc.2014.03.012. 40

[127] Christine Lætitia Lisetti and Fatma Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals. *Eurasip Journal on Applied Signal Processing*, 2004:1672–1687, 2004. doi:10.1155/S1110865704406192. 37

[128] Christine Lætitia Lisetti and Fatma Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals, 9 2004. doi:10.1155/S1110865704406192. 41, 52, 110, 113

[129] Siyuan Liu, Shuhui Wang, and Qiang Qu. Trajectory mining. *Encyclopedia of GIS*, pages 2310–2313, 2017. URL: https://link.springer.com/referenceworkentry/10.1007/978-3-319-17885-1_1576, doi:10.1007/978-3-319-17885-1_1576. 26

[130] Miranda Loh, Dimosthenis Sarigiannis, Alberto Gotti, Spyros Karakitsios, Anjoeka Pronk, Eelco Kuijpers, Isabella Annesi-Maesano, Nour Baiz, Joana Madureira, Eduardo Oliveira Fernandes, Michael Jerrett, and John W. Cherrie. How sensors might help define the external exposome. *International Journal of Environmental Research and Public Health*, 14:343, 2017. doi:10.3390/ijerph14040434. 5, 21, 22

[131] Germaine M. Buck Louis, Melissa M. Smarr, and Chirag J. Patel. The exposome research paradigm: an opportunity to understand the environmental basis for human health and disease. *Current environmental health reports*, 4:89–98, 2017. doi:10.1007/s40572-017-0126-3. 19

[132] Hendramoorthy Maheswaran, Scott Weich, John Powell, and Sarah Stewart-Brown. Evaluating the responsiveness of the warwick edinburgh mental well-being scale (wemwbs): Group and individual level analysis. *Health and Quality of Life Outcomes*, 10, 12 2012. doi:10.1186/1477-7525-10-156. 36

[133] Panu Maijala, Zhao Shuyang, Toni Heittola, and Tuomas Virtanen. Environmental noise monitoring using source classification in sensors. *Applied Acoustics*, 129:258–267, 1 2018. doi:10.1016/j.apacoust.2017.08.006. 30, 58

[134] Léa Maitre, Jeroen De Bont, Maribel Casas, Oliver Robinson, Gunn Marit Aasvang, Lydiane Agier, Sandra Andrušaitytė, Ferran Ballester, Xavier Basagaña, Eva Borràs, Céline Brochot, Mariona Bustamante, Angel Carracedo, Montserrat De Castro, Audrius Dedele, David Donaire-Gonzalez, Xavier Estivill, Jorunn Evandt, Serena Fossati, Lise Giorgis-Allemand, Juan R. Gonzalez, Berit Granum, Regina Grazuleviciene, Kristine Bjerve Gützkow, Line Småstuen Haug, Carles Hernandez-Ferrer, Barbara Heude, Jesus Ibarluzea, Jordi Julvez, Marianna Karachaliou, Hector C. Keun, Norun Hjertager Krog, Chung Ho E. Lau, Vasiliki Leventakou, Sarah Lyon-Caen, Cyntia Manzano, Dan Mason, Rosemary McEachan, Helle Margrete Meltzer, Inga Petraviciene, Joane Quentin, Theano Roumeliotaki,

Eduard Sabido, Pierre Jean Saulnier, Alexandros P. Siskos, Valérie Siroux, Jordi Sunyer, Ibon Tamayo, Jose Urquiza, Marina Vafeiadi, Diana Van Gent, Marta Vives-Usano, Dagmar Waiblinger, Charline Warembourg, Leda Chatzi, Muireann Coen, Peter Van Den Hazel, Mark J. Nieuwenhuijsen, Rémy Slama, Cathrine Thomsen, John Wright, and Martine Vrijheid. *Human Early Life Exposome (HELIX) study: A European population-based exposome cohort*, volume 8, page e021311. BMJ Publishing Group, 9 2018. doi:10.1136/bmjopen-2017-021311. 2, 17, 18, 20

[135] Alexandru Ion Marinescu. Bach 2.0 - generating classical music using recurrent neural networks. *Procedia Computer Science*, 159:117–124, 1 2019. doi:10.1016/J.PROCS.2019.09.166. xi, 44

[136] Daisuke Mashima, Stephen Kobourov, and Yifan Hu. Visualizing dynamic data with maps. *IEEE Transactions on Visualization and Computer Graphics*, 18:1424–1437, 2012. doi:10.1109/TVCG.2011.288. 123

[137] Gavin McArdle, Urška Demšar, Stefan van der Spek, and Seán McLoone. Classifying pedestrian movement behaviour from gps trajectories using visualization and clustering. *Annals of GIS*, 20:85–98, 2014. doi:10.1080/19475683.2014.904560. 17, 24

[138] Michael J. McGrath, Cliodhna Ní Scanaill, Michael J. McGrath, and Cliodhna Ní Scanaill. *Environmental Monitoring for Health and Wellness*, pages 249–282. Apress, 2013. doi:10.1007/978-1-4302-6014-1_11. 5, 58

[139] Ravi Mehta, Rui(Juliet) Zhu, and Amar Cheema. Is noise always bad? exploring the effects of ambient noise on creative cognition. *Journal of Consumer Research*, 39:784–799, 2012. doi:10.1086/665048. 17, 131, 133

[140] Fatima A. Merchant, Shishir K. Shah, and Kenneth R. Castleman. Object measurement. *Microscope Image Processing, Second Edition*, pages 153–175, 1 2023. doi:10.1016/B978-0-12-821049-9.00017-4. 48

[141] Tauno Metsalu and Jaak Vilo. Clustvis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic acids research*, 43:W566–W570, 2015. URL: https://pubmed.ncbi.nlm.nih.gov/25969447/, doi:10.1093/NAR/GKV468. 62, 123

[142] Elena Mocanu, Phuong H. Nguyen, and Madeleine Gibescu. *Deep Learning for Power System Data Analysis*, pages 125–158. Elsevier, 1 2018. doi:10.1016/B978-0-12-811968-6.00007-3. 46

[143] movisens GmbH. movisens gmbh, 2022. URL: https://www.movisens.com/en/http://www.movisens.com/. 39

[144] M. M. Mukaka. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, 24:69, 2012. URL: /pmc/articles/PMC3576830//pmc/articles/PMC3576830/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/. 96

[145] Inez Myin-Germeys, Zuzana Kasanova, Thomas Vaessen, Hugo Vachon, Olivia Kirtley, Wolfgang Viechtbauer, and Ulrich Reininghaus. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*, 17:123, 6 2018. URL: /pmc/articles/PMC5980621//pmc/articles/PMC5980621/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5980621/, doi:10.1002/WPS.20513. 35

[146] Lars Müller, Verónica Rivera-Pelayo, Christine Kunzmann, and Andreas Schmidt. From stress awareness to coping strategies of medical staff: Supporting reflection on physiological data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7065 LNCS:93–103, 2011. URL: https://link.springer.com/chapter/10.1007/978-3-642-25446-8_11, doi:10.1007/978-3-642-25446-8_11/COVER/. 39

[147] Thomas Münzel, Tommaso Gori, Wolfgang Babisch, and Mathias Basner. Cardiovascular effects of environmental noise exposure, 2014. doi:10.1093/eurheartj/ehu030. 15, 17, 31, 108

[148] Chris Neale, Peter Aspinall, Jenny Roe, Sara Tilley, Panagiotis Mavros, Steve Cinderby, Richard Coyne, Neil Thin, and Catharine Ward Thompson. The impact of walking in different urban environments on brain activity in older people. *Cities Health*, 4:94–106, 2020. doi:10.1080/23748834.2019.1619893. 18

[149] Joseph Needham. Science and civilisation in china: Mathematics and the sciences of the heavens and the earth. *Cambridge University Press*, 3:147, 1959. 88

[150] Andreas Neumann. Spatial statistical visualization. *Encyclopedia of GIS*, pages 2072–2086, 2017. URL: https://link.springer.com/referenceworkentry/10.1007/978-3-319-17885-1_1525, doi:10.1007/978-3-319-17885-1_1525. 122

[151] Apollo Neuro. Science – apollo neuro, 2022. URL: https://apolloneuro.com/pages/science. 38

[152] Joanne B Newbury, Robert Stewart, Helen L Fisher, Sean Beevers, David Dajnak, Matthew Broadbent, Megan Pritchard, Narushige Shiode, Margaret Heslin, Ryan Hammoud, Matthew Hotopf, Stephani L Hatch, Ian S Mudway, and Ioannis Bakolis. Association between air pollution exposure and mental health service use among individuals with first presentations of psychotic and mood disorders: retrospective cohort study. 2021. URL: https://doi.org/10.1192/bjp.2021.119,, doi:10.1192/bjp.2021.119. 2

[153] Binh Nguyen, Martin Ivanov, Venkat Bhat, and Sri Krishnan. Digital phenotyping for classification of anxiety severity during covid-19. *Frontiers in Digital Health*, 4:203, 10 2022. doi:10.3389/FDGTH.2022.877762/BIBTEX. 58

[154] Vinh T. Nguyen, Kwanghee Jung, and Vibhuti Gupta. Examining data visualization pitfalls in scientific publications. *Visual Computing for Industry, Biomedicine, and Art*, 4:1–15, 12 2021. URL: https://vciba.springeropen.com/articles/10.1186/s42492-021-00092-y, doi:10.1186/S42492-021-00092-Y/TABLES/5. 122

[155] Mark J. Nieuwenhuijsen, David Donaire-Gonzalez, Maria Foraster, David Martinez, and Andres Cisneros. Using personal sensors to assess the exposome and acute health effects. *International Journal of Environmental Research and Public Health*, 11:7805–7819, 2014. doi:10.3390/ijerph110807805. 18, 21

[156] Ali Al Nima, Kevin M. Cloninger, Björn N. Persson, Sverker Sikström, and Danilo Garcia. Validation of subjective well-being measures using item response theory. *Frontiers in Psychology*, 10:1–33, 2020. doi:10.3389/fpsyg.2019.03036. 33

[157] Clemens Noelke, Mark McGovern, Daniel J. Corsi, Marcia P. Jimenez, Ari Stern, Ian Sue Wing, and Lisa Berkman. Increasing ambient temperature reduces emotional well-being. *Environmental Research*, 151:124–129, 11 2016. doi:10.1016/j.envres.2016.06.045. 32, 52

[158] Tales P. Nogueira, Reinaldo B. Braga, Carina T. de Oliveira, and Hervé Martin. Framestep: A framework for annotating semantic trajectories based on episodes. *Expert Systems with Applications*, 92:533–545, 2 2018. URL: https://doi.org/10.1016/j.eswa.2017.10.004, doi:10.1016/J.ESWA.2017.10.004. 25, 92

[159] Tales P. Nogueira, Hervé Martin, and Rossana M.C. Andrade. A statistical method for detecting move, stop, and noise episodes in trajectories. *Proceedings of the Brazilian Symposium on GeoInformatics*, 2017-Decem:210–221, 2017. 25, 133

[160] Mawutorli Nyarku, Mandana Mazaheri, Rohan Jayaratne, Matthew Dunbabin, Md Mahmudur Rahman, Erik Uhde, and Lidia Morawska. Mobile

phones as monitors of personal exposure to air pollution: Is this the future? *PLoS ONE*, 13, 2018. doi:10.1371/journal.pone.0193150. 30, 37

[161] Henrik Nygård, Soile Oinonen, Heidi A. Hällfors, Maiju Lehtiniemi, Eija Rantajärvi, and Laura Uusitalo. Price vs. value of marine monitoring. *Frontiers in Marine Science*, 3:205, 10 2016. doi:10.3389/FMARS.2016.00205/BIBTEX. 5

[162] The University of Warwick. Warwick-edinburgh mental well-being scale (wemwbs) user guide-version 2. 2015. URL: www.healthscotland.com/documents/458.aspx. 35

[163] Kelly Polido Kaneshiro Olympio, Fernanda Junqueira Salles, Ana Paula Sacone da Silva Ferreira, Elizeu Chiodi Pereira, Allan Santos de Oliveira, Isabelle Nogueira Leroux, and Flávia Bosquê Alves Vieira. The human exposome unraveling the impact of environment on health: Promise or reality? *Revista de Saude Publica*, 53, 2019. doi:10.11606/S1518-8787.2019053000649. 2, 20, 21

[164] World Health Organisation. Ambient air pollution, 2018. 30

[165] World Health Organisation. Urban health, 10 2021. URL: https://www.who.int/news-room/fact-sheets/detail/urban-health. 1

[166] World Health Organisation. Health and well-being, 2022. URL: https://www.who.int/data/gho/data/major-themes/health-and-well-being. 33

[167] World Health Organisation. Urban health, 2023. URL: https://www.who.int/health-topics/urban-health#tab=tab_1. 1

[168] Oura. Shop oura ring, 2023. URL: https://ouraring.com/product/heritage-silver. 39

[169] Costas Panagiotakis, Nikos Pelekis, Ioannis Kopanakis, Emmanuel Ramasso, and Yannis Theodoridis. Segmentation and sampling of moving

object trajectories based on representativeness. segmentation and sampling of moving object trajectories based on representativeness.. ieee transactions on knowledge and data engineering segmentation and sampling of moving object trajectories based on representativeness. 24:1328–1343, 2011. URL: https://hal.archives-ouvertes.fr/hal-00585610, doi:10.1109/TKDE.2011.39ï. 25

[170] Kanawat Paoin, Kayo Ueda, Xerxes Tesoro Seposo, Junichiro Hayano, Ken Kiyono, Norihiro Ueda, Takashi Kawamura, Akiko Honda, and Hirohisa Takano. Association between pm2.5 exposure and heart rate variability for the patients with cardiac problems in japan. *Air Quality, Atmosphere and Health*, 13:339–347, 3 2020. doi:10.1007/s11869-020-00797-8. 108

[171] Reinhard Pekrun. Using self-report to assess emotions in education. *Methodological Advances in Research on Emotion and Education*, pages 43–54, 1 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-29049-2_4, doi:10.1007/978-3-319-29049-2_4/COVER. 35

[172] Livia Petrescu, Cătălin Petrescu, Ana Oprea, Oana Mitruț, Gabriela Moise, Alin Moldoveanu, and Florica Moldoveanu. Machine learning methods for fear classification based on physiological features. *Sensors (Basel, Switzerland)*, 21, 7 2021. doi:10.3390/s21134519. 37, 41

[173] Tuan D. Pham. Time–frequency time–space lstm for robust classification of physiological signals. *Scientific Reports*, 11, 12 2021. doi:10.1038/s41598-021-86432-7. 118

[174] Joseph Pizzorno and Walter Crinnion. Particulate matter is a surprisingly common contributor to disease. *Integrative Medicine: A Clinician's Journal*, 16:8, 8 2017. 30

[175] Daniel Plugge, Daniel Kübler, Prem Raj Neupane, Konstantin Olschofsky, and Laura Prill. Measurement, reporting, and verifications systems in forest assessment. *Tropical Forestry Handbook, Second Edition*, 1:839–882, 1 2016. doi:10.1007/978-3-642-54601-3_73. 127

[176] Wojciech Pokojski and Paulina Pokojska. Voronoi diagrams – inventor, method, applications. *Polish Cartographical Review*, 50:141–150, 12 2018. doi:10.2478/pcr-2018-0009. 125, 127

[177] Jyoti Prakash, Suprakash Chaudhury, and Kaushik Chatterjee. Digital phenotyping in psychiatry: When mental health goes binary. *Industrial Psychiatry Journal*, 30:191, 2021. URL: /pmc/articles/PMC8709510/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8709510/, doi:10.4103/IPJ.IPJ_223_21. 58

[178] Anjoeka Pronk, Miranda Loh, Eelco Kuijpers, Maria Albin, Jenny Selander, Lode Godderis, Manosij Ghosh, Roel Vermeulen, Susan Peters, Ingrid Sivesind Mehlum, Michelle C. Turner, Vivi Schlünssen, Marcel Goldberg, Manolis Kogevinas, Barbara N. Harding, Svetlana Solovieva, Tina Garani-Papadatos, Martie Van Tongeren, and Rob Stierum. Applying the exposome concept to working life health: The eu ephor project. *Environmental Epidemiology*, 6:E185, 4 2022. URL: https://journals.lww.com/environepidem/Fulltext/2022/04000/Applying_the_exposome_concept_to_working_life.2.aspx, doi:10.1097/EE9.0000000000000185. 22

[179] Vivian C. Pun, Justin Manjourides, and Helen Suh. Association of ambient air pollution with depressive and anxiety symptoms in older adults: Results from the nshap study. *Environmental Health Perspectives*, 125:342–348, 2017. doi:10.1289/EHP494. 30

[180] QuerGiorgio, DaftariJoshal, and RaoRamesh R. Heart rate wavelet coherence analysis to investigate group entrainment. *Pervasive and Mobile Computing*, 28:21–34, 6 2016. URL: https://dl.acm.org/doi/abs/10.1016/j.pmcj.2015.09.008, doi:10.1016/J.PMCJ.2015.09.008. 49

[181] Akash M R and Sai G Ananda Krishnan. Comparative analysis of heat maps over voronoi diagram in eye gaze data visualization. 2017. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8321960. 123

[182] David Rabin and Greg Siegle. S87. toward emotion prosthetics: Emotion regulation through wearable vibroacoustic stimulation. *Biological Psychiatry*, 83:S380–S381, 5 2018. doi:10.1016/J.BIOPSYCH.2018.02.978. 38

[183] Chotirat Ann Ratanamahatana and Eamonn Keogh. Making time-series classification more accurate using learned constraints. pages 11–22, 2004. doi:10.1137/1.9781611972740.2. 48

[184] ReliaSENS. Reliasens 19-15 environmental monitoring system, 2023. URL: https://www.eurotech.com/en/products/intelligent-sensors/environmental-monitoring-systems/reliasens-19-15. 28

[185] C RENSO. Semantic trajectories modeling and analysis. 2011. URL: http://www.uhasselt.be/Documents/datasim/Papers/Semantic-Trajectories-Modeling-and-Analysis.pdf. 24

[186] Aaron Reuben, Louise Arseneault, Andrew Beddows, Sean D Beevers, Terrie E Moffitt, Antony Ambler, and Rachel M Latham. Association of air pollution exposure in childhood and adolescence with psychopathology at the transition to adulthood. 4:1–14, 2021. doi:10.1001/jamanetworkopen.2021.7508. 17

[187] Ana Isabel Ribeiro, Carla Tavares, Alexandra Guttentag, and Henrique Barros. Association between neighbourhood green space and biological markers in school-aged children. findings from the generation xxi birth cohort. *Environment International*, 132:105070, 11 2019. doi:10.1016/J.ENVINT.2019.105070. 131

[188] Beate Ritz, Barbara Hoffmann, and Annette Peters. The effects of fine dust, ozone, and nitrogen dioxide on health. *Deutsches Arzteblatt International*, 116:881–886, 12 2019. doi:10.3238/arztebl.2019.0881. 31

[189] James A. Rosenthal. *Statistics and Data Interpretation*. 2012. 105

[190] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 12 1980. doi:10.1037/h0077714. xi, 33, 34

[191] David Scott. Q-q plots. URL: http://onlinestatbook.com/2/advanced_graphs/q-q_plots.html. 107

[192] Pavel Senin. Dynamic time warping algorithm review. *Science*, 2007:1–23, 2008. URL: http://129.173.35.31/~pf/Linguistique/Treillis/ReviewDTW.pdf. 48, 97

[193] Sensirion. Particulate matter sensor sps30 | sensirion. URL: https://www.sensirion.com/en/environmental-sensors/particulate-matter-sensors-pm25/. 30

[194] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine*, 14:410–417, 3 2010. doi:10.1109/TITB.2009.2036164. 3, 40

[195] Fred Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms, 9 2017. URL: /pmc/articles/PMC5624990//pmc/articles/PMC5624990/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/, doi:10.3389/fpubh.2017.00258. 40

[196] Siqing Shan, Xijie Ju, Yigang Wei, and Zijin Wang. Effects of pm2.5 on people's emotion: A case study of weibo (chinese twitter) in beijing. *International Journal of Environmental Research and Public Health*, 18, 5 2021. doi:10.3390/IJERPH18105422. 30, 63

[197] Nandita Sharma and Tom Gedeon. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, 108:1287–1301, 12 2012. doi:10.1016/j.cmpb.2012.07.003. 98, 112, 114

[198] Arthur A. Shaw and N. P. Gopalan. Finding frequent trajectories by clustering and sequential pattern mining. *Journal of Traffic and Transportation Engineering (English Edition)*, 1:393–403, 12 2014. doi:10.1016/S2095-7564(15)30289-0. 26

[199] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4:1–32, 2008. URL: https://www.gov.uk/guidance/ecological-momentary-assessment, doi:10.1146/ANNUREV.CLINPSY.3.022806.091415. 36

[200] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. Ecological momentary assessment. *Annual review of clinical psychology*, 4:1–32, 2008. URL: https://pubmed.ncbi.nlm.nih.gov/18509902/, doi:10.1146/ANNUREV.CLINPSY.3.022806.091415. 36

[201] Katharina Sielemann, Alenka Hafner, and Boas Pucker. The reuse of public datasets in the life sciences: Potential risks and rewards. *PeerJ*, 8, 2020. URL: /pmc/articles/PMC7518187//pmc/articles/PMC7518187/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7518187/, doi:10.7717/PEERJ.9954/SUPP-1. 130

[202] Thiago H. Silva, Pedro O.S. Vaz De Melo, Jussara M. Almeida, and Antonio A.F. Loureiro. Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21:42–51, 2014. doi:10.1109/MWC.2014.6757896. 18

[203] Valérie Siroux, Lydiane Agier, and Rémy Slama. The exposome concept: A challenge and a potential driver for environmental health research. *European Respiratory Review*, 25:124–129, 2016. doi:10.1183/16000617.0034-2016. 22

[204] Valérie Siroux, Lydiane Agier, and Rémy Slama. The exposome concept: A challenge and a potential driver for environmental health research. *Eu-*

ropean Respiratory Review, 25:124–129, 6 2016. doi:10.1183/16000617.0034-2016. 22

[205] Katharine A Smith, Charlotte Blease, Maria Faurholt-Jepsen, Joseph Firth, Tom Van Daele, Carmen Moreno, Stephane Mouchabac, John Torous, and Andrea Cipriani. Digital mental health: challenges and next steps open access introduction: State of the art and beyond in digital mental health, and current challenges. *BMJ Ment Health*, 26:1–7, 2023. URL: http://mentalhealth.bmj.com/, doi:10.1136/bmjment-2023-300670. 58

[206] Johannes Smolander, Matthias Dehmer, and Frank Emmert-Streib. Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders. *FEBS Open Bio*, 9:1232–1248, 7 2019. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/2211-5463.12652, doi:10.1002/2211-5463.12652. 46

[207] Jennifer Sorinas, Jose Manuel Ferrández, and Eduardo Fernandez. Brain and body emotional responses: Multimodal approximation for valence classification. *Sensors (Switzerland)*, 20, 1 2020. doi:10.3390/s20010313. 6, 41

[208] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola. Field calibration of a cluster of low-cost available sensors for air quality monitoring. part a: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, 215:249–257, 8 2015. doi:10.1016/J.SNB.2015.03.031. 5, 30

[209] Asimina Stamatelopoulou, D. Chapizanis, S. Karakitsios, P. Kontoroupis, D. N. Asimakopoulos, T. Maggos, and D. Sarigiannis. Assessing and enhancing the utility of low-cost activity and location sensors for exposure studies. *Environmental Monitoring and Assessment*, 190, 3 2018. doi:10.1007/s10661-018-6537-2. 5, 18

[210] Stephen A. Stansfeld and Mark P. Matheson. Noise pollution: Non-auditory effects on health. *British Medical Bulletin*, 68:243–257, 2003. doi:10.1093/bmb/ldg033. 17

[211] Susanne Steinle, Stefan Reis, Clive E. Sabel, Sean Semple, Marsailidh M. Twigg, Christine F. Braban, Sarah R. Leeson, Mathew R. Heal, David Harrison, Chun Lin, and Hao Wu. Personal exposure monitoring of pm2.5 in indoor and outdoor microenvironments. *Science of the Total Environment*, 508:383–394, 3 2015. doi:10.1016/j.scitotenv.2014.12.003. 30

[212] Sarah Stewart-Brown, Alan Tennant, Ruth Tennant, Stephen Platt, Jane Parkinson, and Scott Weich. Internal construct validity of the warwick-edinburgh mental well-being scale (wemwbs): A rasch analysis using data from the scottish health education population survey. *Health and Quality of Life Outcomes*, 7, 2 2009. doi:10.1186/1477-7525-7-15. 35

[213] Jeanette A. Stingone, Germaine M. Buck Louis, Shoji F. Nakayama, Roel C.H. Vermeulen, Richard K. Kwok, Yuxia Cui, David M. Balshaw, and Susan L. Teitelbaum. Toward greater implementation of the exposome research paradigm within environmental epidemiology. *Annual Review of Public Health*, 38:315–327, 2017. doi:10.1146/annurev-publhealth-082516-012750. 19, 20

[214] Arthur A. Stone, Saul Shiffman, Joseph E. Schwartz, Joan E. Broderick, and Michael R. Hufford. Patient non-compliance with paper diaries. *BMJ*, 324:1193–1194, 5 2002. URL: https://www.bmj.com/content/324/7347/1193https://www.bmj.com/content/324/7347/1193.abstract, doi:10.1136/BMJ.324.7347.1193. 36

[215] Jason D. Stone, Hana K. Ulman, Kaylee Tran, Andrew G. Thompson, Manuel D. Halter, Jad H. Ramadan, Mark Stephenson, Victor S. Finomore, Scott M. Galster, Ali R. Rezai, and Joshua A. Hagen. Assessing the accuracy of popular commercial technologies that measure resting heart rate

and heart rate variability. *Frontiers in Sports and Active Living*, 3, 2021. doi:10.3389/fspor.2021.585870. 39

[216] Waleed M. Sweileh, Samah W. Al-Jabi, Sa'Ed H. Zyoud, and Ansam F. Sawalha. Outdoor air pollution and respiratory health: A bibliometric analysis of publications in peer-reviewed journals (1900 - 2017), 6 2018. doi:10.1186/s40248-018-0128-5. 17

[217] Roberto Sánchez-Reolid, Francisco López de la Rosa, María T. López, and Antonio Fernández-Caballero. One-dimensional convolutional neural networks for low/high arousal classification from electrodermal activity. *Biomedical Signal Processing and Control*, 71:103203, 1 2022. doi:10.1016/J.BSPC.2021.103203. 38

[218] Jianhua Tao and Tieniu Tan. Affective computing: A review. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3784 LNCS:981–995, 2005. doi:10.1007/11573548_125. 2

[219] Yaguang Tao, Alan Both, Rodrigo I Silveira, Kevin Buchin, Stef Sijben, S Purves, Patrick Laube, Dongliang Peng, Kevin Toohey, and Matt Duckham. A comparative analysis of trajectory similarity measures. 62

[220] Romain Tavenard. Introduction to dynamic time warping, 2021. URL: https://rtavenar.github.io/blog/dtw.htmlhttp://www.mblondel.org/journal/2009/08/31/dynamic-time-warping-theory/. xi, 48

[221] Ruth Tennant, Louise Hiller, Ruth Fishwick, Stephen Platt, Stephen Joseph, Scott Weich, Jane Parkinson, Jenny Secker, and Sarah Stewart-Brown. The warwick-dinburgh mental well-being scale (wemwbs): Development and uk validation. *Health and Quality of Life Outcomes*, 5, 11 2007. doi:10.1186/1477-7525-5-63. 35

[222] Dušan B. Topalović, Miloš D. Davidović, Maja Jovanović, A. Bartonova, Z. Ristovski, and Milena Jovašević-Stojanović. In search of an opti-

mal in-field calibration method of low-cost gas sensors for ambient air pollutants: Comparison of linear, multilinear and artificial neural network approaches. *Atmospheric Environment*, 213:640–658, 9 2019. doi:10.1016/j.atmosenv.2019.06.028. 28

[223] John Torous, Mathew V. Kiang, Jeanette Lorme, and Jukka Pekka Onnela. New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health 2016;3(2):e16 https://mental.jmir.org/2016/2/e16*, 3:e5165, 5 2016. URL: https://mental.jmir.org/2016/2/e16, doi:10.2196/MENTAL.5165. 58

[224] Fraser Tough. Statistical tools in environmental impact assessment. 2013. URL: https://eleanor.lib.gla.ac.uk/record=b3004018. 62

[225] TSI. Aerosol and dust monitors | tsi, 2023. URL: https://tsi.com/products/aerosol-and-dust-monitors/aerosol-and-dust-monitors/. 5, 28

[226] Michelle C. Turner, Mark Nieuwenhuijsen, Kim Anderson, David Balshaw, Yuxia Cui, Genevieve Dunton, Jane A. Hoppin, Petros Koutrakis, and Michael Jerrett. Assessing the exposome with external measures: Commentary on the state of the science and research recommendations. *Annual Review of Public Health*, 38:215–239, 2017. doi:10.1146/annurev-publhealth-082516-012802. 2, 21

[227] Maximilian Ueberham and Uwe Schlink. Wearable sensors for multifactorial personal exposure measurements – a ranking study. *Environment International*, 121:130–138, 2018. doi:10.1016/j.envint.2018.08.057. 21, 37

[228] Paolo Vineis. What is the exposome and how it can help research on air pollution. *Emission Control Science and Technology*, 5:31–36, 2019. doi:10.1007/s40825-018-0104-8. 20, 21

[229] Paolo Vineis, Christiana A. Demetriou, and Nicole Probst-Hensch. Long-term effects of air pollution: an exposome meet-in-the-middle approach. *International Journal of Public Health 2020 65:2*, 65:125–127, 1 2020. URL: https://link.springer.com/article/10.1007/s00038-019-01329-7, doi:10.1007/S00038-019-01329-7. 21

[230] Cynthia Hudson Vitale and Heather Moulaison Sandy. Data management plans a review. *DESIDOC Journal of Library Information Technology*, 39:322–328, 12 2019. doi:10.14429/DJLIT.39.06.15086. 87

[231] Martine Vrijheid. The exposome: A new paradigm to study the impact of environment on health. *Thorax*, 69:876–878, 2014. doi:10.1136/thoraxjnl-2013-204949. 20

[232] Markus S. Wahl, Harald I. Muri, Rolf K. Snilsberg, Jacob J. Lamb, and Dag R. Hjelme. Temperature and humidity measurements. *Micro-Optics and Energy: Sensors for Energy Devices*, pages 31–43, 2020. doi:10.1007/978-3-030-43676-6_3. 32

[233] Di Wang, Tomio Miwa, and Takayuki Morikawa. Big trajectory data mining: A survey of methods, applications, and services. URL: www.mdpi.com/journal/sensors, doi:10.3390/s20164571. 23

[234] Yan Wang, Farhang Tahmasebi, Elizabeth Cooper, Samuel Stamp, Zaid Chalabi, Esfandiar Burman, and Dejan Mumovic. Exploring the relationship between window operation behavior and thermal and air quality factors: A case study of uk residential buildings. *Journal of Building Engineering*, 48:103997, 5 2022. doi:10.1016/J.JOBE.2022.103997. 67

[235] David Watson, Lee A. Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54:1063–1070, 1988. URL: https://pubmed.ncbi.nlm.nih.gov/3397865/, doi:10.1037//0022-3514.54.6.1063. 33, 35

[236] Julian Wienert, Tina Jahnel, and Laura Maaß. What are digital public health interventions? first steps toward a definition and an intervention classification framework. *Journal of Medical Internet Research*, 24, 6 2022. URL: /pmc/articles/PMC9277526//pmc/articles/PMC9277526/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9277526/, doi:10.2196/31921. 130

[237] Christopher Paul Wild. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers and Prevention*, 14:1847–1850, 2005. doi:10.1158/1055-9965.EPI-05-0456. 2, 19, 20

[238] Christopher Paul Wild. The exposome: From concept to utility. *International Journal of Epidemiology*, 41:24–32, 2012. doi:10.1093/ije/dyr236. 19, 20, 22, 54

[239] Osnat Wine, Alvaro Osornio Vargas, Sandra M. Campbell, Vahid Hosseini, Charles Robert Koch, and Mahdi Shahbakhti. Cold climate impact on air-pollution-related health outcomes: A scoping review. *International Journal of Environmental Research and Public Health*, 19, 2 2022. URL: /pmc/articles/PMC8835073//pmc/articles/PMC8835073/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8835073/, doi:10.3390/IJERPH19031473/S1. 139

[240] Kieran Woodward and Eiman Kanjo. ifidgetcube: Tangible fidgeting interfaces (tfis) to monitor and improve mental wellbeing. *IEEE Sensors Journal*, pages 1–1, 2020. doi:10.1109/jsen.2020.3031163. 42

[241] Kieran Woodward, Eiman Kanjo, David Brown, T. M. McGinnity, Becky Inkster, Donald J. Macintyre, and Athanasios Tsanas. Beyond mobile apps: A survey of technologies for mental well-being. *arXiv*, 2019. 37

[242] Kieran Woodward, Eiman Kanjo, David J. Brown, T. M. McGinnity, and Gordon Harold. In the hands of users with intellectual disabilities: co-designing tangible user interfaces for mental wellbeing. *Personal*

and Ubiquitous Computing 2023, pages 1–21, 10 2023. doi:10.1007/S00779-023-01752-X. 66

[243] Kieran Woodward, Eiman Kanjo, Andreas Oikonomou, and Samuel Burton. Emoecho: A tangible interface to convey and communicate emotions. pages 746–749. Association for Computing Machinery, Inc, 10 2018. doi:10.1145/3267305.3267705. 37

[244] Tao Wu, Jianxin Qin, and Yiliang Wan. Tost: A topological semantic model for gps trajectories inside road networks. *ISPRS International Journal of Geo-Information*, 8, 2019. doi:10.3390/ijgi8090410. 23

[245] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9:611–629, 8 2018. URL: https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9, doi:10.1007/S13244-018-0639-9/FIGURES/15. 42

[246] Zhixian Yan. Semantic trajectories: Computing and understanding mobility data. *Phd, Ecole Polytechnique Fédérale de Lausanne*, 5144, 2011. URL: http://infoscience.epfl.ch/record/167178. 25, 53

[247] Eman M.G. Younis, Eiman Kanjo, and Alan Chamberlain. Designing and evaluating mobile self-reporting techniques: crowdsourcing for citizen science. *Personal and Ubiquitous Computing*, 2019. doi:10.1007/s00779-019-01207-2. 121

[248] Antonella Zanobetti and Marie S O'neill. Longer-term outdoor temperatures and health effects: A review compliance with ethical standards human and animal rights and informed consent hhs public access. *Curr Epidemiol Rep*, 5:125–139, 2018. URL: https://www.ncbi.nlm.nih.gov/pubmed/. 32

[249] Nianyin Zeng, Han Li, and Yonghong Peng. A new deep belief network-based multi-task learning for diagnosis of alzheimer's disease. *Neural Computing and Applications*, 6 2021. doi:10.1007/S00521-021-06149-6. 46

[250] He Zhang, Ravi Srinivasan, and Vikram Ganesan. Low cost, multi-pollutant sensing system using raspberry pi for indoor air quality monitoring. *Sustainability (Switzerland)*, 13:1–15, 1 2021. doi:10.3390/SU13010370. 121

[251] Pei Zhang, Manish Arora, Romanas Chaleckis, Tomohiko Isobe, Mohit Jain, Isabel Meister, Erik Melén, Matthew Perzanowski, Federico Torta, Markus R. Wenk, and Craig E. Wheelock. Tackling the complexity of the exposome: Considerations from the gunma university initiative for advanced research (giar) exposome symposium. *Metabolites*, 9:106, 6 2019. URL: /pmc/articles/PMC6631702//pmc/articles/PMC6631702/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6631702/, doi:10.3390/metabo9060106. 22

[252] ZhengYu. Trajectory data mining. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6, 5 2015. URL: https://dl.acm.org/doi/abs/10.1145/2743025, doi:10.1145/2743025. 23

[253] Justas Šalkevicius, Robertas Damaševičius, Rytis Maskeliunas, and Ilona Laukienė. Anxiety level recognition for virtual reality therapy system using physiological signals. *Electronics (Switzerland)*, 8, 9 2019. doi:10.3390/electronics8091039. 40