UNIVERSITY OF LIVERPOOL

Detection, Prediction and Modelling of Mental Fatigue in Naturalistic Environment

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor of Philosophy

by

Hilal Al-Libawy

December 2017

Abstract

Operator mental fatigue in workplace can result in serious mistakes which have dangerous and life-threatening consequences. Fatigue assessment and prediction are, therefore, considered critical safety requirements that cut across modes and operations of numerous high-risk environments and industries such as nuclear and transportation. However, robust, accurate and timely assessment of fatigue (or alertness) is still a challenging task for many reasons. The majority of operator fatigue studies are still being carried out in simulation environments, overlooking operator's naturalistic behaviour and fatigue growth. Moreover, most of the available systems rely on using a single fatigue-related data source, which is clearly a major drawback that affects operation, performance, accuracy and reliability of the system in case this source fails. With multi-data sources in an integrated system, the system might stop working in the event of losing one or more data sources or at least becomes inaccurate or unreliable. Furthermore, paying no attention to human individual differences working as an operator in mission-critical jobs related to fatigue growth and in response to fatigue deleterious effect is another serious issue with the current fatigue assessment and prediction systems.

The research work presented in this thesis proposes a novel fatigue assessment approach, which addresses the aforementioned issues with fatigue detection and prediction system. This is achieved by developing and realising algorithms based on data collected from participants in naturalistic environments. Numerous experiments have been conducted to cover a wide range of fatigue-related tasks which are broadly grouped into two categories: biological and behavioural (performance) experiments. The biological-based experiments employ various data types such as heart rate, skin temperature, skin conductance and heart rate variability. These fatigue-related data types are used to build the proposed fatigue detection system, and the obtained results have demonstrated high accuracy and reliability (94.5% accuracy in naturalistic environments). The behavioural-based category includes two experiments: keyboard typing and driving task. The typing experiments have been carried out using computer keyboard and smartphone virtual keyboard, and have confirmed enhanced operator fatigue detection accuracy (94%). The driving experiments were conducted in naturalistic driving environments, and the used algorithms have demonstrated a new framework for

driver fatigue detection using smartphone inertial sensors based on a novel vehicle heading algorithm.

A prototype system was designed and built with a modular structure so as to allow the addition of multiple fatigue-related biological and behavioural sources. This modular structure was tested under different situations that involve losing one or more sources. In addition, the circadian rhythm, which is a main input to fatigue/alert regulators, was customised for each operator and modelled based on biological data collected from wearable devices. The constructed model captures individual differences of operators, which is a challenge in current systems. Such multi-source, modular and non-intrusive approach for fatigue/alertness assessment and prediction is expected to be of superior performance, low-cost and favourable by users compared to existing systems. Furthermore, it addresses other challenges of current fatigue systems by carrying out fatigue assessment in naturalistic environments and considering operator individual differences in response to fatigue. In addition, the modular structure of the proposed system helps improving robustness and accuracy against losing one or more input sources (accuracy for 4 sources: 91%, 3 sources: 87%, 2 sources: 77%). Following the proposed approach will definitely enhance the reliability of fatigue assessment systems, improve operator safety, productivity and reduce financial fatigue impacts. Moreover, the proposed system has proven to be non-intrusive in nature and of low implementation cost. The results obtained after testing the proposed system have been very promising to support the aforementioned benefits.

To my

Beloved mother

Darling wife

Lovely children

Great brothers

and

Wonderful sisters

Acknowledgements

I would like to express my gratefulness to all people who helped me achieve this work.

My deep thanks and appreciations go to my supervisor, Dr. Ali Al-Ataby, who supervised and guided this work; his advice and suggestions have been invaluable. Many thanks and gratitude go to Dr. Waleed Al-Nuaimy for his encouragement, advice and constructive suggestions. His support has been invaluable for me. Special thanks also go to Prof. Majid Al-Taee for his valuable advice.

My deepest gratitude goes to my lovely family: my mother for pleasant words, my wife for her unconditional love and support, great brothers and sister and to my children.

I gratefully acknowledge financial support from the University of Babylon, Ministry of Higher Education and Scientific Research, Iraq, that made this research possible and enable me to obtain this degree.

To all I would like to say, **thank you very much.**

Hilal Abbood Al-Libawy
Liverpool - UK
2017

Contents

List of Figures

Nomenclature

ACT-R	Adaptive Control of Thoughts -Rational
ANN	Artificial Neural Network
AUC	Area Under Curve
BC	Bayesian Combiner
BFM	Bio-mathematical Fatigue Model
CR	Circadian Rhythm
BKS_R	BacKSpace Rate
ECG	ElectroCardioGram
EGG	ElectroEncephaloGram
EMG	ElectroMyoGram
GSR	Galvanic Skin Response
GUI	Graphical User Interface
HF	High Frequency
HT	Hold Time
HR	Hold Rate
HRV	Heart Rate Variability
kNN	k-Nearest Neighbour
KSS	Karolniska Sleepiness Scale
KSPC	KeyStroke Per Character
LF	Low Frequency
PSO	Particle Swarm Optimisation
PVT	Psychomotor Vigilance Test
PSD	Power Spectral Density
ROC	Receiver Operating Characteristics
RR	Interval between two successive heart beats
RT	Reaction Time
SIM	Shape Invariant Model
SC	Skin Conductance
SVM	Support Vector Machine
SWM	Steering Wheel Movements
VAS	Visual Analogue Scale
WT	Wrist Temperature

Chapter 1

Introduction

1.1 Mental Fatigue

Mental fatigue is a feeling of tiredness or weakness which is usually accompanied by slow reaction time and may lead to negative outcomes and *"decrements in performance on tasks requiring alertness and the manipulation and retrieval of information stored in the memory"* [1]. It is mainly driven by the body natured circadian rhythm (daily cycle), sleep deprivation, time on task and type of task [2].

Research in mental fatigue is motivated by the fact that fatal errors made by human operators and drivers have dangerous and life-threatening consequences [3]. Fatigue detection is therefore considered a critical safety issue that cuts across numerous high-risk environment including; nuclear industry, healthcare, transportation, and others. Workplace safety and productivity have also been gaining increased attention in industry and insurance companies. In these, mental fatigue is considered as a main cause of degradation in the work safety. It can

lead to reduction in performance, productivity, and ultimately fatal errors and accidents [4]. The deleterious effect of fatigue on operator behaviour raises the need for fatigue management, which in turn requires a robust approach to detect, quantify and predict fatigue.

1.2 Current Solutions

Numerous methods have been used to detect and quantify operator fatigue. These methods are based on a number of operator fatigue symptoms such as yawing and slowed reflexes [3], [5]. Generally, operator fatigue can be monitored using [6], [7], [8]: (i) self-reporting, (ii) physiological data (i.e. heart rate, heart rate variability, skin temperature and skin conductivity) and (iii) behavioural changes of motor skill tasks that can be used to detect operator fatigue such as walking, typing, computer interactions or riding a bike.

Many research studies, ideas and methods for fatigue quantification have been investigated and implemented for academic and commercial goals. However, some of important fatigue symptoms have not yet been implemented in real systems. This mainly because existing measuring methods cannot stand the noisy environment of real world or due to practical challenges of using sensitive measurement devices in such environment.

1.3 Problem Statement and Motivation

1.3.1 Problem Statement

Many studies have sought to build a fatigue detection and assessment system. However, the existing systems are still facing several challenges. Of these, the real environment validation, personalisation and robustness are the most outstanding. These challenges are briefly outlined as follows.

a) The need for real life validation

From a practical perspective, validating a fatigue monitoring system in real world environment presents a major challenge [9]. The challenge arises due to health and safety concerns of conducting validation trials in the workplace especially at later stages of fatigue to avoid the risk of vigilance failure. This risk was reported in driving accidents when the driver sleeps behind steering wheel [10], [11].

b) Failure of data sources

Reliability and robustness of fatigue assessment systems are among the existing issues that need to be studied in more detail [12]. Statistics of road accidents highlighted the impact of poor quality fatigue assessment regimes on safety, public health and economics. This suggests an urgent need for reliable fatigue detection and assessment systems [13]. When compared to single-source, the multi-source and modular system designs also improves robustness and reliability. Existing fatigue detection systems which mainly depend on single fatigue-related metric (or multi-sensors included in an integrated system design) are generally prone to failure in the event of system failure.

c) Lack of personalisation

The reported fatigue assessment systems and research did not account for the individual differences of operators in term of their circadian rhythm and response to fatigue causes [14], [15], [16]. For example, the existing fatigue management systems and bio-mathematical fatigue models average the fatigue ratings across range of individuals while these individuals are clearly different from one another [17], [18]. Furthermore, the model does not offer information on group variance around the average performance [19]. Hence, *"Until all significant, individual fatigue-related factors are able to be modelled, absolute and individual measures of fatigue are unable to be provided by bio-mathematical models"* [20].

1.3.2 Motivation

Most academic research has been implemented in laboratory environments only without field testing while commercial solutions do not rely on solid theoretical background or do not have accredited validation requirement. The work presented in this thesis can be considered timely due to:

a) Emergence of low-cost and privacy preserving wearable devices which are capable of measuring fatigue-related metrics in real life.

b) The growing need for fatigue assessment solutions due to increasing of workplace fatigue-related accidents.

c) Lack of fatigue-related studies that are based on naturalistic environment. Most of existing academic researches were implemented in laboratory or simulator environment.

1.4 Aim and Objectives

The work reported in this thesis aims addressing the standing challenges of real-life validations, failure of data-source and lack of personalisation in the existing fatigue detection systems. This is achieved through research, design and development of a novel fatigue detection and prediction system. The main objectives of this research are summarised as follows:

a) Collecting a new fatigue-related datasets based on measurements carried in naturalistic environment.

b) Evaluating and validating the proposed system and algorithm in naturalistic environment by conducting a wide range of experiments involving real-world activities to check the ability of the proposed system to identify and quantify early signs of operator fatigue.

c) Improving system reliability by combining multiple data sources in a modular structure and test the system resilience against losing one or more data sources.

d) Personalising the circadian rhythm of operator by predicting fatigue-growth trend with time and building customising a model for each individual to build more accurate fatigue prediction system based on operator physiological parameters.

1.5 Novelty and Contributions

A novel fatigue detection system is proposed to detect and assess operator fatigue status. Unlike exiting systems, the proposed one is based on new models developed by the author. These models are based on physiological and behavioural data collected from real subjects in a naturalistic environment rather than using laboratory simulators. The developed system also addresses the individual differences between operators depending on their physiological and behavioural response to fatigue. In addition, robustness of the developed system is further improved by adopting a modular design approach that makes it self-immune system against the possibility of single-point failure. To the best of the author's knowledge, these novel attributes of the proposed fatigue detection system have not thoroughly explored in literature.

The contributions of this work include the following tasks:

a) Experimental assessment of the developed system and algorithm in naturalistic work environment validates the findings especially in fatigue area where human behaviour is sensitive to the environment within which it occurs.

b) Increasing in the acceptance of fatigue detection solutions by improving the robustness and reliability to identify and quantify the early signs of operator fatigue based on multiple data sources combined in a modular system structure.

c) Solution customisation for individual operators to overcome the inaccuracies which arise from generalisation and enhance prediction of the fatigue-growth trend with time.

The contribution of this work is supported by publishing high-quality publications in peer-reviewed journal and referred conferences. A list of publications is shown below.

1.6 Publications

[1] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, " A New Approach to Capturing Individual Differences in Circadian Rhythms Using Wearable Devices," *IET Systems Biology*, IET, (submitted: under review).

[2] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, "Detection of Early Signs of Driver Fatigue in Naturalistic Environments Using Smartphone Inertial Sensors," *IATSS Research*, Elsevier, (first revision).

[3] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, "Modular Design of Fatigue Detection in Naturalistic Driving Environments," *Accident Analysis & Prevention*, Elsevier, (first revision).

[4] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, "Multisensor Data Fusion in Fatigue Detection Using Wearable Devices," *Advances in Systems, Signals and Devices (ASSD)*, vol. 6 (SCI), 2018, (In Print). DOI 10.1515/ 9783110448399-012

[5] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, "A modular approach to personalise driver fatigue prediction," 6^{th} International Symposium on Naturalistic Driving Research (NDRS), Den Haag, Netherlands, Jun 2017.

[6] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and, M. A. Al-Taee and Ahmad M. Al-Taee, "A Fatigue Prediction Cognitive Model for Naturalistic Typing Environment ," 2017 10^{th} International Conference on Developments in eSystems Engineering (DeSE), Paris, France. Jun 2017.

[7] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, "Enhanced Operator Fatigue Detection Method Based on Computer-Keyboard Typing Style," 2016 14^{th} International Multi-Conference on Systems, Signals & Devices (SSD), Marrakech, Morocco, March 2017.

[8] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy and M. A. Al-Taee, "HRV-based operator fatigue analysis and classification using wearable sensors," 13^{th} International Multi-Conference on Systems, Signals & Devices (SSD), Leipzig, Germany, 2016, pp. 268-273. doi: 10.1109/SSD.2016.7473750

[9] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy, M. A. Al-Taee and Q. Al-Jubouri, "Fatigue Detection Method Based on Smartphone Text Entry Performance Metrics," 9^{th} International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK, 2016, pp. 40-44. doi: 10.1109/DeSE.2016.9

[10] **H. Al-Libawy**, A. Al-Ataby, W. Al-Nuaimy, M. A. Al-Taee and H. S. AlZu'bi, "Estimation of driver alertness using low-cost wearable devices," 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, 2015, pp. 1-5. doi: 10.1109/AEECT.2015.7360528 .

[11] **H. Al-Libawy**, W. Al-Nuaimy, A. Al-Ataby, S. A. Salem and H. S. AlZubi, "Prediction of driver fatigue: Approaches and open challenges," 2014 14th UK

Workshop on Computational Intelligence (UKCI), Bradford, UK, 2014, pp. 1-6. doi: 10.1109/UKCI.2014.6930193

[12] Q. Al-Jubouri , W. Al-Nuaimy, M. Al-Taee, B. Al-Saaidah, **H. Al-Libawy**, I. Young,"Identification and Extraction of New Feature for Zebrafish Discrimination ," in 9^{th} International Conference on the Developments on eSystems Engineering (DeSE), Liverpool, UK, IEEE, 2016.

[13] A. A. Ameri, C. Nichita, **H. Al-Libawy** and A. A. Atabi, "Fast Estimation Method for Selection of Optimal Distributed Generation Size Using Kalman Filter and Graph Theory," 2015 17^{th} UKSim-AMSS International Conference on Modelling and Simulation (UKSim), Cambridge, UK, 2015, pp. 420-425. doi: 10.1109/UKSim.2015.105

1.7 Thesis Organization

The remaining part of this thesis is organised as follows:

Chapter 2 lists the related work and state of the art with the mathematical and theoretical background and overviews the work outlined.

Chapter 3 describes biological-based fatigue symptoms and its effect on operator and the methods to quantify it using wearable devices for measuring fatigue-related metrics in real environments. This chapter also presents a new mathematical model for operator alertness that is based on bio-data collected from real subjects using low-cost wearable devices.

Chapter 4 continues with quantifying and monitoring fatigue symptoms in real environments but for behavioural symptoms of a typing task. This task is monitored for fatigue symptoms via typing style by conducting three experiments with three different datasets; smartphone typing, two-status computer-keyboard typing and multi-level status computer-keyboard typing

Chapter 5 discusses the last experiment of behavioural symptoms of capturing driver fatigue status based on their driving style. A new smartphone fatigue detection approach including an application for data collection is outlined in this chapter. In addition, new vehicle heading based on smartphone sensors is detailed.

Chapter 6 presents a scalable modular design approach to build a robust system using Bayesian combiner and particle swarm optimiser. This chapter shows the validation of system robustness against losing one or more of the input modules.

Chapter 7 concludes the work and presents some suggestions for areas of further investigations.

Chapter 2

Mental Fatigue Assessment

2.1 Introduction

Workplace safety and productivity have been gaining a growing attention in many industries and insurance companies. Mental fatigue is considered as a main factor of degradation of work safety and can lead to fatal errors and accidents [21] with reduction of performance and productivity impairment.

Many people are suffering from fatigue with different severity levels that are considered high in ratio range of 18.3-27% [22]. Different ratios (between 7% and 45%) of workers have been reported to be fatigued in industries depending on cognitive skills [23]. More accurate statistical results have also been reported in transportation industry in which fatigue is considered the second major cause of fatal crashes after speed [24]. For example, about 65% of accidents are attributed to driver fatigue [25]. On British roads, 25%-33% of all serious crashes are caused mainly by fatigue [26]. A similar situation is reported on Australian roads where

10%-17% of accidents are involving fatigue [8], [24]. Fatigue is the primary reason behind at least 100,000 crashes and more than 1,500 deaths annually in New Zealand. The estimated cost for these crashes is around $12.5 billion every year. Furthermore, the disastrous consequences of operator exhaustion are not limited to vehicle drivers; they also appeared to be one of major causes of fatal accidents in commercial aviation (i.e. air, sea, road and rail) [27].

The deleterious effect of fatigue on operator behaviour generally emphasises the need for fatigue risk assessment, which in turn requires a robust approach to manage, detect, quantify and predict fatigue.

This chapter presents theoretical foundation and state -of-the art review for mental fatigue symptoms and assessment techniques. It also presents an overview for the scope of work reported in this thesis.

2.2 Mental fatigue

There is no simple definition of fatigue as the definitions overlap with how to operationalise fatigue [28]. However, researchers agree on general aspects; regarding the harmful consequences of fatigue, as well as the basic definition that covers fatigue meaning in everyday use which is *"extreme tiredness arising from mental or physical effort"* [28].

Fatigue is a mental state and usually combined with slower response times. The circadian rhythm, sleep deprivation, time on task and type of task are considered the main drivers of this natural state. For example, air transportation is one of the fields that is thoroughly covered for risk assessment in which including fatigue and

sleepiness are considered as the main risk factors [29]. Generally, mental fatigue is a serious problem and results in operating errors, especially in transportation and aviation area. Although some studies suggest that aviation transportation is the safest form of transportation, human errors (including fatigue) remain the main cause of accidents [18], [30].

Fatigue manifests itself with signs such as: headache, slowed reflexes and reaction time (RT), reduced ability to pay attention to the situation at hand, impaired decision-making and judgement, sensation of weakness, feeling of inhibition, impaired activity and others. However, not all of these signs are measurable and can be used for fatigue quantification.

A wide variety of sleep research reported various assessment metrics and signs of fatigue. These metrics and signs are grouped under several categories such as subjective, objective (physiological and behavioural) and visual. The visual approach in which visual signs on driver face used to detect drowsiness uses several metrics [31] including eye closure [32], blinking rate [33], head nodding [34], yawning rate [35] and others. This approach has been considered accurate in driver fatigue assessment, but only at late stages of fatigue (sometimes can be too late). In addition, the way in which the visual sign is collected (via cameras installed inside the vehicle) may becreate privacy inconvenience thus degrading the acceptability level of this approach. The visual approach is out of the scope of this work.

The subjective and objective fatigue assessment methods are adopted in this work. Theoretical foundation and practical consideration of these methods are described in the next sections.

2.3 Subjective Fatigue Assessments

Subjective fatigue assessment has been carried out using different scales. In this section, the most common scales are described briefly as follows: [36]:

a) Visual analogue scales

Typically, Visual analogue scales (VAS) is a line of 10 cm long with labelled end points [37] as show in Figure 2.1. The subject needs to mark a sign on this line at appropriate distance between the two ends of the line. Although VAS is simple to use and sensitive to small changes, but this sensitivity has a negative impact on the assessment credibility due to individuals' differences.

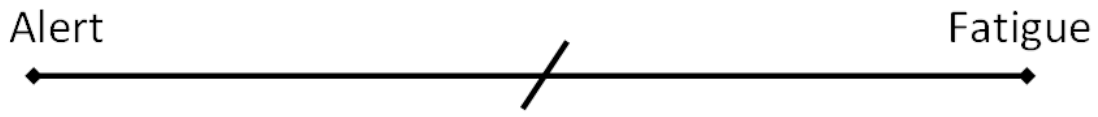


FIGURE 2.1: Example of VAS

b) Fatigue assessment scale

Fatigue Assessment Scale (FAS) is a questionnaire of ten statements. These statements are answered by the subject through selecting one category out of five [38]. As listed in Table 2.1, the ten statements investigate mental and physical status of a targeted subject and it can show, after been answered, the response of this subject to fatigue effect. While this response is usually used in self-assessment of health problems such as stroke [39], it can not be considered as an instantaneous mental fatigue quantification. Moreover, it is hard to use in many times in naturalistic environment due to practical issues.

c) Karolinska sleepiness scale

Table 2.1: FAS [38]

No	Question	cat 1	cat 2	cat 3	cat 4	cat 5
1	I am bothered by fatigue	1	2	3	4	5
2	I get tired very quickly	1	2	3	4	5
3	I do not do much during the day	1	2	3	4	5
4	I have enough energy for everyday life	1	2	3	4	5
5	Physically, I feel exhausted	1	2	3	4	5
6	I have problems to start things	1	2	3	4	5
7	I have problems to think clearly	1	2	3	4	5
8	I feel no desire to do anything	1	2	3	4	5
9	Mentally, I feel exhausted	1	2	3	4	5
10	When I am doing something, I can concentrate quite well	1	2	3	4	5

1 = Never, 2 = Sometimes; 3 = Regularly; 4 = Often and 5 = Always
cat= category.

Karolinska sleepiness scale (KSS) [40] is the most commonly used scales in fatigue studies although it is not the most precise scale. This is mainly due to ease of use and it does not disturb the subject especially in naturalistic environment. Table 2.2 lists the 9 levels of sleepiness status.

Table 2.2: Modified version of KSS [41]

Rate	Verbal description
1	Extremely alert
2	Very alert
3	Alert
4	Fairy alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep alert
8	Sleepy, some effort to keep alert
9	Very sleepy, great effort to keep alert, fighting sleep

Numerous specialised research centres have studied the awareness and signs of drowsiness; some of these signs are found to be strongly correlated to fatigue while others are not. Key symptoms of fatigue can be categorised into two major categories: physiological and behavioural symptoms. The symptoms are described in the next two sections.

2.4 Physiological Symptoms

A range of measures have been investigated in many research studies for an association with mental fatigue [42]. These include body temperature, heart rate and several indexes of heart rate variability, various variables calculated from Electroencephalography and others. Brief description of these measures provided as follows.

a) Electroencephalography

Electroencephalography (EEG) is a medical measuring technique that reads electrical brain activity. It is a scalp surface electrical record and it can be picked up using metal electrodes and conductive media [43]. It can be measured without penetrating the skin and is therefore non-invasive, safe and can be applied repeatedly without interfering with the activity carried out [44]. EEG is considered one of the most direct, valid and effective physiological measures for assessing the state of mental arousal. Its rhythms reflect the mental state of the brain, and the measured signals show fluctuation in its components for different mental states. Table 2.3 shows the different bands of the EEG signal. Experimental research associates mental fatigue with increased power in theta

(θ) and alpha (α) bands of EEG spectrum [45]. Despite that some studies are a reporting slight difference of band frequency boundaries, it is well accepted that EEG correlates with fluctuations in performance during sustained attention tasks [45], [46]. The measured EEG signal is generally weak and noisy, and thus it is not easy to extract this signal in naturalistic environment especially with reducing spacial resolution (i.e. using fewer electrodes) of mobile versions of EEG devices [47].

TABLE 2.3: EEG rhythm bands and related mental activities [41]

Band Name	Frequency Range (Hz)	Mental Activity
alpha (α)	8 -13	Relaxation
beta (β)	14 -26	Thinking and attention
theta (θ)	4 -7.5	Unconscious or deep meditation
delta (δ)	0.5 -4	Deep sleep
gamma (γ)	30 -45	Finger, toes and tongue movements

b) Core and skin temperatures

Core and skin temperatures of the human body are affected by internal and external factors. While the core temperature is regulated within a small range of variation and following the circadian rhythm, the skin temperature is considered another indication of the circadian rhythm, and it is almost a mirror of the core body temperature [48], [49]. In spite of the good correlation between body temperature and the circadian rhythm which in turn captures the fatigue development due to sleep pattern, only few researchers have investigated the effect of fatigue on the body temperature. The researchers agreed that the temperature during and after performing cognitive tasks tends to be higher [50], [51]. As cognitive tasks are also a cause of fatigue evolution, it can conclude that changes in mental fatigue reflect a corresponding change in the body temperature [52].

c) Skin conductivity

Skin conductance and resistance levels (also called galvanic skin repose (GSR)), are metrics of electrodermal activity of the human body. They are proved to have a correlation with fatigue, sleep deprivation and circadian rhythm [53], [54]. An example of skin resistance level variation during sleep deprivation experiment for 28 hours was previously reported in [53]. In this study, skin resistance level has increased with period of total sleep deprivation (main reason of mental fatigue). This metric has been widely gaining interest in recent years due to technology advances in developing wearable skin conductance measuring devices [55], [56].

d) Heart rate and heart rate variability

The correlation between heart rate and fatigue has not yet confirmed, especially when heart rate is used as the only metric [57]. Heart rate monitoring is therefore insufficient by itself for reliable monitoring of fatigue [50]. This conclusion is derived due to the sensitivity and complex nature of heart rate metric. However, other studies reported that operator heart rate decreases when experiencing a light workload during extended periods [58], [59]. Thus, a more robust fatigue metrics, called heart-rate variability (HRV), has been proposed instead.

HRV, the duration between two heartbeats, is typically obtained from the distance between two adjacent peaks in electrocardiogram (ECG) signal [60]. These peaks are a part of the QRS complex which is a name for the combination of three of a graphical deflections seen on a typical ECG. It corresponds to the depolarisation of the right and left ventricles of the human heart. An example of an ECG signal with some details about QRS complex and RR intervals is depicted in Figure 2.2. HRV indexes are calculated from the RR periods that

reflect the variation between heartbeats intervals.These can be categorised into five domains: time, frequency, complexity, fractal and nonlinear [61].

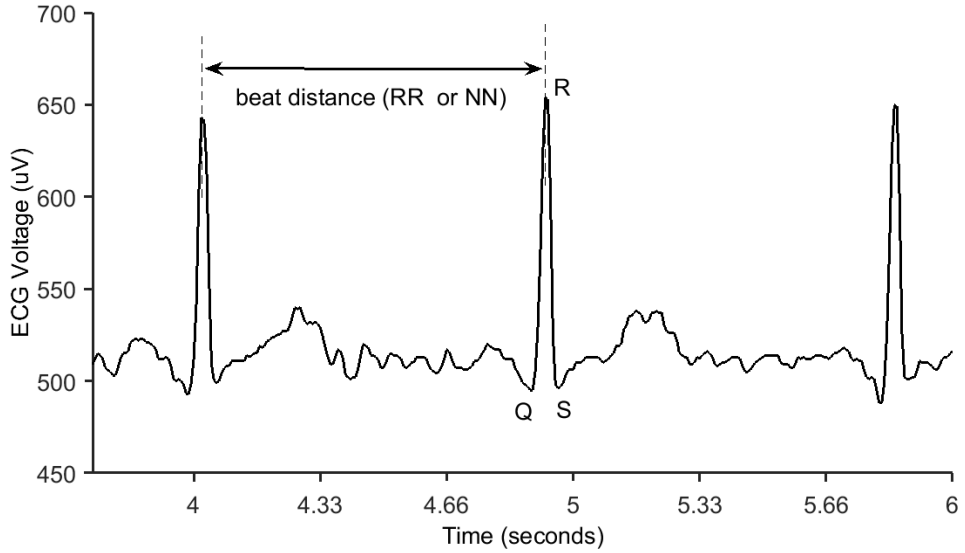


FIGURE 2.2: Example of ECG signal with QRS complex

A measurement period of at least 5 minutes is considered sufficient to calculate the HRV from ECG [62]. Table 2.4 summarises the indexes of common use in time and frequency domains. The relationship between the frequency power ranges of HRV investigated in [63]. It was reported that the ratio of low frequency (LF) to high frequency (HF) components is inversely proportional with fatigue evolution (i.e. high ratio reflects low-fatigue level and vice versa). The LF/HF is therefore considered of a particular interest in present work.

2.5 Behavioural and Performance Symptoms

Continuous performance of human is affected by fatigue [64]. Thus the lack of awareness of fatigue status may lead to serious workplace injuries and accidents. One of many definitions of fatigue [1]., that is related to performance impairment,

Table 2.4: HRV Indexes

Domain	Index	Description
Time	SDNN	Standard Deviation of NN (RR) intervals
	RMSSD	Root Mean Square of Successive Differences
	NN50	Number of pairs of successive NNs (RR) that differ by >50 ms
Frequency	VLF	Very Low Frequency power from 0.0033 - 0.04 Hz
	LF	Low Frequency power from 0.04 - 0.15 Hz
	HF	High Frequency power from 0.15 - 0.4 Hz
	LF/HF	Ratio of low to high frequency power

states that *"decrements in performance on tasks requiring alertness and the manipulation and retrieval of information stored in the memory."*

Most commonly practised activities such as walking, typing, driving, computer interaction, or riding a bike are typically known as motor skills. These skills need to be repeatedly practised in order to be performed automatically [65]. The effect of fatigue on these skills is studied in different work areas including medical surgeries [66], traffic safety and transportation area [67] and others. Behavioural symptoms such as response latency and variability, speed and accuracy, and decision-making and memory form a relation between fatigue and performance degradation and thus have been used as metrics for fatigue quantification [68]. These symptoms are typically measured in laboratory environments using psychomotor vigilance test [69].

2.5.1 Psychomotor vigilance test

Psychomotor skills are the skills behind movement-oriented activities [70], such as walking, climbing a stair, typing and steering a car. Proficiency of these skills

is significantly deteriorated by fatigue and sleep deprivation [66]. Hence, several metrics can be calculated from these skills for fatigue detection and quantification. Reaction time and lapses (failure to respond or late response) have been used as the main features for testing psychomotor performance degradation [69].

Psychomotor vigilance test (PVT) is an objective and instrument-based test. It measures attention and alertness based on the reaction-time metric [69]. In this test, subjects are asked to press a button in response to visual stimuli. Failure to react or any reaction exceeds 500 ms or beyond 100 ms is considered as lapses. The mean of reaction time and number of lapses with 2-10 minutes of the test duration are used in PVT to quantify operator fatigue [71]. Negative impact of the fatigue increases both features of the PVT: reaction time and number of lapses [72].

For a realistic environment,however, the text entry and vehicle driving tasks have been the most common tasks used to capture behavioural pattern changes of the operator. These tasks are considered of a particular interest in the present work and thus further detailed in the next two sections.

2.5.2 Text-entry task

Dealing with computing devices such as PCs , tablets and smartphones has been increasing dramatically. In these devices, the text entry is one of the main interaction activities performed by the between human. Typing speed and error rate are two major categories that are mainly used to assess typing accuracy and efficiency [73]. Timing events that are extracted from user interaction with such devices are called keystroke dynamics. These events contain many features which are believed to be measures of cognitive qualities [74]. As a psychomotor skill,

typing reflects performance degradation by increasing errors, key-down time and time between clicks [75], [73].

Nowadays,assessment of typing performance has become more and more difficult due to auto-correction utility and auto-prediction tools. New performance metrics (speed and accuracy)have therefore been developed depending on large experimental paired data of presented text (what participants are asked to enter) and transcribed text (typing output) [76].

Different reasons have motivated research on keystroke dynamics including quantifying typing performance [77], assessing continuous authentication [78] and detecting some diseases [75]. However, to our knowledge, just one study has been reported in the literature so far on fatigue detection using keyboard and mouse experiment [79]. The results reported in this study has demonstrated an increase in the key-down time and time between keys, and reduction in mouse speed upon fatigue inducing. Despite these important-findings, the fatigue assessment method of this study solely rely on time of the day. The individual differences between participants has not taken into consideration, and thus the reported results do not accurately reflects the actual fatigue level of participants .

2.5.3 Driving task

Mental fatigue increases the reaction time and may lead to falling asleep behind the steering wheel. Many studies have been investigated road safety [80]; of these, the driver-fatigue have been taking an important share of behavioural studies. The behavioural approach depends on indirectly collecting symptoms of fatigued drivers through their driving styles. Unlike the biological, this approach does

not directly collect drivers data that may violate their privacy. Assessment of drivers fatigue through detection of lane departure [81], [82], steering wheel movements (SWM) [83], [84] and symptoms measured by inertial sensors [85] can be considered common examples for the behavioural approach. However, the lane departure that depends on camera detection of road-marks is still facing some practical challenges including clarity of the road marks, weather circumstances and others. In addition, the SWM approach reported in [83], [84] was implemented and tested in a laboratory environment using a constructed car simulator with a single participant. This approach also requires a third-party licence to insert a new device in the car.

More recent works on fatigue detection systems have used multiple fatigue-related metrics to improve detection accuracy [5]. These systems fuse and combine the calculated metrics in three main levels. The first is a raw or filtered data level, which was used when different sensors measure the same parameter [86]. The second is a features level which detect fatigue status from the features even when they are extracted from different types of sensors [87]. The last level is the abstract (also called the decision level), which combines multi-module outputs to calculate a more accurate output. Several driver fatigue integrated systems were also developed through combining different classifiers decisions, using multiple fatigue-related metrics [88]. Despite importance of findings that were obtained from simulated environment, these decision-combining systems lack robustness testing when loosing one or more of the sources, and validation in real environments.

Some other experiments were also conducted to measure and quantify driver behaviour in real environments. A small fraction of these experiments focused

on driver fatigue in naturalistic environment [89], [90]. Main limitations of these works are: (i) measurement of fatigue-related metrics are impractical and (ii) need third-party devices to be installed on board.

2.6 Fatigue Modelling

Fatigue prediction has gained an increasing interest over the last few decades due to the need of accurate assessment of early onset of fatigue and to deliver better overview for fatigue management plans. Fatigue prediction needs robust mathematical modelling has challenges to generate accurate and reliable estimation based on objective measurements [91]. Fatigue modelling has been investigated using a bio-mathematical approach. This approach essentially depends on a two or three-process model of sleep regulation [92], or empirical model.

A number of bio-mathematical models of fatigue have been developed within the fatigue and sleep research communities. Table 2.5 lists number of existing alertness models that are based on two-process model and validated in different studies [16], [93], [94].

Table 2.5: Examples of existing BFM models

No.	BFM name
1	Circadian Alertness Simulator (CAS)
2	Fatigue Assessment by InterDynamic (FAID)
3	Advisory System for Tired Drivers (ASTiD)
4	Sleep, Activity, Fatigue and Task Effectiveness Mode (SAFTE)
5	System for Aircrew Fatigue Evaluation (SAFE)

Several other methods based on the two and three process models have also been reported including System for Aircrew Fatigue Evaluation (SAFE) [94] and Circadian Neurobehavioural Performance and Alertness (CNPA) [93]. However, depending on the application, each of these models has its own shortcomings. The major limitations are [16]: (i) lack of an estimate of group variance about the average performance prediction and (ii) unavailability of individual-difference parameters, such as age, morningness/eveningness, or sleep requirement for full performance. The prediction accuracy of individuals in these models is therefore inaccurate even when the fatigue trend is predicted correctly [91].

Bio-mathematical models are currently built on either two- or three-process models. The two-process model was first introduced to cover the sleep-wake regulation in 1982 by Alexander Borbely [95]. This model includes two processes: process S and process C which are outlined briefly as follows:

(a) Process S: A homoeostatic process which is an exponential growth of the sleep propensity with time awake and this will drop in same manner at the moment of sleeping.

(b) Process C: It represents the circadian rhythm of the biological clock of human. This part of the model is not related to sleep-wake activity. The Circadian rhythm is a physical and mental process that displays an endogenous oscillation of about 24 hours. A higher performance level can be achieved through sufficient duration and good quality/depth of sleep [96]. The circadian role in fatigue growth is clear in aviation industry since circadian rhythms of the pilots is shifted frequently when they travel across different time zones. It was shown that the circadian rhythm is one of the sources of individual differences [29].

A more recent but very similar model to two-process model called the mutual inhibition model or three-process model [97] has been proposed. This model is considered as an extension to the two-process model where the sleep inertia part referred to as (Process W) [98]. The sleep inertia represents the drop in performance that happens immediately after awaking. Two processes (S and C) were used in this model to regulate the arousal system and sleep promotion neurons. This model showed some improvements over the two-process model in explaining the low-vigilance levels directly after wakefulness, even with good quality of sleep.

Some of fatigue bio-mathematical models have also been developed based on empirical approaches from the biomedical research community data [93]. Other bio-mathematical models found that it is better to model the circadian part by 24-hour and 12-hour components [99]. The circadian rhythm depends mainly on one of the well-approved alertness biometric, which is the core temperature change around the day. It has two harmonics of sinusoidal components thus, it is called the two-harmonic model. It has the fundamental 24 hour sinusoidal and the second harmonic 12 hour period [99]..

2.7 Naturalistic vs Laboratory Environment

In studies where humans are the subject of experiments, including fatigue, the ability to generalise experimental findings beyond laboratory environment remains questionable [100]. Similarly, psychologists believe that [101] *"many small effects from the laboratory will turn out to be unreliable"* and conclude that *"a surprising*

number of laboratory findings may turn out to be affirmatively misleading about the nature of relations among variables outside the laboratory."

In a laboratory study, the participant's behaviour is usually different from that in naturalistic environments. This is mainly due to the participant's sensitivity to the environment in which the experiment is carried out. Thus, findings at naturalistic environments can play a key role in validating and interpreting the laboratory findings [102].

As humans are the main subjects of experiments and fatigue is an environmental-sensitive feeling, the naturalistic environment has been of particular interest in this work. Driving, keyboard typing and other naturalistic daily activities are considered the main tasks to asses human fatigue.

2.8 Work Overview

Unlike previously reported studies, which were mainly based on laboratory or simulator environment, a particular focus is given in this work on fatigue detection and modelling based on realistic environment. The individuality differences among different users has also been taken in consideration in the proposed models. Robustness against the partial loss of data resources is another important aspect of this work. These contribution pillars are shown in block diagram of Figure 2.3 and are outlined briefly as follows.

a) Fatigue Detection

Fatigue-related symptoms can be monitored using several metrics that are collected from real environment. In real environment, hardware and software carefully selected requirements need to be taken into consideration to collect biological and behavioural data of interest. Ease of use, mobility, accuracy, user privacy and acceptability are considered of a particular concern in assessing these requirements.

Two categories of experiments are carried out: biological and behavioural. The biological experiments are based on data collected directly from operator body such as heart rate, skin temperature and heart rate variability. In the behavioural experiments, the text-entry activity is identified as a new metric especially with smartphone using its inertial sensors. Furthermore, detecting fatigue symptoms through the smartphone inertial sensors is another novel contribution of this work in the context of driving task in naturalistic environment.

b) Modular design

A system architecture based on the modular design approach is proposed and implemented. This contributed in improving system reliability, scalability and robustness against single-point of failure that exist in equivalent systems.

c) Circadian customisation

New fatigue detection biological models are proposed and customised for each subject. These models are derived from heart-rate data that is collected by the author over several days. The collected dataset allowed personalising the operator's circadian rhythms, thus more realistic and accurate fatigue detection are made possible.

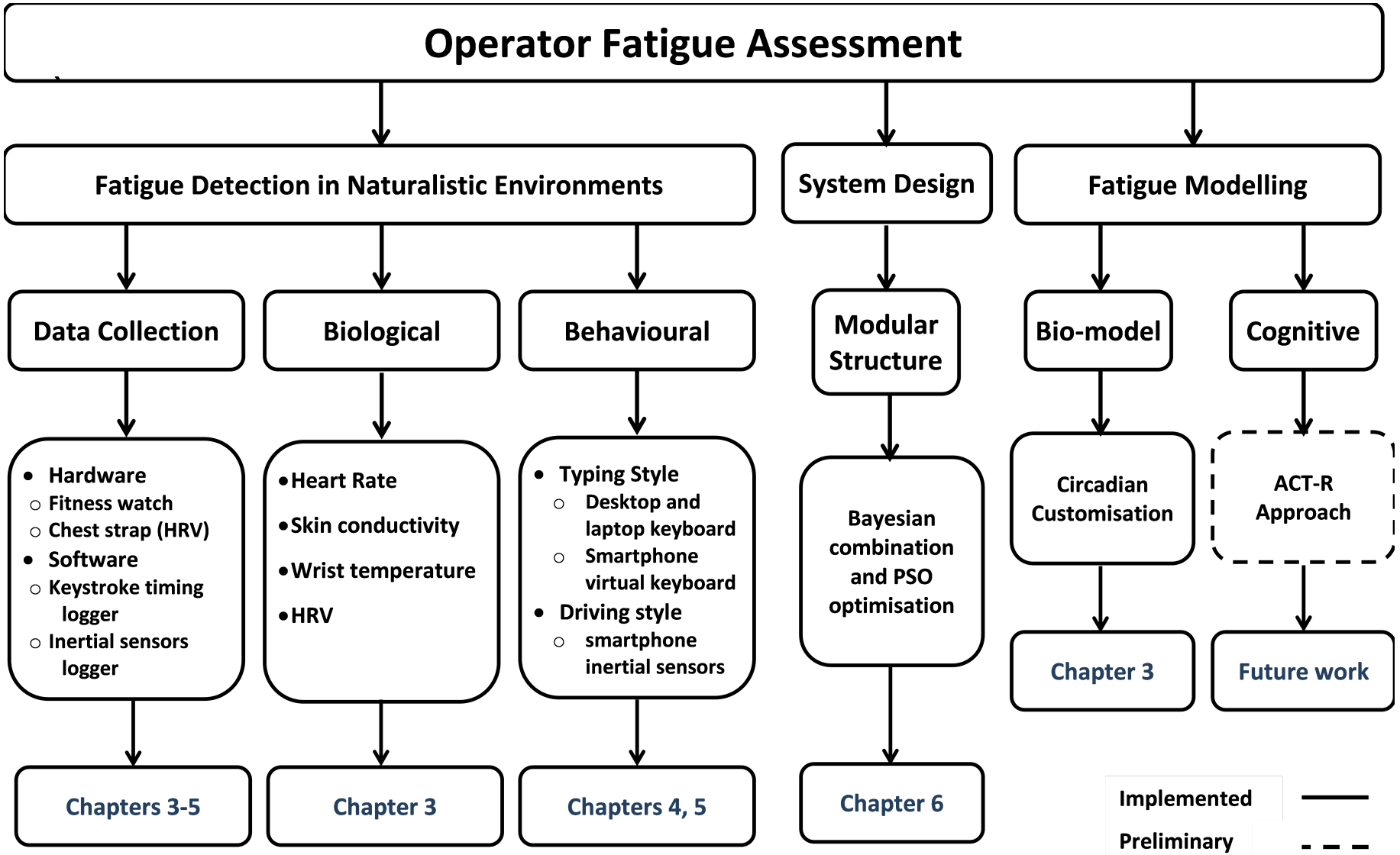


FIGURE 2.3: Proposed Work Overview

Chapter 3

Fatigue Detection and Circadian Modelling Customisation

3.1 Introduction

Fatigue assessment and quantification in naturalistic environments are considered essential requirements to reduce risks that occur as a consequence of a fatigued operator. The new developed wearable devices enable accurate measurement of multiple of fatigue-related physiological signals, thus making assessment of fatigue levels in real-life environment feasible and cost-effective.

Numerous fatigue/alertness models, especially models related to sleep deprivation, have been derived from a two-process model [103]. Most of these models focus on general and consistent features that are reflected by certain bio-data like the circadian rhythm. The circadian part of t the two-process model is represented by a pure sinusoidal wave or multiple in-phase sinusoidal harmonics. More

recently, a more detailed model has also been suggested [99]. The latter model investigate some additional patterns such as the post-lunch dip (sleep needs after midday hours), morningness and eveningness. A more flexible circadian model (two-harmonics) is then developed in [104] which further investigated the effect of these patterns on the circadian rhythm.

This chapter deals with two main objectives: (i) fatigue detection and (ii) circadian modelling customisation using physiological wearable devices. In the fatigue detection, a new approach based on physiological data is collection presented. Various types of physiological sensors are considered including heart rate, skin temperature, skin conductance and heart rate variability (HRV). In this work, the data is collected in a naturalistic environment, thus no particular tasks are required from the participantsto perform.

In the circadian modelling customisation, the Shape-Invariant Model (SIM) is used to represent the circadian rhythmwhich is then customised for more accurate representation of the individual users .

3.2 Wearable Physiological Sensors

Recent technology developments have led to the design and development of low-cost wearable devices that are capable of accurately measuring and collecting various physiological data including the heart rate and ECG, which are known to be influenced by mental fatigue [50], [58] as explained earlier in Section 2.4. Four physiological-data sensors for heart rate (HR), wrist temperature (WT), skin

conductance (SC) and RR heart beat intervals are adopted in this work for data collection and generation of fatigue-related metrics.

Two wearable devices are used in this study; a fitness tracker watch and heart-rate sensor strap. The fitness-tracker watch shown in Figure 3.1(a), is adopted in this study because of its: (i) capability of saving physiological-data in its internal memory, (ii) good performance in measuring heart rate (correlation coefficient of 0.83 in comparison with standard electrocardiographic device [105]), and (iii) low-cost (around £140 as listed in Appendix B) when compared to the equivalent devices. This fitness-tracker watch is used to collect several bio-data, including heart rate, wrist temperature, and skin conductance. This watch collects data at a rate of 1 reading per minute for each sensor.

(a) Fitness tracker watch (Basis peak) [106].

(b) Heart-rate sensor strap (Polar H7) [107].

FIGURE 3.1: Fitness wearable devices

The heart-rate sensor strap, shown in Figure 3.1(b), is used to measure the heart-rate beat intervals (RR). This device which is equipped with a Bluetooth communication facility, can therefore measure and transmit data wirelessly

transmitted in real-time to a wide variety of handheld devices (e.g. smartphones and tablets). A publicly available smartphone application is used to collect data from sensor and upload it to a remote web-server called Fluxtream [108]. Because of its high accuracy in measuring heart-rate (correlation coefficient of 0.99 in comparison with standard electrocardiographic device [105]), this strap is selected and used in this study.

3.3 The Two-Process Model

Alertness is usually measured by performance metrics like reaction time, motor tasks and perception times or by bio-data like core body temperature [109]. Some research papers showed that Heart Rate (HR) hourly changes are affected by many external factors like physical activity, job-demand and environmental cues. Despite of these changes, the circadian rhythm is still displayed in the trend of the collected bio-data along with other bio-data metrics like oral temperature [110]. Some studies have attempted to correlate the skin conductance with driver fatigue level [111]. Others studies have connected the skin resistance level with relaxation and pre-somnolence state. It was reported that skin conductance level is negatively correlated with circadian variation in core body temperature [112].

Based on the two-process model [92], [95] of the sleep regulation, bio-mathematical models have been built [113] for alertness, fatigue and performance assessments. The two-process model consists of two processes: (i) a sleep dependant S-process and (ii) a Circadian C-process. These processes are described as follows.

3.3.1 Sleep-process (Homeostatic)

The S-process (sometimes called homeostatic) depends on sleep evolution process. It follows exponential trend that grows in wake time with time constant $\tau_w = 18.2$ hrs and fall down in the same trend in sleep time with time constant $\tau_s = 4.2$ hrs [97]. During wake, S-process is governed by:

$$S_w(t) = 1 + (1 - S_{wo})e^{\frac{t_{wo}-t}{\tau_w}} \tag{3.1}$$

where time t is in hours, S_{wo} is the S value at the previous wake onset. and t_{wo} is the time of the previous wake onset. While during sleep, S-process is regulated by:

$$S_s(t) = S_{so}e^{\frac{t_{so}-t}{\tau_s}} \tag{3.2}$$

where S_{so} is the S value at the previous sleep time. and t_{so} is the time of the previous sleep.

3.3.2 Circadian process

The C-process along with with high (H) and low (L) limits, which is almost sinusoidal shape, is a free running process of a period of 24 hours [114]. The process ($C(t)$) in its original form is given by:

$$C(t) = sin(\frac{\pi}{12}(t - \phi)) \tag{3.3}$$

where ϕ is time-shift (in hours) from midnight.

Transition between wake and sleep occurs when the sleep pressure $S(t)$ reaches its upper threshold (S^+),

$$S^+(t) = S_{so} + C(t) \tag{3.4}$$

In contrast, transition between sleep and wake occurs when the sleep pressure $S(t)$ reaches its lower threshold (S^-),

$$S^-(t) = S_{wo} + C(t) \tag{3.5}$$

Based on some typical parameter settings, given in [97], the two-process model is generated as shown in Figure 3.2.

The circadian rhythm is a physical and mental process that displays an endogenous oscillation of about 24 hours as well as 12 hours rhythms [104]. Based on these types of rhythms, some researchers used physiological and behavioural data to model the rhythm interaction as well as the alertness daily pattern based on a shape invariant model (SIM) [115], [116]. A recent version of SIM has been developed to nicely fit the circadian driven data [115], [117]. This version is based on multiple harmonics rather than a pure sinusoidal shape.

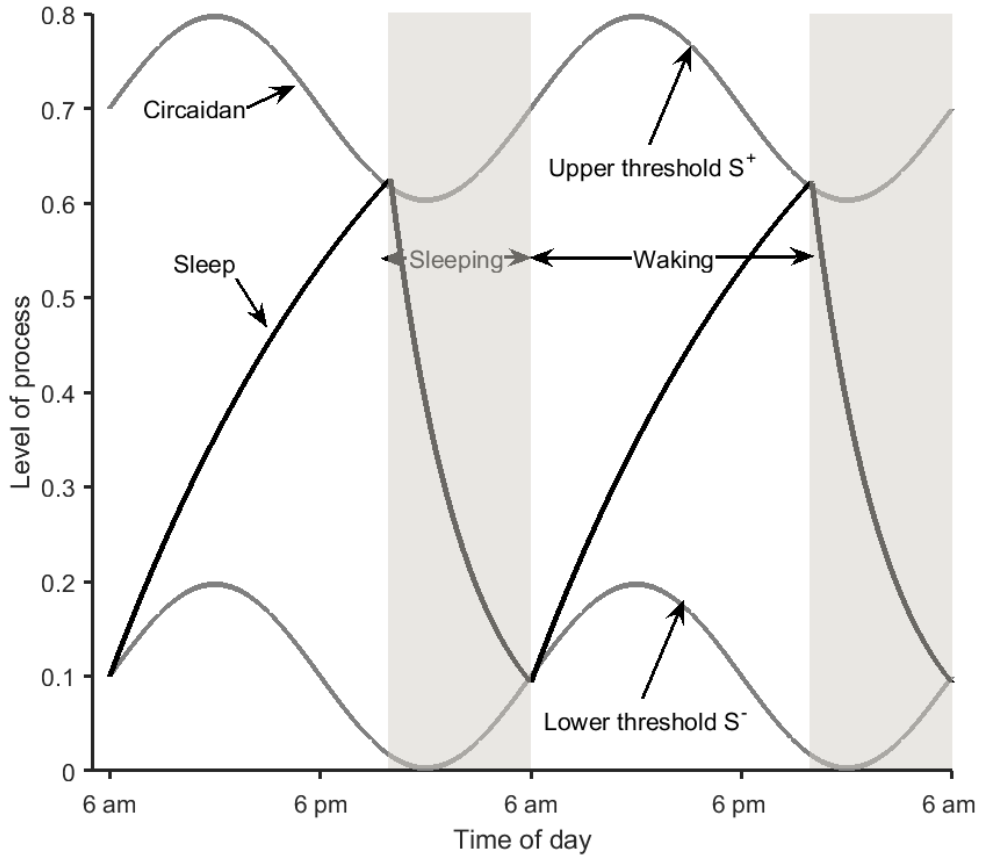


FIGURE 3.2: Two-process model

3.4 Shape-Invariant Model

Shape-Invariant Model (SIM) is a statistical approach developed to capture common patterns between individuals from a specific observations of natural phenomena over time [118]. It assumes that all participants have the same shape function after applying appropriate shift and scale transformations. The shift and scale parameters provides individuality differences among participants [119]. In mathematical form, SIM (y_{ij}) is giving by

$$y_{ij} = \mu_i + B_i S_H(t_{ij} - \varphi_i) + e_{ij}, \;\; i = 1, ..., m; \; j = 1, ..., n_i, \tag{3.6}$$

where y_{ij} is the response of i^{th} participant in the j^{th} time, μ_i is the daily average of each participant, m is the number of participants, n is the number of measures for participant i, B_i is the amplitude of the i^{th} participant, φ_i is the phase of the i^{th} participant and $S_H(t_{ij}-\varphi_i)$ is the shape function which represents the common pattern across participants. This function is expressed as harmonic function and is mathematically given by [117]:

$$S_H(t_{ij} - \varphi_i) = \sum_{k=1}^{N} A_k \, cos(2\pi k(t_{ij} - \varphi_i)) \tag{3.7}$$

where A_i is the amplitude of each harmonic, N is the number of harmonics.

In this work two harmonics are considered for several reasons, including [99], [117]: (i) they are widely used in modelling physiological and behavioural data, (ii) they can be used to interpret some phenomena such as morningness/eveningness and post-lunch dip. In addition, the higher-order harmonics do not consistently exist in the dataset used in this study.

The two harmonic shape function ($S_H(t - \varphi_i)$) which represents the Circadian rhythm of each participant, is given by:

$$S_H(t - \varphi_i) = A_{1i} cos(2\pi(t - \varphi_i)) + A_{2i} \, cos(4\pi(t - \varphi_i)) \tag{3.8}$$

Unlike previously reported work [118], parameters of the individual participants (amplitude, phase and frequency) are obtained through performing frequency analysis rather than using statistical iterative methods such as non-linear

regression. The frequency analysis method can be considered more efficient in terms of computational power especially with the increase of the parameters.

3.5 The Proposed System and Methodology

The proposed operator fatigue detection system is based on physiological data. This systems includes three distinct stages: (i) data collection and pre-processing, (ii) feature extraction and labelling including HRV measurements and (iii) fatigue detection and circadian customisation. It should be mentioned here that the last two stages of the fatigue detection system, shown in Figure 3.3, are implemented offline post completion of the practical data collection experiment.

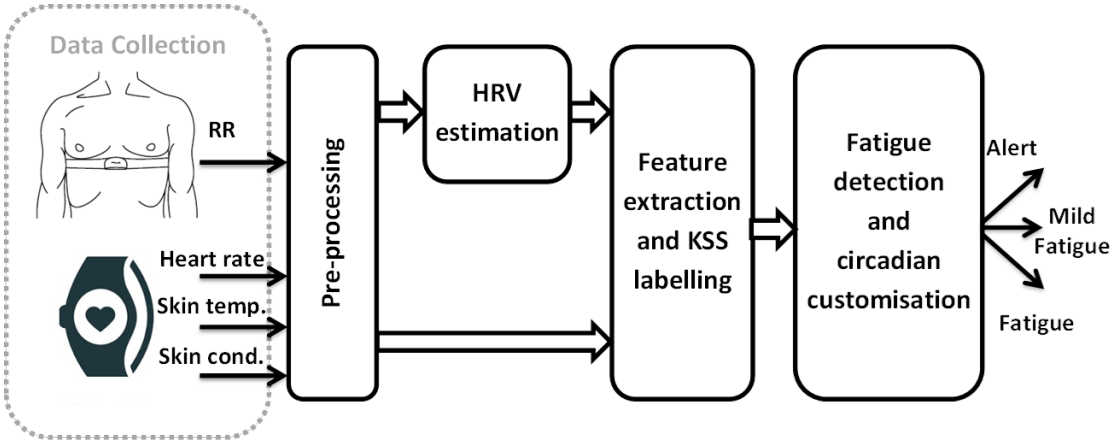


FIGURE 3.3: Block diagram for the proposed wearable-based system

Development methodology of the model comprises several stages, including: data collection, pre-processing, HRV estimation, feature extraction, and fatigue detection and circadian customisation. These stages are shown in the block diagram of Figure 3.3 and are described as follows.

3.6 Data Collection and Pre-Processing

Using low-cost and portable bio-sensors allow the model to be tested in real working conditions with minimal impact on the operators.

3.6.1 Data collection

In this work, two experiments were carried out, one is for fatigue detection and the other is for circadian customisation.

a) Experiment 1

A total of 9 participants, aged 16-50 years with body mass indexes of 21-35 participated in this study. Each participant was provided with two wearable devices: a fitness tracker watch and a heart-rate sensor strap. The data was collected around 16 hours (awake period) daily. The participants were asked to record their KSS sleepiness level every hour while they were practising their daily routine in university such as attending lectures, working on computer and other educational tasks in addition activities during day hours such as walking, driving or watching TV. Numerous experiments have been carried out over a period of 9 weeks; (each experiment takes about a week to complete). The participants were instructed to collect and synchronize data with a remote server. A user-friendly data collection and management application was used on the handheld device for this purpose.

The watch collects the data from the participant and sends to his smartphone by installed application. The smartphone uploads this data to the cloud whenever internet connection is available. This procedure is almost the same

for heart-rate sensor strap except the strap itself dose not record data so it need to be close to the phone within 3 meters to ensure data constant collection.

b) Experiment 2

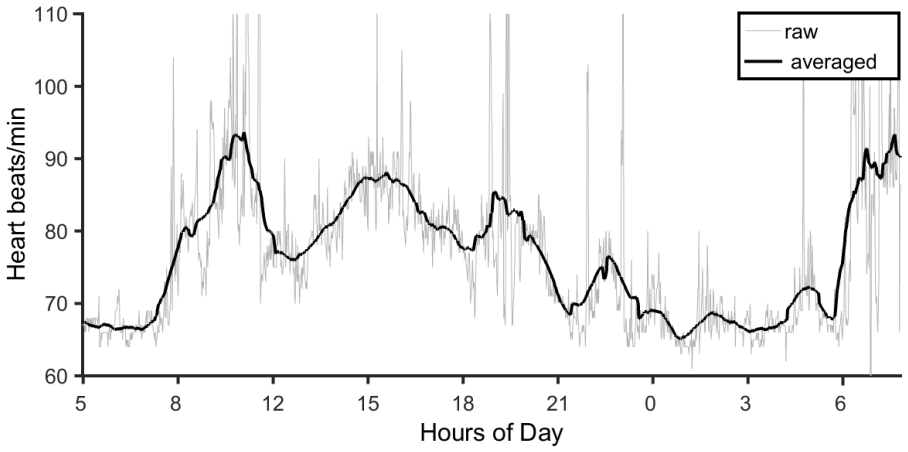
Six participants, aged 18-50 years old participated in this study for a period of 14 days using smart watches around the day. Each participant is asked to synchronize his data daily and upload it to the cloud using his mobile phone with installed application of the watch. The data is a stream of samples with a resolution of a sample per minute for each type of data (i.e. heart-rate, skin conductance and wrist temperature).

3.6.2 Per-processing

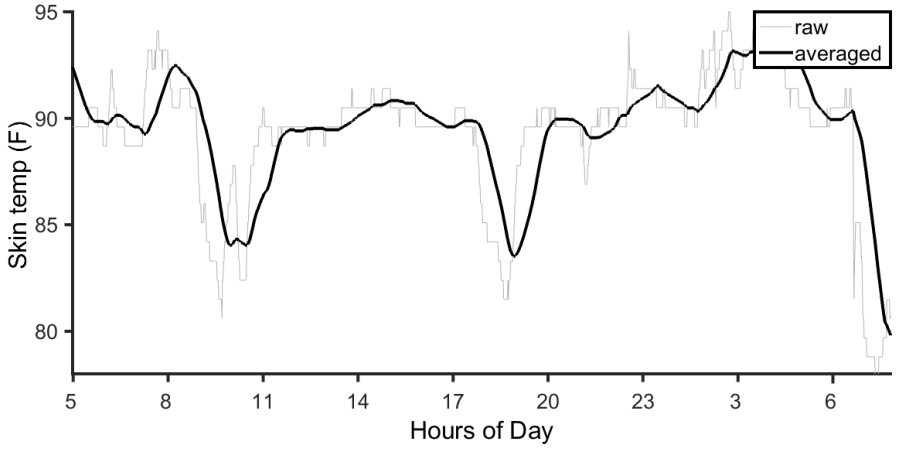
Missing data is a common problem in data collection; it can be caused by a sensor or a human failure. For example, a participant inability of wearing the watch or the chest strap during battery charging or shower time. To mitigate the impact of this problem, some practical arrangements were considered: (i) dealing with slots of less than 30 samples through interpolating the missing data and (ii) using unequally spaced frequency domain analysis when data-missing slot was greater than 30 samples. The 30 sample window (30 min) is chosen based on typical selection of self-reported fatigue scores which vary between 30-120 min [120].

The collected data from tracking watch was pre-processed and analysed to generate statistical features like 30 min (30 samples) moving average mean (shown in Figure 3.4) and standard deviation also a frequency domain analysis was conducted to generate several features such as maximum frequency and area under power

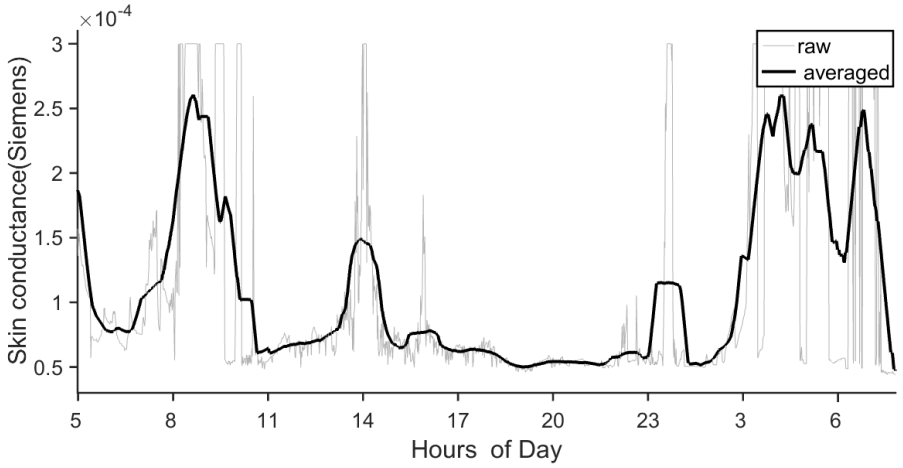
spectrum density curve. Finally, a set of several features were selected. The selected set of features are discussed in Section 3.8.



(a) Heart-rate data



(b) Wrist-temperature data



(c) Skin-conductance data

FIGURE 3.4: Examples of the collected physiological signals

3.7 HRV Estimation

An example of 10-minute RR intervals for one of the participant involved in this study is shown in Fig. 3.5. The collected RR data from chest strap is analysed and used to estimate HRV. An average record duration of 10 mins is considered adequate to obtain reasonably accurate and reliable HRV indexes [121]. A moving average window is used to calculate HRV indexes with an overlap of 9 minutes to match the resolution of data collected from watch sensors (one measurement per minute).

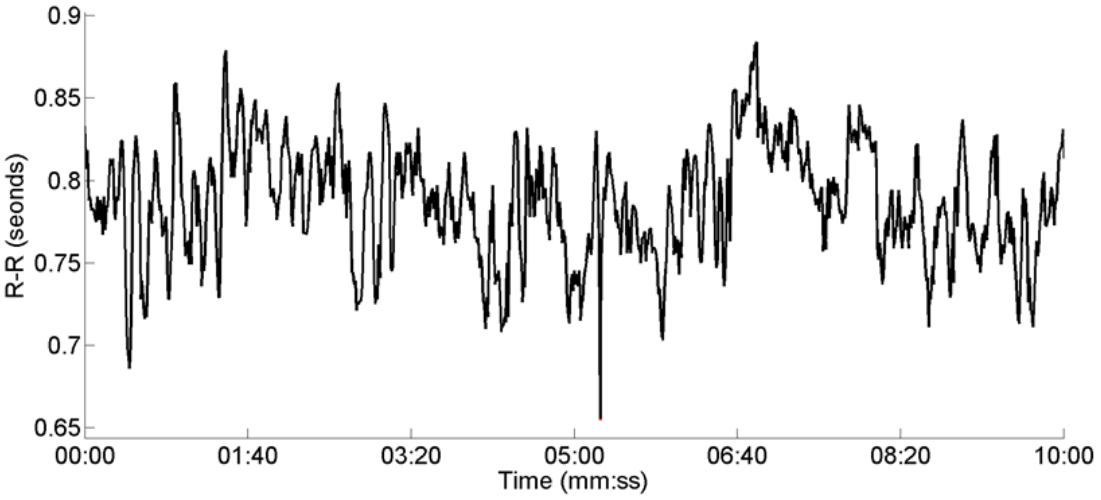


FIGURE 3.5: RR intervals extracted from heart-rate monitor

Based on estimation method, HRV indexes can be categorised mainly into two domains time and frequency domains as listed previously in Table 2.4. The ratio of low frequency power to high frequency power (LF/HF) of RR power spectral density (PSD) is widely used in research as fatigue and stress indicator [122, 123, 124, 121]. While less frequent research papers use the whole set of HRV indexes for fatigue detection [125]. The time domain HRV indexes are simpler to calculate than frequency domain set and the ratio of standard deviation of NN (RR) intervals (SDNN) index over root mean square differences of successive RR (RMSSD) index

can be considered as a surrogate for LF/HF ratio [126]. Hence, in this work, two HRV indexes are used: LF/HF ratio and SDNN/RMSSD ratio. The time domain indexes (SDNN and RMSSD) are calculated as follow [126]:

$$\overline{RR_n} = \frac{1}{N-1}\sum_{n=1}^{N} RR_n \tag{3.9}$$

$$SDNN = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left[RR_n - \overline{RR_n}\right]^2} \tag{3.10}$$

$$RMSSD = \sqrt{\frac{1}{N-2}\sum_{n=2}^{N}\left[RR_n - RR_{n-1}\right]^2} \tag{3.11}$$

where RR_n is the time series of RR values within duration of 10 minutes and N is the number of samples in RR_n.

The frequency domain HRV indexes is calculated by transform RR time series data into frequency domain. Power Spectral Density (PSD) of time series is usually calculated using many methods and analysis approaches such as Welch, Burg and Lomb-Scargle [127, 128]. In this work, Welch approach is used to calculate PSD of RR signal because it widely used in HRV indexes calculation [129, 130]. Welch PSD approach is described mathematically as follows:

$$PSD_{Welch}(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} RR_n w_n e^{-i2\pi f_n}\right|^2 \tag{3.12}$$

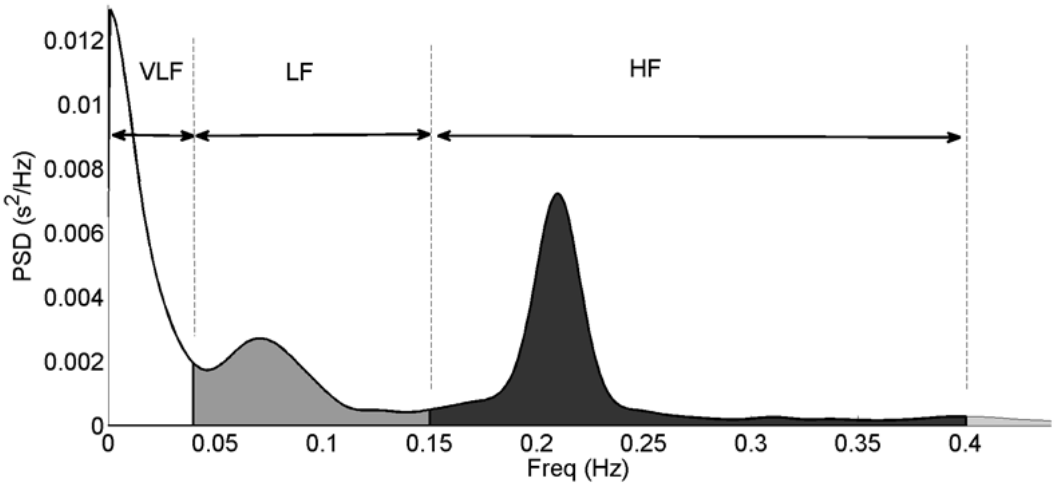
where $U = \frac{1}{MU}\sum_{n=0}^{M-1} w_n^2$, w_n is weighted Hamming window and M is the window size.

Figure 3.6 shows two examples of PSD in which frequency band limits are identified. LF band (0.04-0.15Hz) is represented by grey area while HF band (0.15-0.4Hz) is represented by black area. Changes in bands power are clearly shown between two example due to change in operator's mental status. It is worth mentioning that the range of PSD also changes between two examples which reflects changes in total power of deviation in RR data.

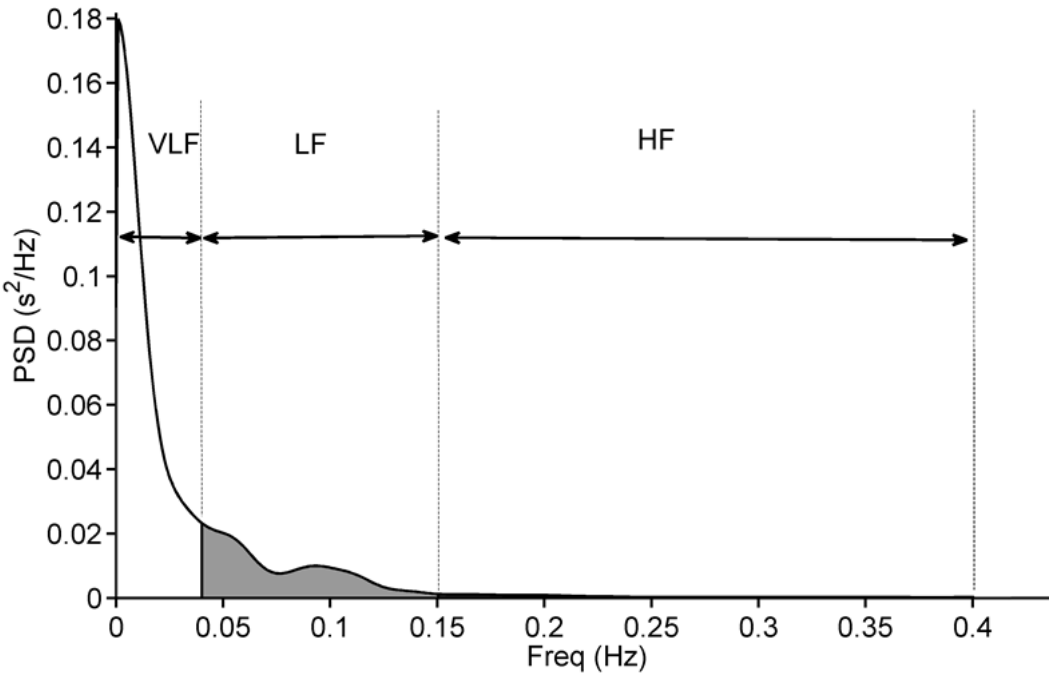
An example of calculated LF/HF ratio around waking hours is illustrated in Figure 3.7. In this example, LF/HF fluctuates widely over hours of day and does not show consistent trend and this is may come from physical fatigue which changes frequently around the day due to different activities. However, higher values of this index is reported in early hours of the day while lowest values appear at early hours of night.

3.8 Feature Extraction and Data Labelling

The collected data from the fitness watch (hear-rate, wrist temperature and skin conductance) are passed to feature extraction stage after pre-processing stage. 14 features are extracted from the three physiological metrics including time domain and frequency domain features. Another set of data has been offered by heart-rate sensor strap to be used HRV indexes (LF/HF and SDNN/RMSSD) calculation. Also, time of day is used as an extra feature because it is involved directly with circadian rhythm of operator which is the main contributor of operator alertness state.



(a) PSD example at 23:15 (fatigued)



(b) PSD example at 14:15 (alert)

FIGURE 3.6: Example record of power spectral density for RR data

Practical implementation needs of real-time fatigue detection system and classification accuracy are used as a guide for feature selection. Hence, a trade off between computational power and classification accuracy is followed to choose

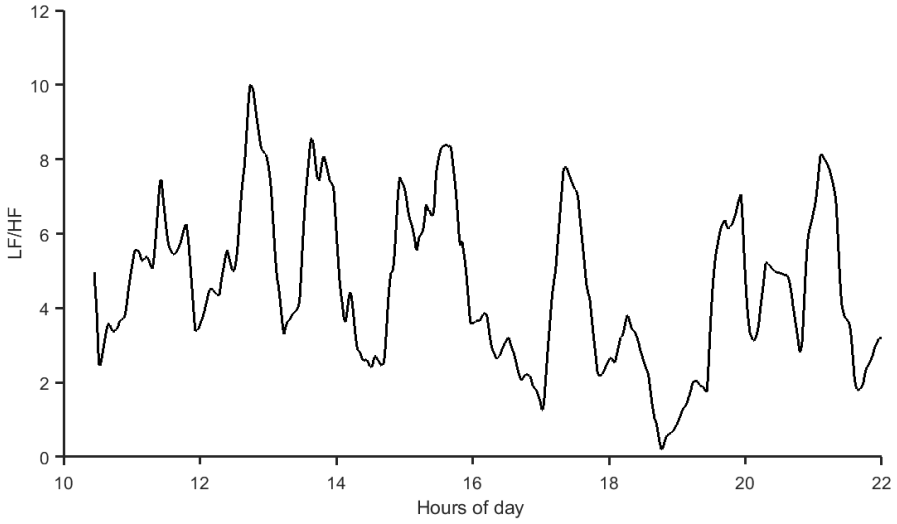


FIGURE 3.7: Example of HRV index, LF/HF PSD

best set of features. The following 9 out of 14 features are chosen after feature extraction including time of the day:

a) Heart rate 30 sample window average

b) Heart rate standard deviation

c) Wrist temperature 30 sample window average

d) Wrist temperature standard deviation

e) Skin conductance 30-sample window average

f) Skin conductance standard deviation

g) LF/HF ratio

h) SDNN/RMSSD ratio

i) Time of the day

The self-reported KSS scores are used to label the selected features. Although using two labels has shown better performance, three labels (classes) are adopted in this experiment to capture fatigue onset and development. KSS , which consists of nine level, are divided equally into three classes: Alert, Mild fatigue and Fatigue.

3.9 Fatigue Detection

This stage proceeds in two steps: classification and Bayesian combiner. The labelled features are fed to detection stage which in turn generates decision on the operator fatigue status (i.e. Alert, Mild fatigue or Fatigued). Figure 3.8 shows a block diagram for the proposed two-steps detector, starting from four ternary sub-classifiers and ending with Bayesian combiner stage.

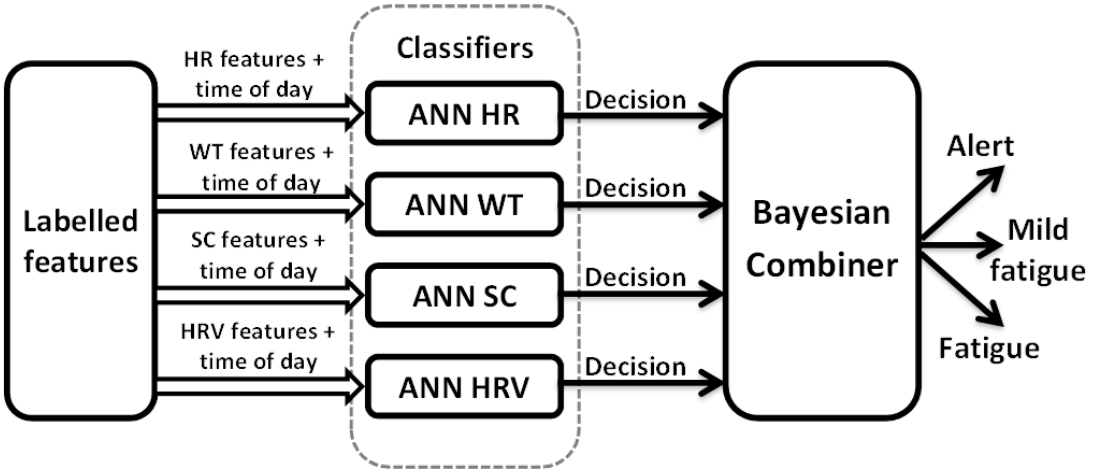


FIGURE 3.8: Two-stage fatigue detector (classifier and combiner)

3.9.1 Classification

The classification is the process of mapping a new input to one of the output category. Large number of statistical and heuristic classification methods have

been used in wide range of research areas [131]. The classifiers have different groups of categorisation based on learning method (supervised and unsupervised), number of classes (e.g. binary, ternary or multi - classes) or algorithm type (e.g. linear, kernel estimation or quadratic). The classifier can be functionally described in several steps as follows. Starting with data collection from object using data acquisition tools. The collected raw data is preprocessed and de-noised if needed. then, the features is extracted and the best combination of features are selected. Next, the chosen features with their labels is divided in to training and test sets to train and evaluate the classifier. The classifier performance is assessed by different metrics such as accuracy, sensitivity, precision, confusion matrix and area under curve (AUC) of receiver operation characteristics (ROC) [132].

Numerous classifier types have been used in operator fatigue detection studies such as k-nearest neighbour (kNN), decision tree (DT), random forests (RF), artificial neural network (ANN), support vector machine (SVM) and others [133], [90], [134]. However, the most commonly used classifiers in fatigue detection are ANN and SVM . These classifiers normally used for pattern recognition which is the case in fatigue detection as well as their ability to capture complex non-linear relationships between different fatigue-related physiological metrics [135]. In this work, both classifiers are used as required in various fatigue detection experiments.

The collected dataset is used to classify the operator status into Alert, Mild fatigue and Fatigue states. The dataset is divided into four subsets; each subset is selected based on data type (heart rate, wrist temperature, skin conductance, HRV) in addition to time of day which is included in each subset. Four ANN ternary classifiers are then trained by 65% and validated by 10% of individual feature subset while the rest of 25% are used for test. The structure of the ANNs is based

on the feed-forward with an input layer, two hidden layers and three output units with a tangent-sigmoid transfer function.

Different configurations were also considered to identify the ANN structure with the best performance. These configurations involved changing the number of hidden layers and associated nodes as well as optimizing the training algorithm and the decision transfer function. Levenberg-Marquardt back-propagation algorithm was eventually selected for the ANNs' training. Figure 3.9 shows an example of the HRV sub-classifiers performance illustrated as confusion matrix.

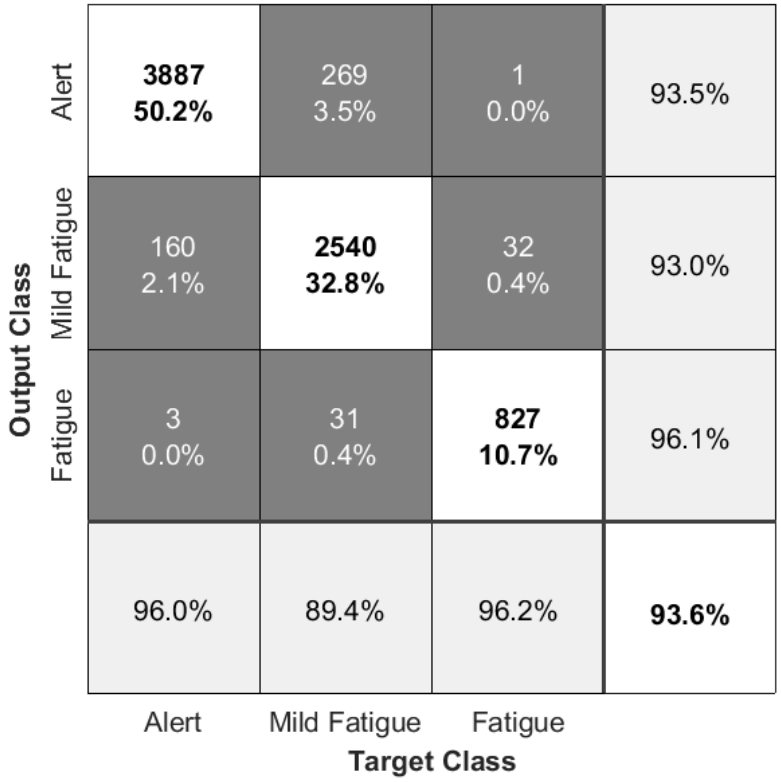


FIGURE 3.9: Example of sub classifiers performance (HRV classifier)

3.9.2 Bayesian combiner

Fusion and combination have been used interchangeably in classification problem. However, fusion term has been used commonly with raw and feature level while combination term has been used with decision or abstracted level [136]. Several definitions for data fusion have been reported in literature. However, the mostly agreed upon definition is [137],

"Information fusion is the synergistic integration of information from different sources about the behaviour of a particular system, to support decisions and actions relating to the system".

Generally, data/ information fusion approaches can be divided into three levels [137]: (i) a data-level that combines multisensor raw data, (ii) feature-level that merges features extracted from raw data, and (iii) a decision-level.

For decision level, several combination algorithms have been proposed to combine heterogeneous sets of modules (e.g. majority voting, weighted majority voting, or Bayesian combiner) [138], [139]. The Bayesian combiner (BC) is ideally suited for problems when the output of the modules is independent even when number of modules are dropped to two [140], [141], [142]. Hence, this approach is adopted in this work to improve accuracy and confidence of the data classification stage.

The BC works on abstract level of the output of L modules, each module M_i predicts class label $b_i \in \boldsymbol{H}, i = 1, ..., L$. So, any input set $\boldsymbol{x} \in \boldsymbol{R^n}$, can be combined, the L module outputs produce a vector $\boldsymbol{b} = [b_1, ..., b_L]^T \in \boldsymbol{H^L}$. The core of the Bayesian combiner relies mainly on Bayes theorem [143]. This theorem of probability computes the posterior probability of module M_i based on

the prior probability $P(h_i)$ and the likelihood $P(\boldsymbol{b} \mid h_j)$ of the evidence $\boldsymbol{b}$. The independence assumption is maintained and allows the conditional probability of the module M_i labels the input $\boldsymbol{x}$ in class $b_i \in \boldsymbol{H}$ to be presented as follows:

$$P(\boldsymbol{b} \mid h_j) = P(b_1, b_2, ..., b_L \mid h_j) = \prod_{i=1}^{L} P(b_i \mid h_j) \tag{3.13}$$

Bayes rule can be described mathematically as follows:

$$P(h_j \mid \boldsymbol{b}) = \frac{P(\boldsymbol{b} \mid h_j)\, P(h_j)}{P(\boldsymbol{b})} \tag{3.14}$$

At this stage, the final decision is generated based on the decisions received from the ANNs. As shown in Figure 3.8, the output of each sub-classifier are combined together by applying Bayesian algorithm.Bayesian combiner follows maximum a posteriori probability (MAP) rule [144, 145] in which the classifiers performance metrics are approximated as posterior probability. The detection results post combiner stage are shown in Table 3.1 . As illustrated, the overall classification accuracy and specificity are improved when compared to the results obtained from the sub-classifiers while sensitivity metric is slightly dropped in comparison with HRV sensitivity but still much higher than other sub-classifier sensitivities.

3.10 Circadian Customisation

3.10.1 Data visualisation

This section presents samples of the collected bio-signals: heart rate, body temperature, and skin conductivity, as illustrated in Figure 3.10. As the data patterns of all participants are found to be almost similar, the results presented in this section will focus on the participant with the largest set of data. Several signal-processing techniques are then applied to the dataset of the participant under study, and the obtained results are also presented and discussed in this section. Unlike Chapter 3, where the purpose of the main objective was fatigue detection, this section focuses on circadian customisation. The physiological metrics in this section are therefore re-represented with a particular orientation to the circadian customisation as follow.

1) Heart rate: The heart rate pattern appears more clearly when a window average of 60 samples is applied for the data. Selection of this reduces activity fluctuations of the heart rate. Figure 3.10(a) shows an example the daily rhythm of the heart rate changing around the day. The after lunch dip for heart rate level is consistently appeared around 12-16 pm.

2) Skin temperature: It is another marker for the circadian rhythm, which is almost as a mirror of the core body temperature. Example of the collected wrist temperature signal is shown in Figure 3.10(b). It is noted that the night temperature is always higher than day temperature, as illustrated.

3) Skin conductance: It is measured in Siemens unit and the data is collected around four days. Example of the collected signal for skin conductivity is

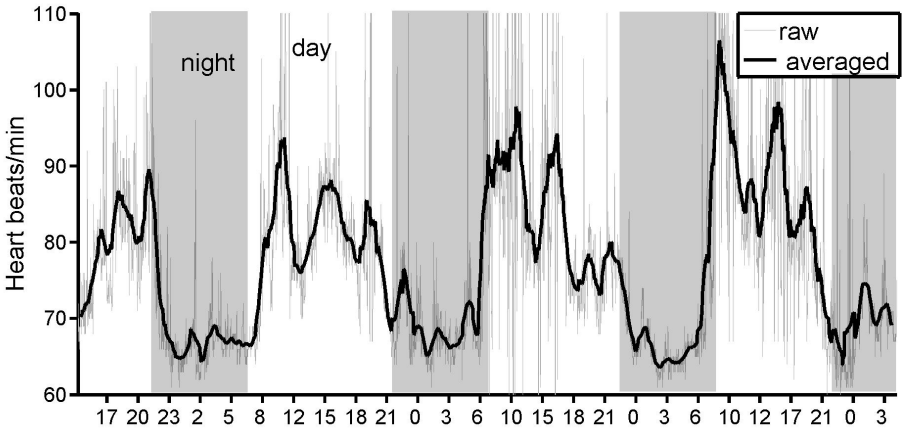
shown in Figure 3.10(c). Some values are clipped and limited to a certain value since these values is impractical and does not reflect the expected data for skin conductance.

The circadian rhythm is shown in Figure 3.10(a) where heart rate increases during the day period (i.e. the operator is active)and drops at night. Biological rhythms of human span a range of cycle lengths, including the daily circadian rhythm and the 12-hour of afternoon dip rhythm. As the cyclic details which can not be shown clearly in time domain, the frequency domain analysis is used here to capture these rhythmicity details.

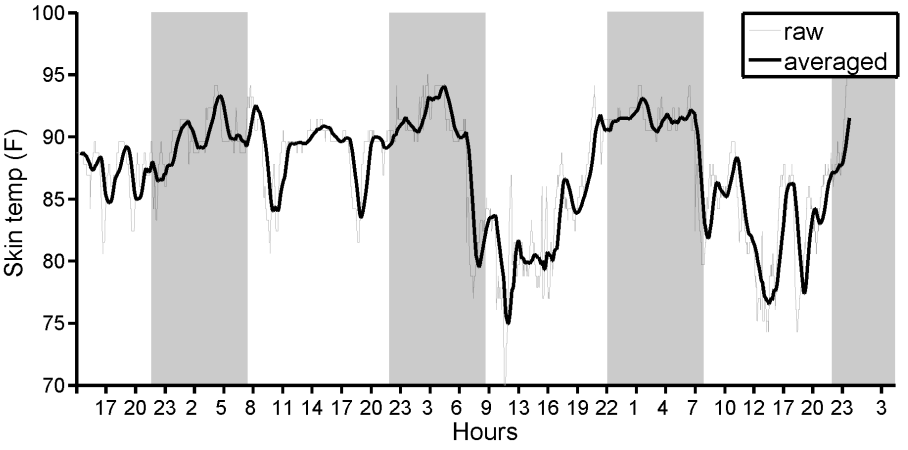
Figure 3.11 shows the frequency domain results for the heart rate signal using wavelet transform analysis which it is applied to the heart rate signal for one of the participants. The expected note in this figure is the rising of heart rate in day period while it decreases at sleeping time. The best results of showing the daily rhythm and post lunch dip were displayed by using wavelet approximation signal based on DMeyer mother wavelet with level 7 and it reflects the daily rhythm with afternoon dip (siesta).

3.10.2 Circadian customisation

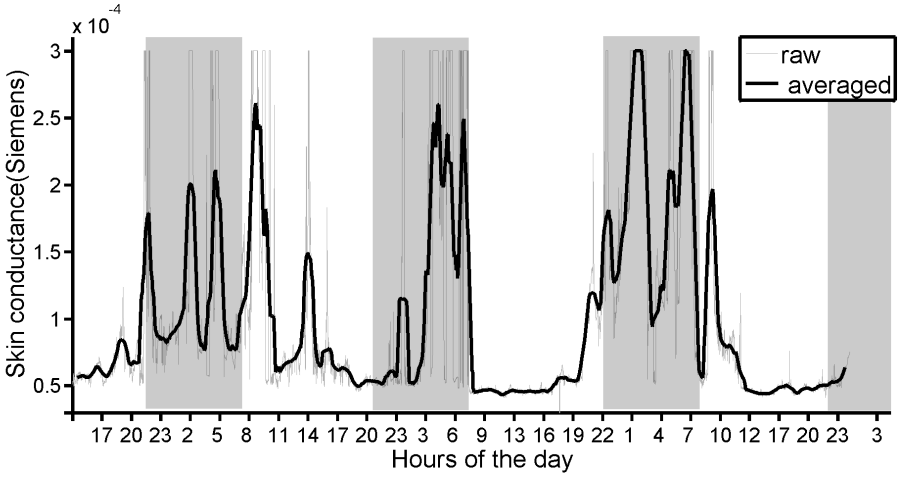
The collected heart rate data is averaged around one 14 days slot to enhance the common daily rhythm of each participant. The circadian rhythm and the midday dip are clearly shown in Figure 3.12. As illustrated, heart rate daily pattern differs among participants and the individual circadian rhythms are clearly appearing. Also, the midday dip varies in terms of its time and period among the participants. The important findings formed a core contribution for the proposed model.



(a) Heart-rate data



(b) Wrist-temperature data



(c) Skin-conductance data

FIGURE 3.10: Examples of the collected bio-signals, with 30-sample window averaged for 4-day period

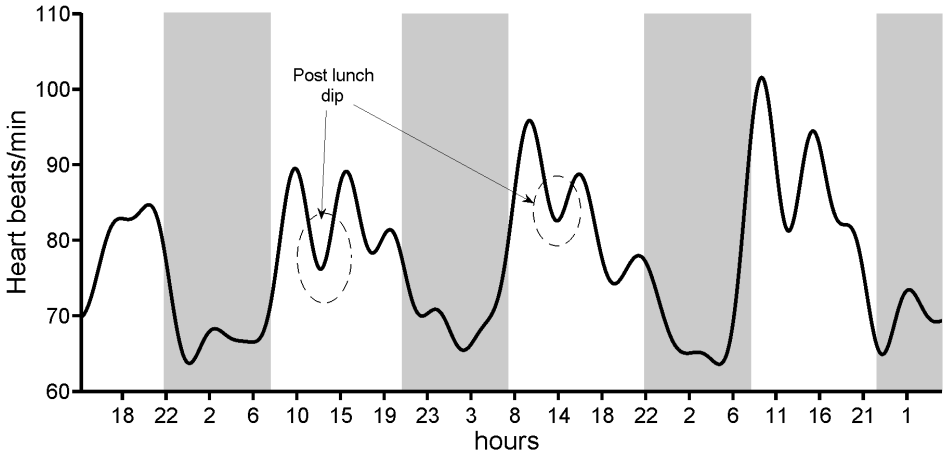


FIGURE 3.11: Wavelet approximation (DMeyer level 7)

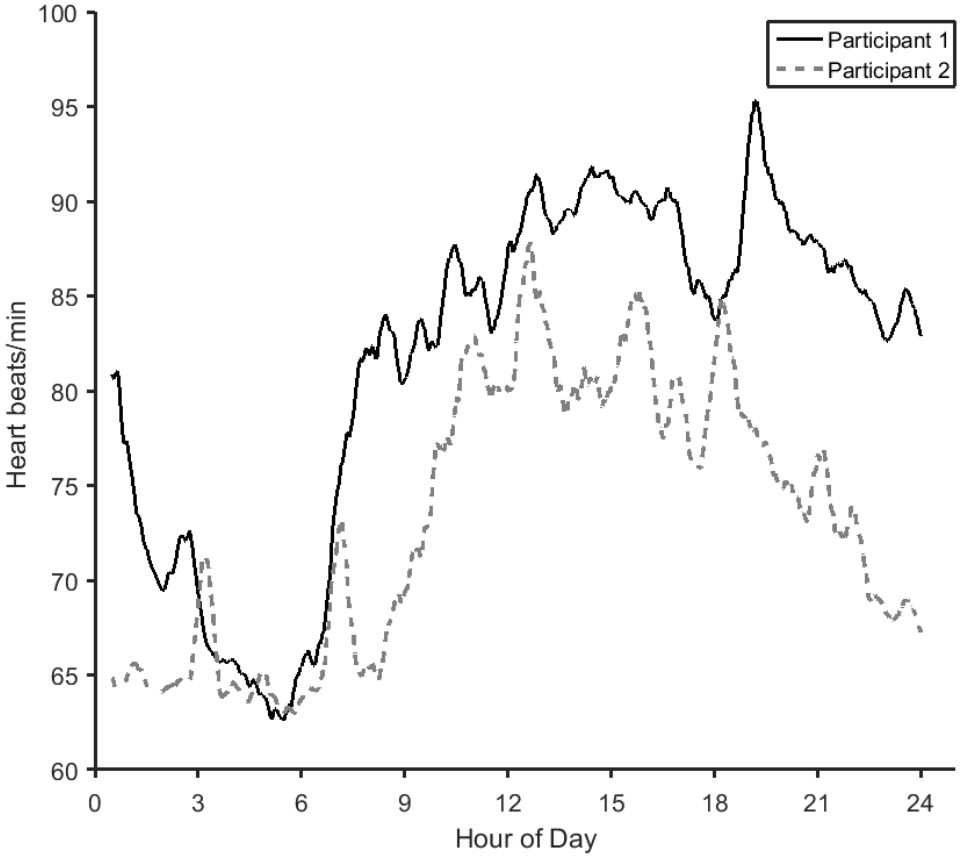


FIGURE 3.12: 24-hour heart rate variation averaged over 14 days

The two-harmonic function, described in Equation 3.8, is adapted in this work by adding the fact that the two harmonics are not 24 and 12 hour exactly but they are vary around these number. Hence, two different frequencies and phase shift

are added to the shape function. The obtained shape function CR for the two sinusoidal functions with almost 24 and 12 hours harmonics is given by

$$CR = A_1 cos(\omega_1 t - \theta_1) + A_2 cos(\omega_2 t - \theta_2) \tag{3.15}$$

where the values of A_1 ,A_2 (the amplitude of the 24 , 12 components respectively) and ω_1, ω_2 (the frequency of the two components), θ_1 and θ_2 (the phase shift of the two components) are unknown parameters and to be calculated using the Fourier transform results (amplitude, frequency and phase).

Fourier transform of heart rate signals is illustrated in Figure 3.13 which gives a good view for the two harmonics in interest. The first harmonic is around 7 cycles/week and its related to the daily rhythm circadian clock. The second component that is with 14 cycles/week can be explained as the 12-hour component. Its worth mentioning that the first harmonic has significant amplitude relative to around harmonics for all participants while this is not the case for second harmonic. The shape function parameters are estimated for each participant from Fourier transform graphs (amplitude and angle) as overviewed in Figure 3.13.

3.11 Results and Discussion

Ten trials of randomly selected records from data sets were conducted to calculate the performance metrics of classification for all classifiers. These performance are calculated using multi-class performance macro-averaging method [132, 146] as follows.

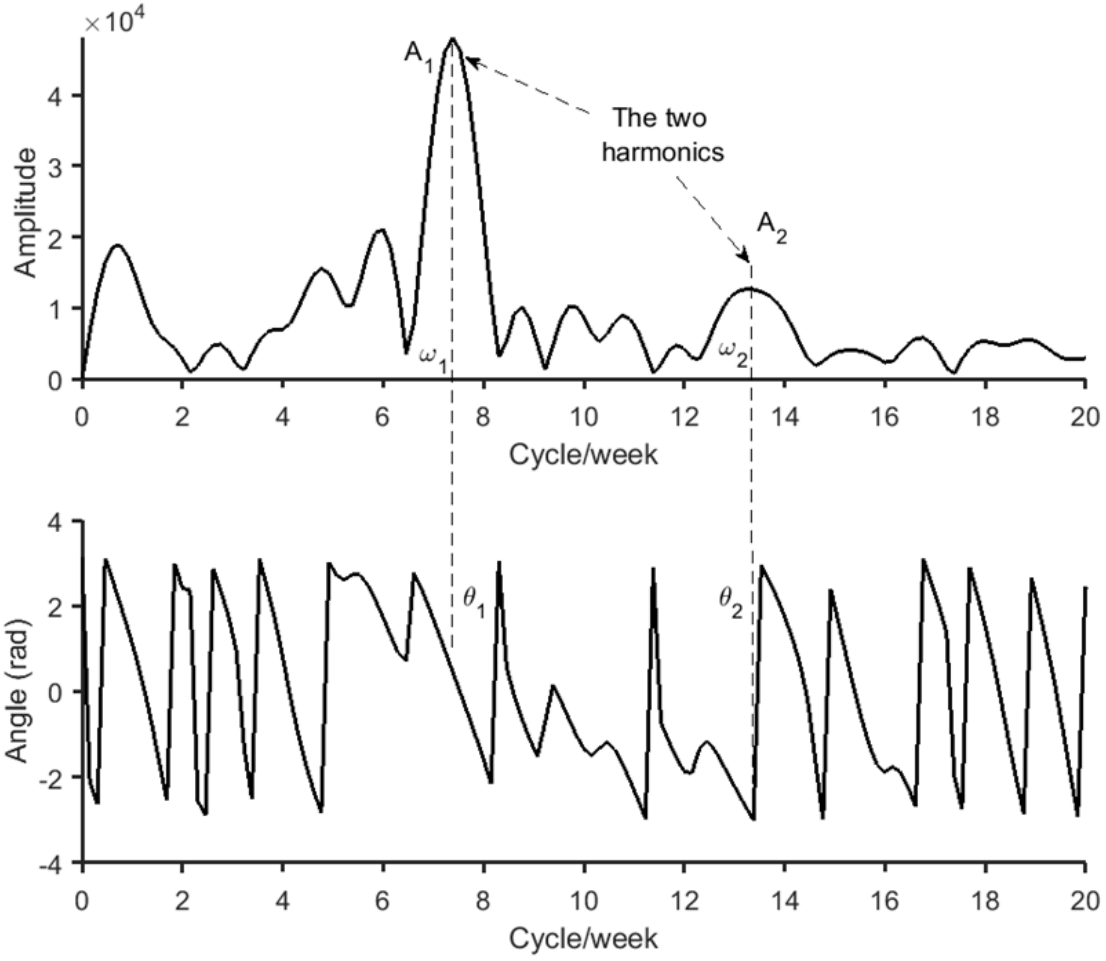


FIGURE 3.13: An example of Fourier analysis for heart rate signal

$$Average\ Accuracy = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{l} \tag{3.16}$$

$$Average\ Sensitivity = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \tag{3.17}$$

$$Average\ Specificity = \frac{\sum_{i=1}^{l} \frac{tn_i}{tn_i + fp_i}}{l} \tag{3.18}$$

were l is the number of classes which is 3 in our case, tp_i is true positive for i^th class, tn_i is true negative for i^th class, fp_i is false positive for i^th class and fn_i is false negative for i^th class.

Table 3.1 clearly shows the superiority of the BC results over the individual

sub-classifiers in terms of accuracy and specificity. The sub-classifiers demonstrated variation in results in terms of accuracy. The HRV sub-classifier demonstrated the highest accuracy (93.7%), while the HR sub-classifiers demonstrated the least accuracy (78.2%). These results are based on the normalised features that were extracted from the dataset of the nine participants.

Accuracy of the developed fatigue detection system is found to be close to that obtained by existing EEG-based fatigue detection systems that have been considered the most direct, valid and effective physiological measure for assessing the state of mental fatigue. Various EEG-based fatigue detection methods have been developed and their detection accuracies are in the range of(88%-99%) [147]. These methods, however, were tested in laboratory environment which is beyond scope of the work presented in this thesis. Furthermore, the EEG-signal is generally weak and noisy, thus it is not easy to detect accurately in a naturalistic environment [47].

Table 3.1: Summary of system performance

Performance metrics	ANN classifiers				BC (%)
	HR (%)	WT (%)	SC (%)	HRV (%)	
Accuracy	78.2	86.7	82.2	93.7	94.5
Sensitivity	72.1	81.0	73.8	94.0	93.1
Specificity	77.3	83.2	86.0	93.5	95.3

Figure 3.14 illustrates an example results obtained from the derived shape function given in Equation 3.15. As illustrated there are two graphs one in grey colour which represents the raw data of heart rate for one participant averaged long 14 days of data. The other graph represents the customised circadian obtained by the shape function where clear similarity are shown between the two graphs.

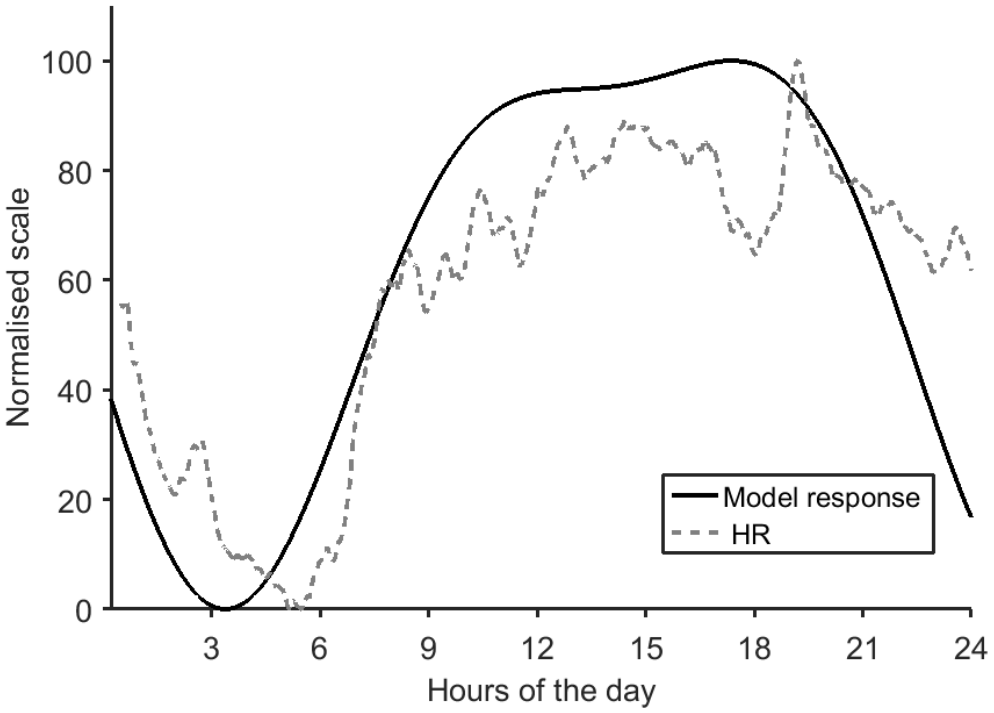


FIGURE 3.14: An example of circadian customisation based on two-harmonic shape function

Similarity between collected and model-based datasets are checked by performing correlation test for both datasets and the results are listed in Table 3.2. Most of participants circadian rhythms are found to have strong and very strong correlation except for participant 4 which is considered as moderate correlation strength (using the guide of correlation strength that reported in [148]) . Differences in correlation coefficient between collected and model-generated datasets can be explained due to variation of second harmonic significance among participant rhythms.

The heart-rate classifier (shown in Figure 3.8) is re-used in this experiment and tested with customised circadian. Table 3.3 compares the classifier performance with and without circadian customisation. It can be noticed that approximatley 4% accuracy improvement has been achieved.

Table 3.2: Correlation strength between collected data and customised CR

Participant number	Correlation Coefficient (r)	Correlation strength [148]
1	0.91	very strong
2	0.84	very strong
3	0.88	very strong
4	0.56	moderate
5	0.86	very strong
6	0.72	strong

Table 3.3: Comparison of system performance with and without circadian customisation (based on heart-rate data)

Performance metrics	Before customisation	After customisation
Accuracy %	79.2	83.2
Sensitivity %	74.3	78.5
Specificity %	78.3	81.1

3.12 Conclusions

A new multi-sensor fatigue detection system has been proposed and implemented successfully. The developed prototype was found to be promising in terms of usage low-cost wearable devices (around £170 as listed in Appendix B) to detect fatigue status of operators in real-life environments with high level of accuracy. Unlike previous studies, this system is tested in naturalistic environments to detect three levels of operator alertness (Alert, Mild fatigue and Fatigue) rather than in laboratory environment with just two levels (Alert, Fatigue) [149].

In this experiments, the results were obtained for 9 participants over a period of 48

hours. These results are close or outperform the state-of-the-art results reported in [122] in which 12 drivers were involved with two hours of driving. A fatigue detection accuracy of 94.5% has been achieved in the present work which is 4.5% higher than that reported in [122]. In addition, unlike the work reported in [122] the developed fatigue detection systems provides three detection levels (Alert, Mild fatigue and Fatigue) rather than two levels (Alert and Fatigue).

A modified Shape-Invariant Model (SIM) has also been developed and used to customise the circadian rhythm of individuals. The developed model is a modified version of the SIM that is based on the two-harmonic function. A frequency analysis technique is used to determine the circadian customisation parameters, thus a new model which strongly correlates to the data collected from experiments at a real-environments.

The circadian customisation method proposed in this study has improved the fatigue assessments through identifying the individuality differences and including them in the developed model. The obtained results have demonstrated that the fatigue detection accuracy, based on the heart-rate data, has increased approximately 4% when compared to that obtained before model customisation. This confirms the importance of taking the individuality deferences into consideration in fatigue assessment.

It should be noted here that the work presented in this chapter had faced some challenges/limitations, including: (i) lack of a close collaboration with domain knowledge experts and (ii) the need for wearing measuring devices for several days which limited the number of participants in this study.

Chapter 4

Fatigue Detection - Naturalistic Typing

4.1 Introduction

Fatigue can be reflected by behavioural changes of operators motor skills [1]. Of these, driving, walking, typing, computer interaction and riding a bike are the most common. These skills need to be repeatedly practised to be done automatically. Performance degrades on psychomotor-related tasks as fatigue develops, and this degradation can have serious health, safety and economic implications. Quantifying the degradation, which is addressed as an indicator of fatigue onset [68], can help to prevent fatigue by scheduling breaks, changing time shift or using any other fatigue countermeasure methods.

In this chapter, two systems with three different typing datasets are proposed to capture the performance degradation of operators. In the first system, two

keyboard typing-based datasets are used and a binary classifier is suggested for fatigue detection. The first dataset is a new dataset generated by the author while the second is an existing dataset that was generated to simulate some mental disease effect [75]. The second system is proposed and tested using an existing smartphone-based dataset.

The proposed systems rely on performance metrics that can be calculated from keyboard usage. These metrics are called keystroke dynamics and are explained in the next sections.

4.2 Keystroke-Dynamics and Error Rate Measures

Keystroke-dynamic metrics are time-related metrics that can be calculated from operator's fingers interactions with a computer keyboard. The main reason behind the interest in increasing research on keystroke dynamics is their use in security and authentication algorithms. This can be simply integrated with existing computer hardware for security systems [74]. Other important applications of the keystroke dynamics include calculation of behavioural metrics [75], [150] and more recently fatigue detection which is of a particular interest in this study [79], [151].

The methodology used in [79], [151] assumes that the fatigue level is only depended on time of the day. Moreover, the practical implementation of these studies challenges may need other device (i.e. mouse) and metrics (i.e. mouse speed, location and others). This requires more computation resources to process the suggested algorithm.

Hold Time (HT) and digraph (time between two consecutive key presses) represent the basic metrics of keystroke dynamics from which the features are extracted. Figure 4.1 depicts the time features related to keystroke dynamics. Latency is another term that is commonly used to describe timing of the keystroke. It includes press-to-press (digraph), release-to-release (RR) and release-to-press (RP) latencies [152]. The typing style is known to be individualistic [153] and external factors, such as keyboard type, keys layout and language affect the typing rhythm [154]. However, psychological factors, such as sleepiness, stress and emotions can also affect typing style [155], [156].

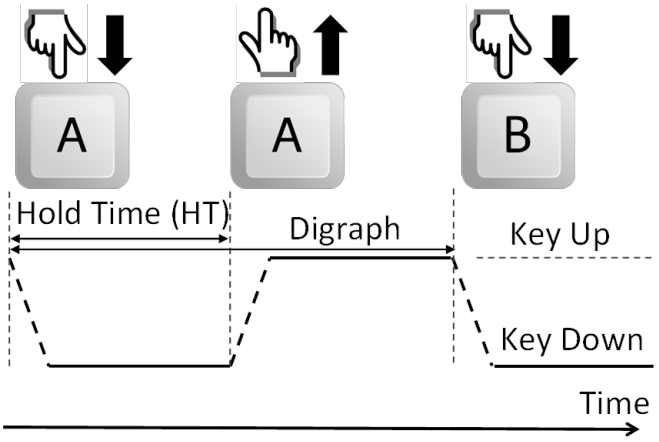


FIGURE 4.1: Graphical description of Keystroke dynamics features

The text entry accuracy metrics and error rate measures have been widely reported in literature, and there has been a consensus on the following set of definitions [74], [157]:

a) Error Rate of Minimum String Distance: Minimum String Distance (MSD) is the minimum number of keystrokes needed to transform transcribed text into presented text [158]. The error rate of minimum string distance ER_{MSD} is given by:

$$ER_{MSD} = (MSD(P,T))/(MAX(|P|,|T|)) \times 100\% \tag{4.1}$$

where presented text (P) is the experiment message, transcribed text (T) is the text that is written by a participant and the symbol $|.|$ is the size of text.

b) Erroneous KeyStroke Rate: It represents the ratio of the total incorrect keystrokes to the presented characters. The erroneous keystroke rate (EKSR) is obtained from:

$$EKSR = (IF + INF)/P \times 100\% \tag{4.2}$$

where incorrect fixed (IF) is the number of the incorrect keystrokes which is noticed by the participant and corrected while incorrect not fixed (INF) is the number of the incorrect characters that appear in the transcribed message.

c) Total Error Rate: The ratio of total incorrect characters (fixed and non-fixed) to the total characters is called total error rate (TER) and it is calculated from:

$$TER = (IF + INF)/(C + F + INF) \times 100\% \tag{4.3}$$

where correct (C) is the number of the correct characters in the transcribed message and fixes (F) is the number of the correction keystrokes (such as delete, backspace, and cursor movement) and, modifier keys (shift, alt, control, etc.)

d) Keystrokes per Character: It represents another important error related metric; its value can be related either to many easy-corrected errors, or to few arduous-corrected errors. In mathematical form, the keystrokes per character

(KSPC) is obtained from:

$$KSPC = |InputStream|/|T| \tag{4.4}$$

e) BackSpace Rate: It is an error related metric. Its value is calculated by counting number of backspace clicks in specified length of text. In mathematical form, the backspace Rate (BKS_R) is calculated from [159]:

$$BKS_R = number\,of\,backspace\,clicks\,in\,certain\,text/|T| \tag{4.5}$$

The last two metrics can be calculated based on counting specific keyboard characters (i.e delete and backspace keys) without the need of addressing each single character or error. This calculation method is considered as essential requirement to protect user privacy. These two metrics are therefore used in the proposed naturalistic study to quantify performance degradation based on the typing error rate [160].

4.3 System 1 (Keyboard Typing)

A new method for operator fatigue detection based on computer-keyboard natural typing style is proposed and implemented as shown in Figure 4.2. The changes in typing behaviour that come from the fatigue evolution are detected using the performance degradation-related metrics: HT and digraph. The proposed method comprises of several stages that are defined briefly, as follows:

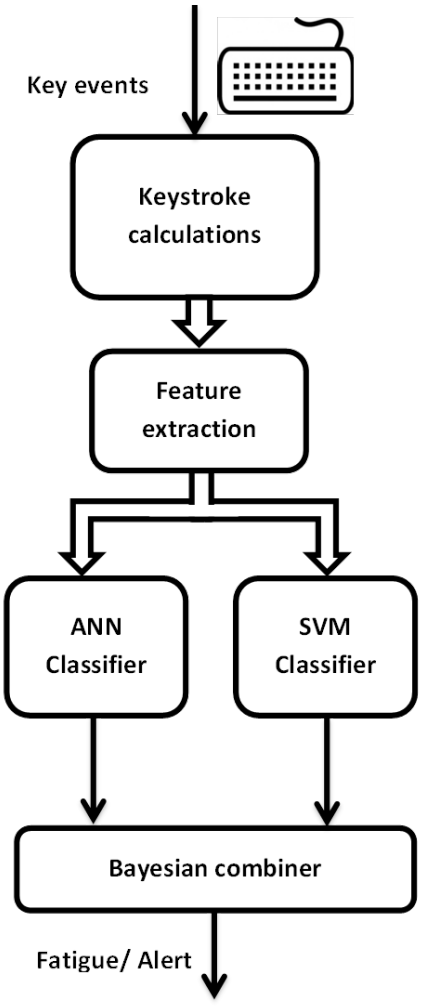


FIGURE 4.2: Block diagram of the implemented system

a) Key-events recording: The timing of key events (key press and key release) are recorded.

b) Keystroke calculation: Keystroke dynamics (i.e. HT and digraph) are calculated based on key events.

c) Feature extraction: Several statistical features, based on keystroke dynamics, are extracted (features are listed in Table 4.1).

d) Classification: The extracted features are used to detect the operator's fatigue status, using two classifiers (ANN and SVM). These classifiers are very common

in typing behaviour studies [74].

e) Bayesian combination: , The outputs are fed to a Bayesian combiner to improve overall classification performance. This combiner is one the most useful algorithms for combining multiple decisions [161] .

4.3.1 Dataset 1: keyboard typing with Karolinska sleepiness scale

This dataset is generated by the author with the aid of volunteer students at the University of Liverpool. 15 participants (13 males and 2 females) aged between 36 and 50 years participated in the development of this dataset; all of them have a good experience from spending at least two hour per day using a computer, but none of them is a professional computer typist. This dataset that is based on personal computer keyboard, is labelled with 9 fatigue levels.

Each participant is asked to install a software package which was originally developed by Massachusetts Institute of Technology researchers [75] to collect keystroke timing over two working days. No specific task are given to the participants to keep them in naturalistic environments. Moreover, the participants are asked to asses their fatigue level and record it on a separate MS Excel sheet with its timestamp, using Karolinska sleepiness scale (KSS), every one hour while they are typing. The experiment is carried out for two days with ate least of 3 hours of daily typing. The total typing time of all participants was around 55 hours.

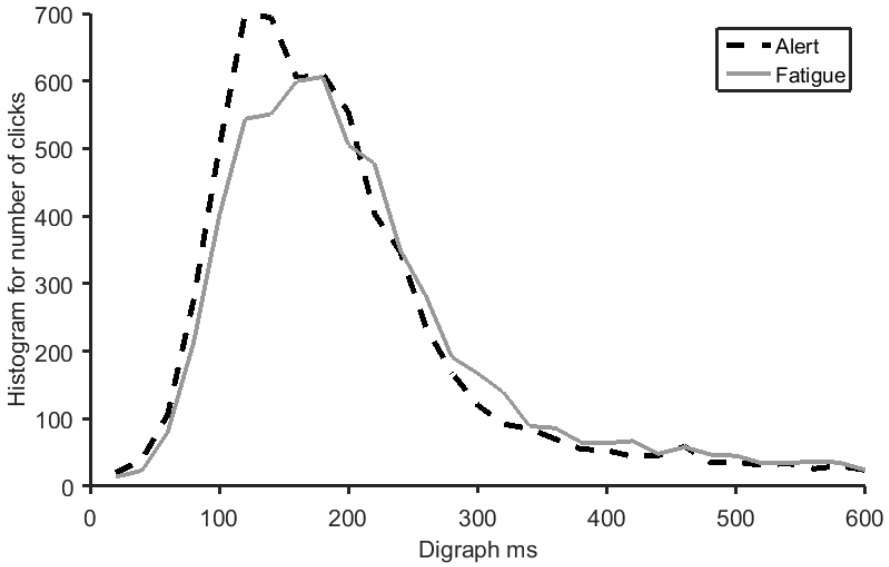
4.3.2 Feature extraction and labelling

The feature extraction method is equally applied to Datasets 1 and 2. Similar feature sets are extracted from both datasets except for an extra feature (called hour of day) which is extracted from Dataset 1. Generally, a small percentage of the key-timing events records shows two typing abnormalities: (i) the long periods of holding times and (ii) the long periods of digraph times. These abnormalities are therefore discarded prior to keystroke calculations.

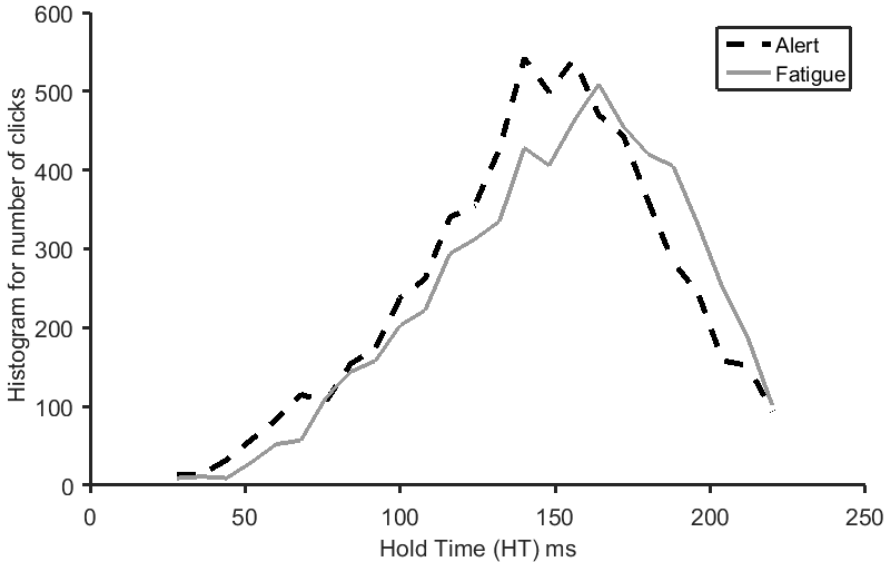
Three metrics of keystroke dynamics are calculated from both datasets: (i) hold time, (ii) digraph and (iii) release to release latency. Statistical analysis and visualisation is implemented on these metrics to select the best collection of fatigue-related features.

Histograms peaks provided in Figure 4.3(a) depict the variation of digraph times between alert and fatigue participant status. Resolution of the timer used in this experiment is 1μs and the variation in the digraph times is 50 ms. This time frame is considered adequate for the classifier to achieve a reasonable performance. The dashed lines belong to the participant in alert status that has peaks with fewer digraph times than the solid lines that belong to the fatigued participant. Similarly, Figure 4.3(b) demonstrates a similar trend of variation in HT histograms peaks between the alert and fatigue states of the participant.

Typing performance degradation due to fatigue effect can be captured by its reflection on error rate. Figure 4.4 shows the relation between one error rate measures (BKS_R) and KSS score. The BKS_R is normalised for each participant to neutralise the differences between participants and between keyboard types. A



(a) Digraph



(b) Hold Time

FIGURE 4.3: Example of keystroke metrics histogram for a participant

correlation coefficient (r=0.31, $p < 0.05$), which is statistically significant, shows an increase in error rate with fatigue growth.

Statistical analysis of the above-mentioned metrics generates some features. A total of 13 features are normalised with their labels (Alert or Fatigue). The metrics of keystroke dynamics and their corresponding features are summarised in

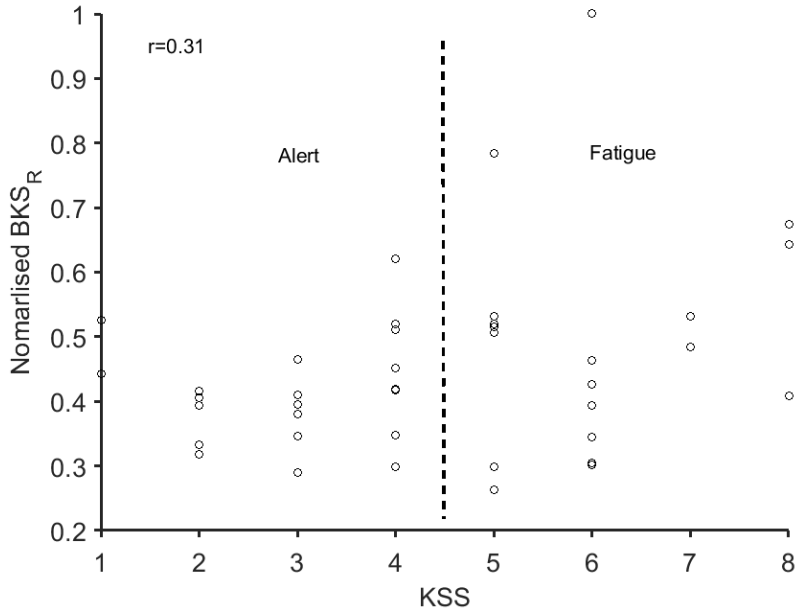


FIGURE 4.4: Relation between BKS_R and KSS score

Table 4.1.

TABLE 4.1: Keystroke metrics and their features

No.	Metircs	Features
1	Hold Time	mean, mode, median and standard deviation
2	Digraph	mean, mode, median and standard deviation
3	Release-to-Release (RR) latency	mean, mode, median and standard deviation
4	Hour of the day	(Only for Dataset 1)

The latency metric press-to-release (time between key press and consecutive key release) is found to be ineffective in terms of classification performance and thus excluded to reduce classifier complexity.

Two labels (Alert and Fatigue) are assigned to the calculated feature set. The "Alert" label is assigned to the record when its subjective assessment (KSS) is

within the first four fatigue levels (i.e. 1-4), and "Fatigue" label is assigned for higher levels (i.e. 5-9).

4.3.3 Fatigue detection

Outputs of the classifiers ANN and SVM are fused by a combiner to improve the classification performance. These fatigue detection techniques are described briefly as follows.

a) Artificial Neural Network

ANN is built and trained with training set of 65% of the available dataset while the rest 35% records are used for validation and testing phases. The proposed ANN is implemented with feed-forward ANN based on one input layer and one hidden layer and one single output unit with a tangent-sigmoid transfer function. Moreover, many trials are adopted for ANN structure to improve the network performance such as changing the number of hidden layer and the nodes in each layer as well as changing in training algorithms and decision transfer function.

b) Support Vector Machine

Support vector machine (SVM) is a supervised machine learning classifier using hyper planes to discriminate output classes. SVM uses a training algorithm to maximise the margins between the classes based on least square method [162]. Similarly to ANN, SVM has a wide range of applications and it is used in large number of classification problems. SVM has gained very good reputation of its superior performance over other classifiers but with slower time in training phase [163].

In this work, the SVM is used to classify the users depending on their collected data into two classes (alert and fatigued). These classes are labelled to indicate the alert/fatigued status. A 2-class SVM classifier is found adequate to detect the fatigue status of the keyboard users. The radial basis function is used as a kernel for the SVM classifier. The dataset is divided into two subsets: the training subset of (75%) and the testing subset of (25%). The latter set is used to evaluate the classifier performance.

c) Bayesian combiner

Bayesian combiner (BC) uses the base classifiers posterior probabilities and their outputs which are provided by the SVM and ANN classifiers to improve the overall classification performance. The posterior probabilities of both classifiers are calculated using confusion matrices. The soft output (not the labels) of the ANN classifier is weighted by the posterior probability of the same classifier. The scores of SVM classifiers represent the distance of the classifier output from the support vector. These scores are generated and weighted by its posterior probability. Then, the summation of the two weighted outputs is used to generate the combined decision of the whole classification stage.

4.3.4 Results of Dataset 1

The receiver operating characteristic (ROC) performance metric is illustrated in Figure 4.5, which compares the ROC curves of both ANN and SVM classifiers. The two curves show close area under curve (AUC) values in agreement with Table 4.2 which reflects the close classification performance. In this experiment, sensitivity of the ANN classifier (74.4%) outperforms that of the SVM classifier (66.0%). This observation is probably due to classes overlapping and gradual

change between classes in this dataset. In addition, the sensitivity values of both classifiers are dropped below the specificity values. This is mainly due to the unbalance of classes shares of the dataset.

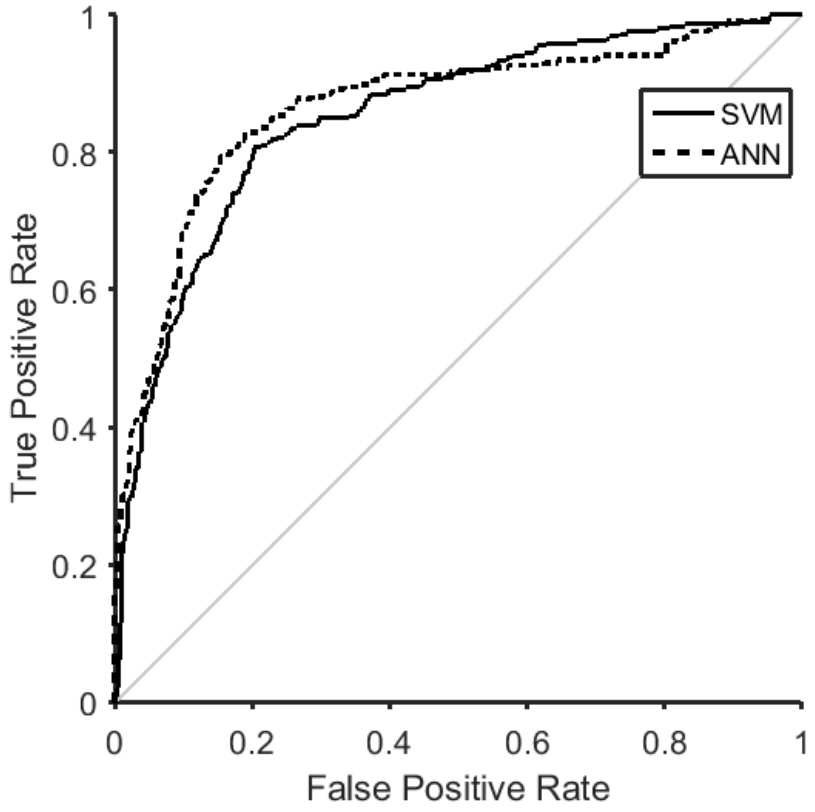


FIGURE 4.5: ANN and SVM classifiers performance (ROC metric) on Dataset 1

Figure 4.6 shows a confusion matrix for the Bayesian combiner post provision of 25% of the dataset. The performance metrics have demonstrated an accuracy of 82.5%, a sensitivity of 75.1% and a specificity of 87.4%. The fatigue-class, which can be calculated from the fusion matrix, is around 40% of the total test-set, and this could justify the low values of sensitivity.

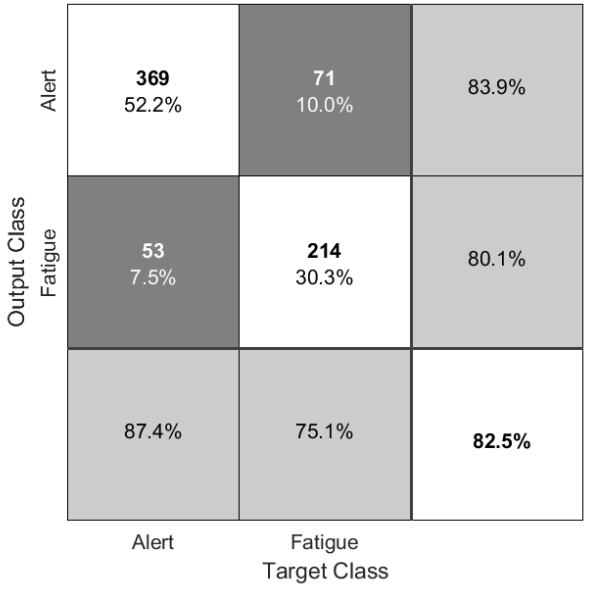


FIGURE 4.6: Bayesian combiner confusion matrix of Dataset 1

TABLE 4.2: Summary of system performance metrics of Dataset 1

Performance metrics	ANN(%)	SVM(%)	BC(%)
Accuracy	81.8	77.9	82.5
Sensitivity	74.7	66.0	75.1
Specificity	86.5	86.0	87.4

4.3.5 Dataset 2: keyboard typing (two states: Rested and Sleepy)

Although sleepiness and fatigue are two distinct phenomena, they are used interchangeably because they are interrelated and have similar effect on psychomotor skill performance [164], [165].

Dataset 2, which is publicly available [75], was generated by Giancardo and his colleagues to quantify the psychomotor impairment generated by sleep loss. Unlike Dataset 1, this dataset is already pre-labelled with rested/sleepy status, thus it

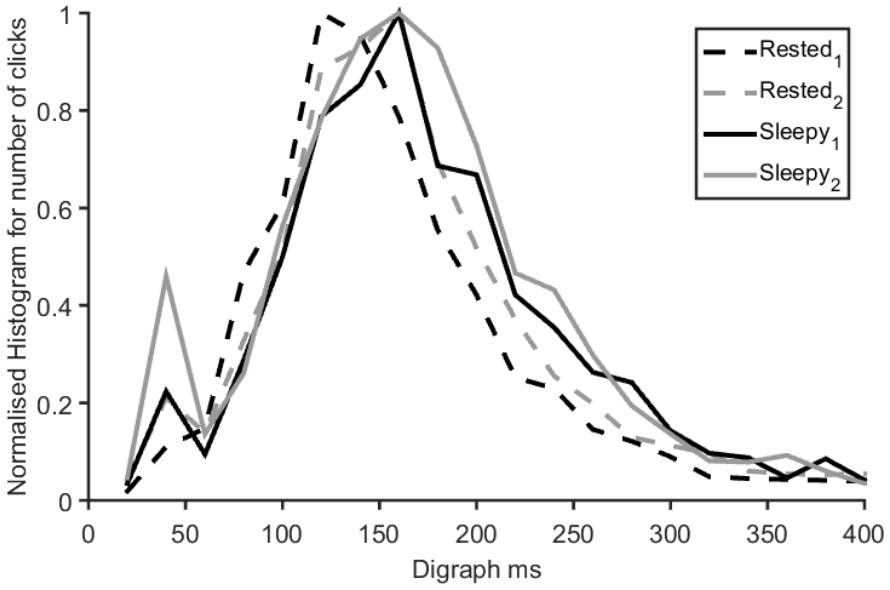
fits well with the need of System 1. In this dataset, 14 volunteers (seven males, seven females) aged between 20 and 39 years (mean 30.8, standard deviation 4.4) participated in the development of this dataset. All the participants had a good experience in typing of at least one hour per day, but none of them is a professional computer typist. Their typing style varies but consistent for each participant.

To ensure naturalistic environment scenario, each participant was asked to use his/her own laptop to type a text from his/her own choice from Wikipedia website for 15 minutes. The first session was conducted in the day time when participants felt rested and were asked to stay awake until they felt sleepy, then were allowed do the night (fatigue) session. This schedule was performed one more time after one week. At the end of this experiment, around 14 hours of key timestamps had been stored in the dataset. Each participant conducted four sessions of typing. Each session was divided into eight subsets as a consequence of the total number of data set records are 448 records. Each record includes a timestamp for key press and key release events. Moreover, each record is labelled with one of two labels (Rested and Sleepy). It is labelled as rested when the data is collected from day sessions while it is labelled as sleepy when data is collected from night sessions with sleep inertia inducing.

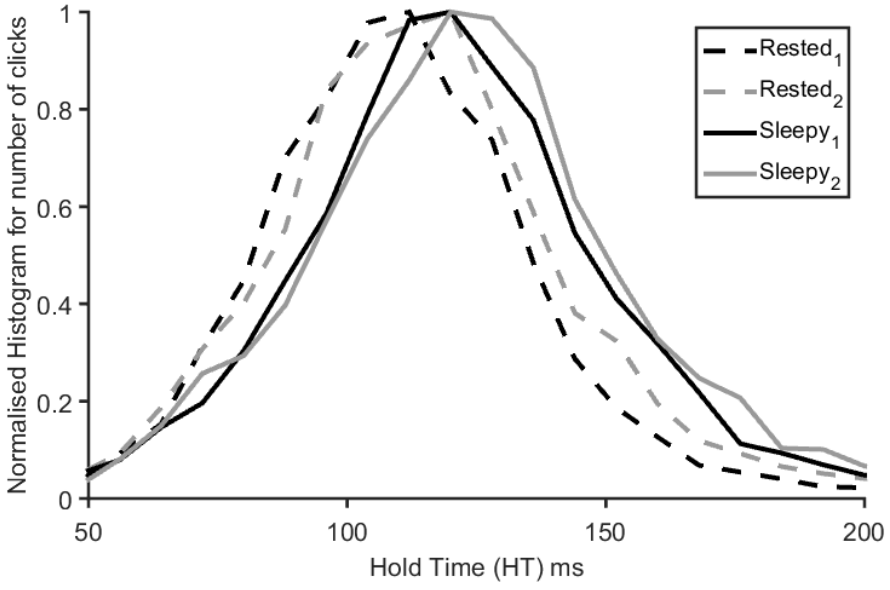
4.3.6 Feature extraction and labelling

Dataset 2 is generated based on two extreme cases of participants mental states: rested and sleepy. Figure 4.7 demonstrates examples of statistical measures, which belong to the same participants, covering four sessions. Histograms peaks appear in Figure 4.7(a) depict the variation of digraph times between the rested and sleepy

participant. The dashed lines that are of the rested participant have peaks with lower digraph times than the solid lines which belong to the rested participant. Similarly, Figure 4.7(b) demonstrates a similar trend of variation in HT histograms peaks between the alert and fatigued participant.



(a) Digraph



(b) Hold Time

FIGURE 4.7: Example of keystroke metrics histogram for a participant, a 4-session experiment

Same features of that listed in Table 4.1 are extracted for Dataset 2. Hence, a total of 12 features (i.e. 4 features per metric) are normalised. The rested and sleepy labels are assigned to each record in Dataset 2, depending on time period (day or night) of the experiment.

4.3.7 Results of dataset 2

Figure 4.8 shows the ROC for both the ANN and SVM classifiers. As illustrated, the AUC of the SMV classifier is better than that of the ANN classifier. Moreover, the SVM performance exhibits higher values of accuracy (91.07%) than ANN, as shown in Table 4.3.

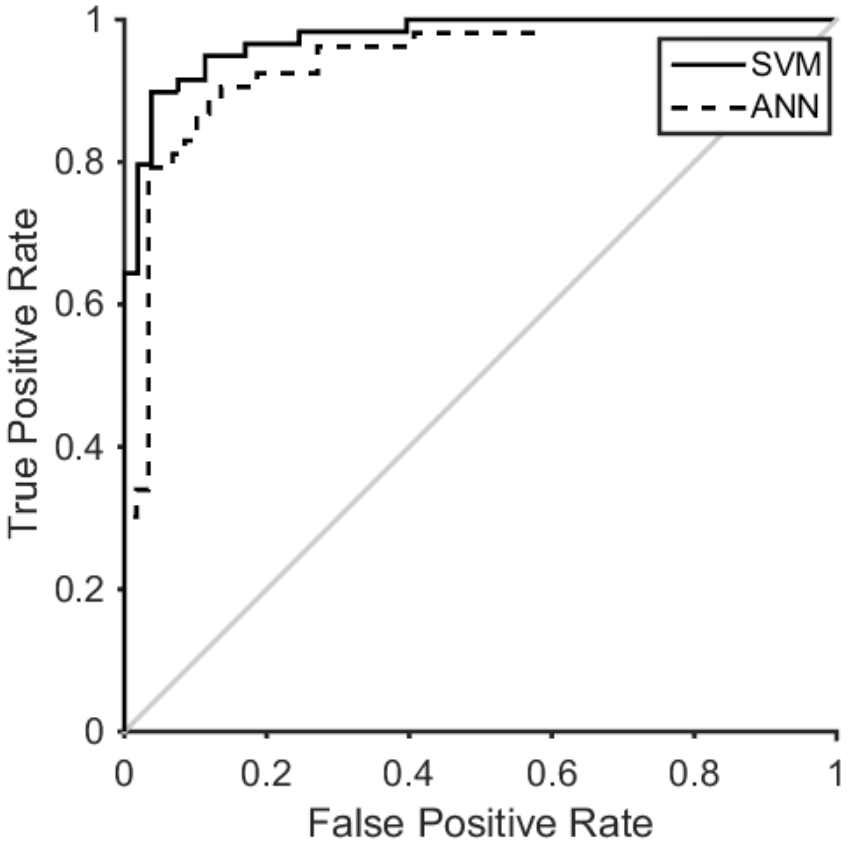


FIGURE 4.8: ANN and SVM classifiers performance (ROC metric)

Figure 4.9 shows the confusion matrix for the Bayesian combiner after feeding it with the test subset data (25% out of the entire dataset). It clearly demonstrates higher performance than that of the ANN and SVM individually. It also slightly outperforms the state of the art findings (94% vs. 92%), reported in [75].

FIGURE 4.9: Bayesian combiner confusion matrix of Dataset 2

Table 4.3 provides classifiers performance metrics for both ANN and SVM algorithms as well as for the BC. The SVM demonstrated a slightly higher performance than the ANN. Also the BC results outperform the individual results of each base classifier (Figure 4.9) of the performance metrics used in this study, the sensitivity metric has been the most outstanding. Around 10% improvement has been achieved in the sensitivity metric of the ANN classifier and around 5% for the SVM classifier, as summarized in Table 4.3.

Table 4.3: Summary of system performance metrics of Dataset 2

Performance metrics	ANN(%)	SVM(%)	BC(%)
Accuracy	87.5	91.07	94.0
Sensitivity	84.9	89.83	94.6
Specificity	89.8	92.45	93.3

4.4 System 2 (Smartphone Typing)

The results and findings of System 1 have demonstrated a relation between the fatigue status and the keystroke metrics. Another relation was also demonstrated in Figure 4.4 between the fatigue status and the error rate. Based on these findings, a performance-degradation detection system using smartphone typing is proposed and implemented.

The proposed system, that comprises several stages, is shown in the block diagram of Figure 4.10. The collected dataset is initially analysed and used to calculate some fatigue-related metrics including keystroke dynamics and inertial sensors measures as illustrated in the left-hand side branch of Figure 4.10. The dataset is initially pre-processed through de-noising inertial sensors data. Next, the fatigue-related metrics and features are extracted and selected features are then fed to a 2-class classifier.

In the right-hand side branch of the diagram, the text-entry dataset is used to calculate the error; keystroke per character (KSPC) that is in turn used as an accuracy metric. Next, a median threshold is set, depending on the KSPC, and used to label the entire dataset. The labelled dataset is then used as a ground truth at the training stage of the classifier. That eventually classifies the user's

fatigue/alert status. In this study the ANN and SVM are used as in System 1.

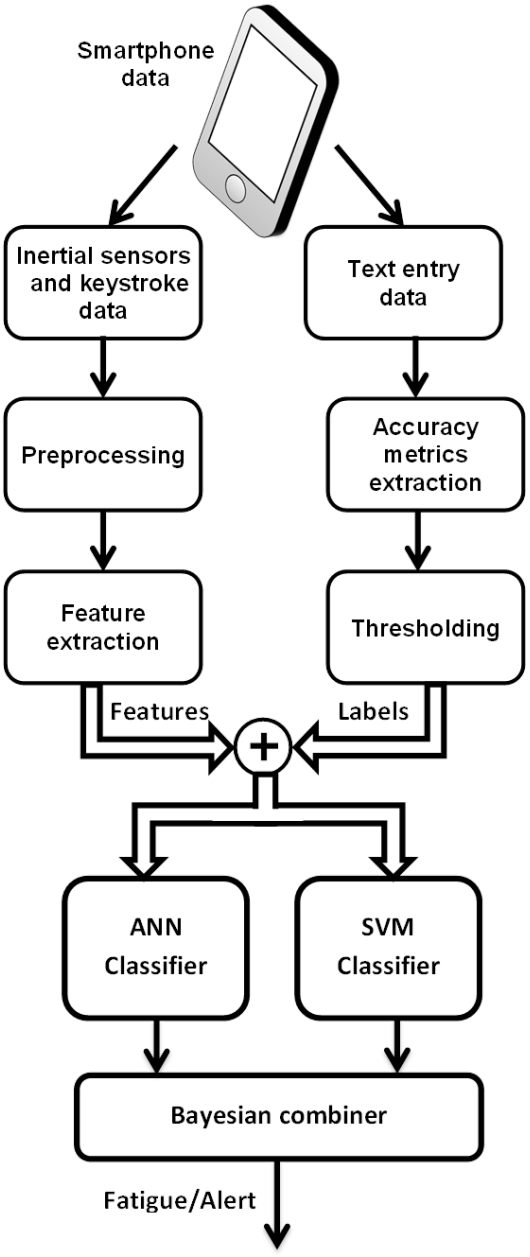


FIGURE 4.10: A block diagram for the developed fatigue detection system using smartphone

4.4.1 Dataset 3: smartphone typing

Publicly available dataset [166] is adopted in this system. The original purpose of this dataset was to investigate behavioural features of typing style to continuously

authenticate smartphone users [78]. The data collection method of this dataset which was implemented using smartphone application named Hand Movement, Orientation and Grasp (HMOG), is reported in [167]. It was collected from 100 smartphone users (47 females and 53 males). Each user was asked to interact with smartphone with three types of tasks (reading session, writing session and using map session).

In System 2 , the writing session of the HMOG is of a particular interest. Its text data elements are used to extract the error rate as a performance metric (as described in Section 4.2). While the corresponding keystroke dynamic metrics and inertial sensor measures are processed to extract features. The timing metrics (hold and digraph times) and the inertial sensors metrics (3D accelerometer and 3D gyroscope measures, with 100 Hz sampling rate) are chosen to capture the performance pattern of the user throughout several writing sessions.

Each one of the 100 participants has eight writing sessions to be completed at different periods. In each session, which has an average of 1193 taps and lasts for 11.6 minutes, the participant is invited to type three pre-specified statements using his/her smartphones keyboard. In order to increase the number of records, each session is divided into 5 segments. The size of dataset used in this study is therefore calculated from:

$$Dataset\ Size = 100\ Participants \times 8\ Sessions \times 5\ Segments = 4000\ records \tag{4.6}$$

4.4.2 Feature extraction

Two groups of metrics is extracted from smartphone users when they interact with virtual keyboards. First, time-event based metrics from which the individual typing pattern can be obtained for each user. This group of metrics is also called keystroke dynamics. Such a pattern can hold much information relevant to the behaviour of a particular user. The hold time (HT) and digraph time (two consecutive keys), which are shown in Figure 4.11, are the basic features that are extracted from keystroke dynamics.

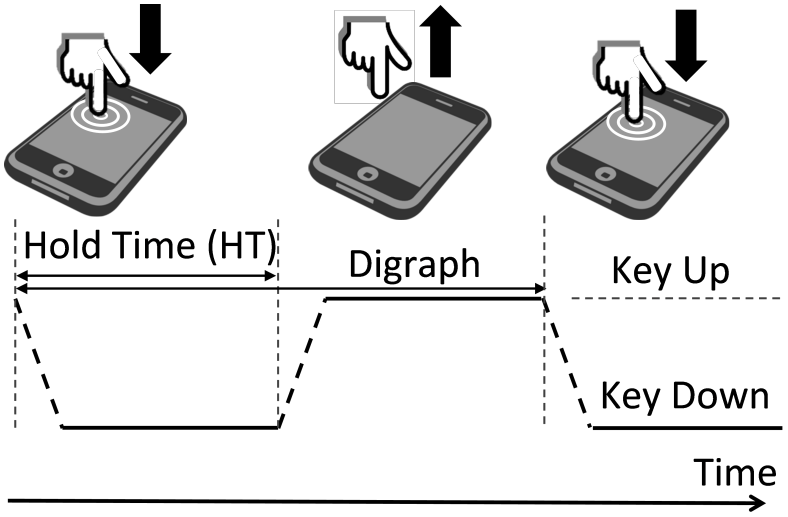


FIGURE 4.11: Graphical description of Keystroke dynamics features of smartphone virtual keyboard

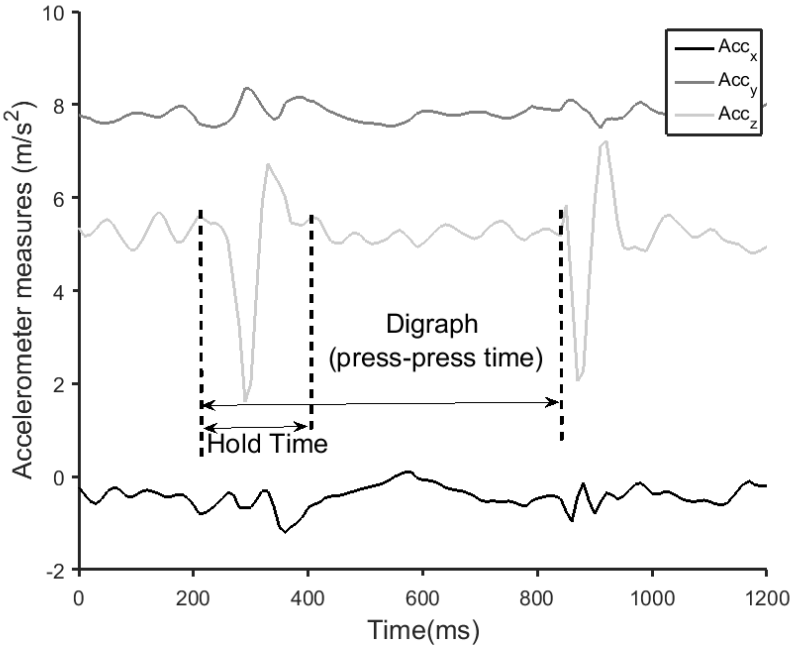
Second, micro-movements metrics can be detected through the touchscreen interactions. Typing on smartphone touchscreen generates micro-movements that can be captured by the built-in inertial sensors (i.e. accelerometer and gyroscope). The recorded micro-movements are used in this work to extract statistical features of users typing style to ward off privacy issue that may arise if typing data is used instead. The accelerometer data, which is generated as an echo for typing two letters on touch screen, is demonstrated in Figure 4.12(a). The grey graph, which represents the z-axis accelerometer data, shows a noticeable fluctuation

in accelerometer measures while typing a letter. These changes helps extracting some statistical and frequency domain features (such as dominant frequencies and sub-band energy ratios). Another example of touchscreen user interaction is the scrolling action for reading multi-page document, Figure 4.12(b).

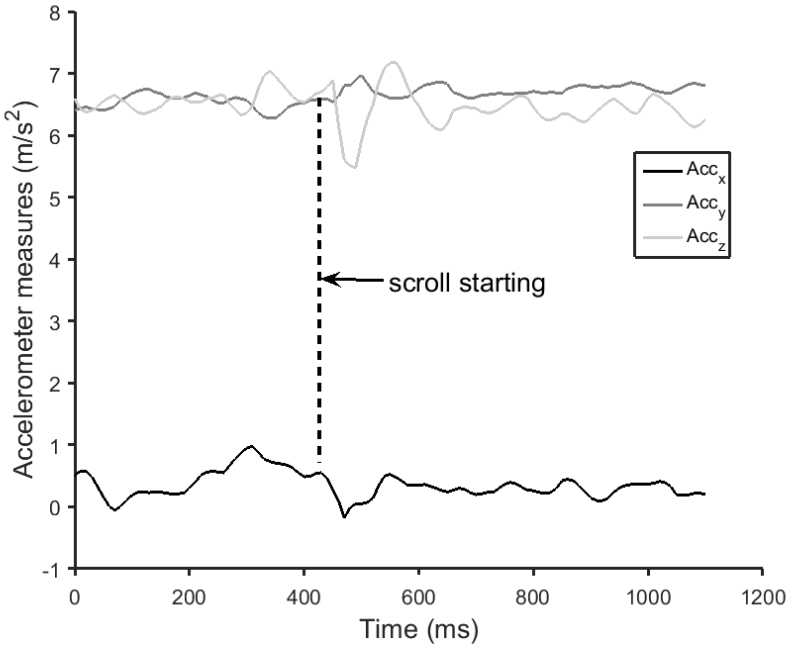
The timing data of text entry keystroke is recorded with a high precision timer (with a resolution of 1μs) to ensure accurate and reliable data. Statistical representations of keystroke dynamics (i.e. the hold-time and the digraph time) are calculated to build a set of features. Some of these features such as HT mode, and digraph mode were found less effective in terms of classification results and thus eliminated. The raw data from 3D accelerometer and 3D gyroscope are pre-processed with average window filter of 30 samples to reduce noise level. Some features (such as median and mode) are eliminated because of redundancy and has no effect on classifier performance. The standard deviation of the hold time has changed between different sessions which is an example of typing feature, as illustrated in Figure 4.13. The metrics and selected features are summarised in Table 4.4.

Table 4.4: Keystroke metrics and their features

No.	Metircs	Features
1	Hold Time	mean and standard deviation
2	Digraph	mean and standard deviation
3	Accelerometer measurements	mean, standard deviation, dominant frequency and sub-band energy ratios
4	Gyroscope measurements	mean, standard deviation, dominant frequency and sub-band energy ratios



(a) Typing



(b) Scrolling

FIGURE 4.12: Examples of triaxial accelerometer data for action responses on smartphone touchscreen (This graph is generated using dataset 3 [166])

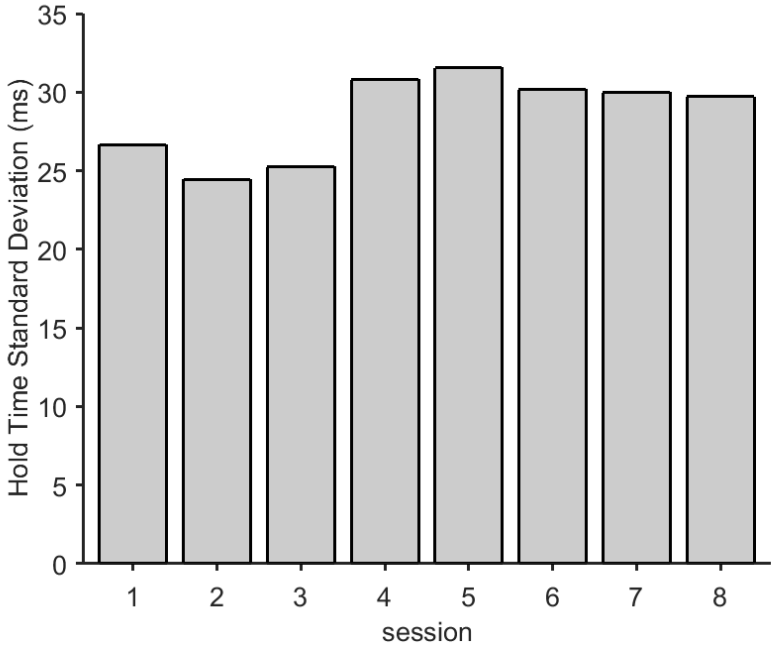


FIGURE 4.13: Hold-time standard deviation for different sessions

4.4.3 Data labelling

Dataset 3 (described in Section 4.4.1) does not have fatigue assessment informations. Hence, performance degradation method is applied to label thee extracted features with two labels: likely alert and likely fatigue. The relation between fatigue levels and the BKS_R (shown in Figure 4.4) has been demonstrated previously in Section 4.3.2.

Two examples of performance metrics for a single participant are demonstrated in Figures 4.14 and 4.15. First, the variation in the number of backspace clicks is presented in Figure 4.14. The backspace clicks is used as an indicator for fixed errors of the user typing. Second, the keystroke per character metric (KSPC) is presented in Figure 4.15. A strong correlation (r=0.72 , $p < 0.01$) for the entire dataset between these two metrics can be noticed in Figure 4.16 , thus, validate the effectiveness of the selected features and performance labelling metrics.

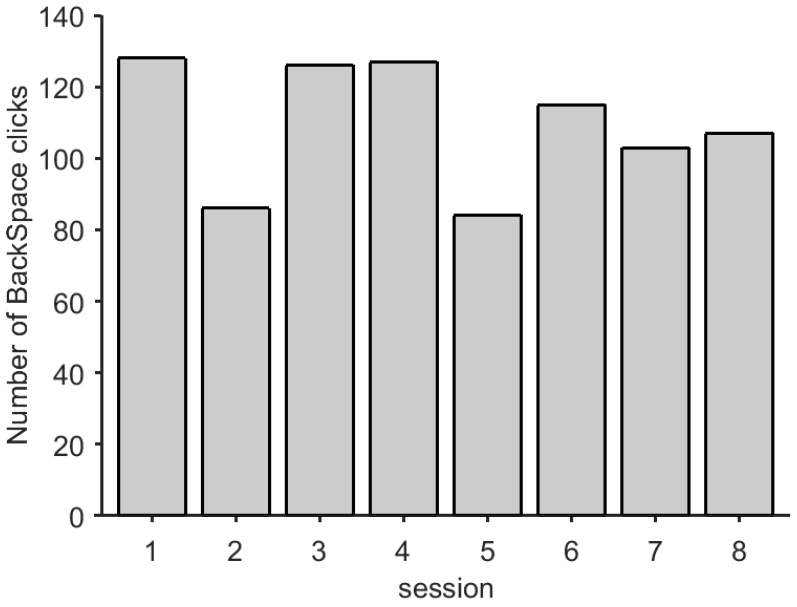


FIGURE 4.14: Backspace-clicks for different sessions

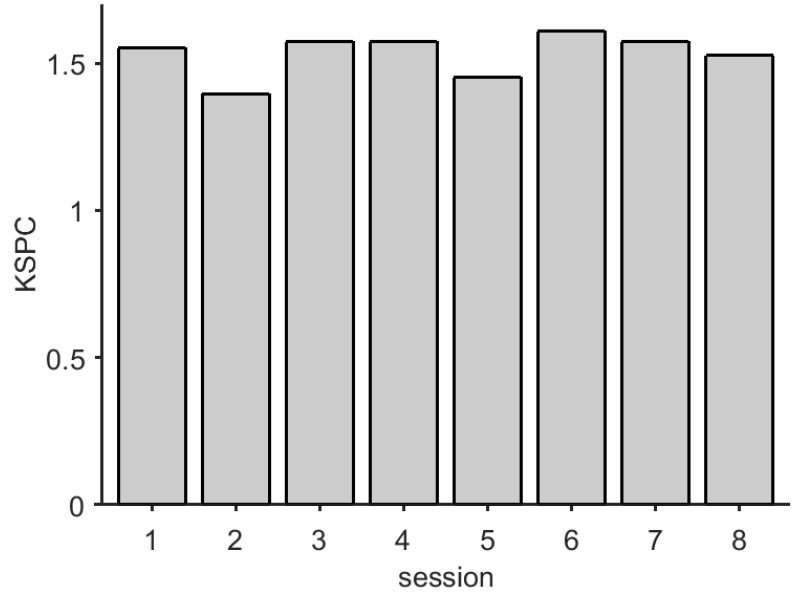


FIGURE 4.15: KSPC for different sessions for a participant

BKS_R, as an error fixer, is a part of KSPC and thus, KSPC is considered a more representative error measure. Due to availability of input stream and transcribed message sizes, KSPC is used with smartphone dataset as an accuracy metric. KSPC range varies between participant because they have different accuracy levels, therefore, a median normalisation [168] is performed. Next, a normalised median

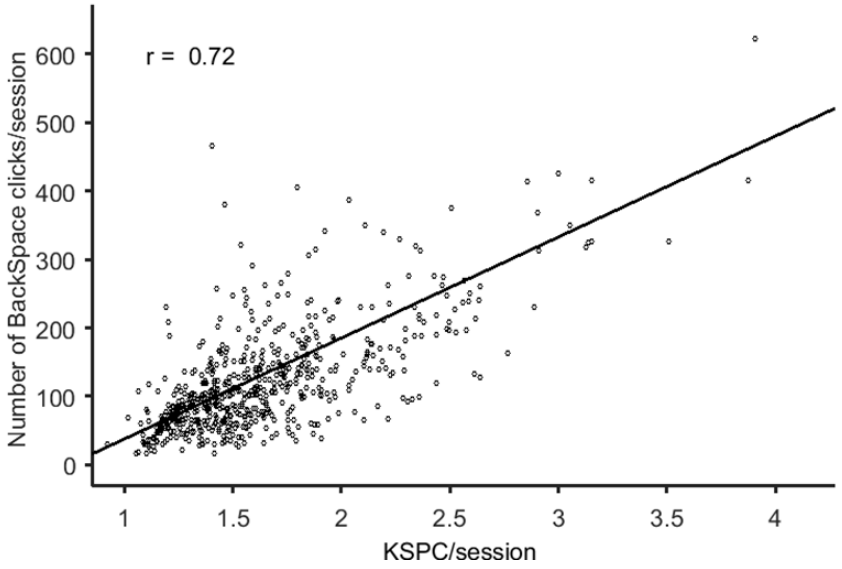


FIGURE 4.16: Relation between two typing accuracy metrics (KSPC and backspace-clicks)

threshold is chosen to divide sessions into two classes: (i) a likely alert and (ii) a likely fatigue. Those two classes are used to label obtained features' dataset. Figure 4.17 shows a histogram for these sessions as well as the threshold setting.

4.4.4 Results of Dataset 3

The developed ANN and SVM classifiers are trained and tested in a similar manner to that previously reported for System 1. Performance of the developed SVM classifier as compared to the ANN classifier performance is shown in the ROC graph of Figure 4.18. The obtained numerical results for this test are summarised in Table 4.5. It can be noticed that the SVM has a superior performance when compared to that of the ANN since the area under the curve of the ROC for the SVM is larger than that of the ANN.

The SVM classifier demonstrated a superior performance as compared that of the ANN classifier in identifying fatigue/alertness conditions. When compared to the

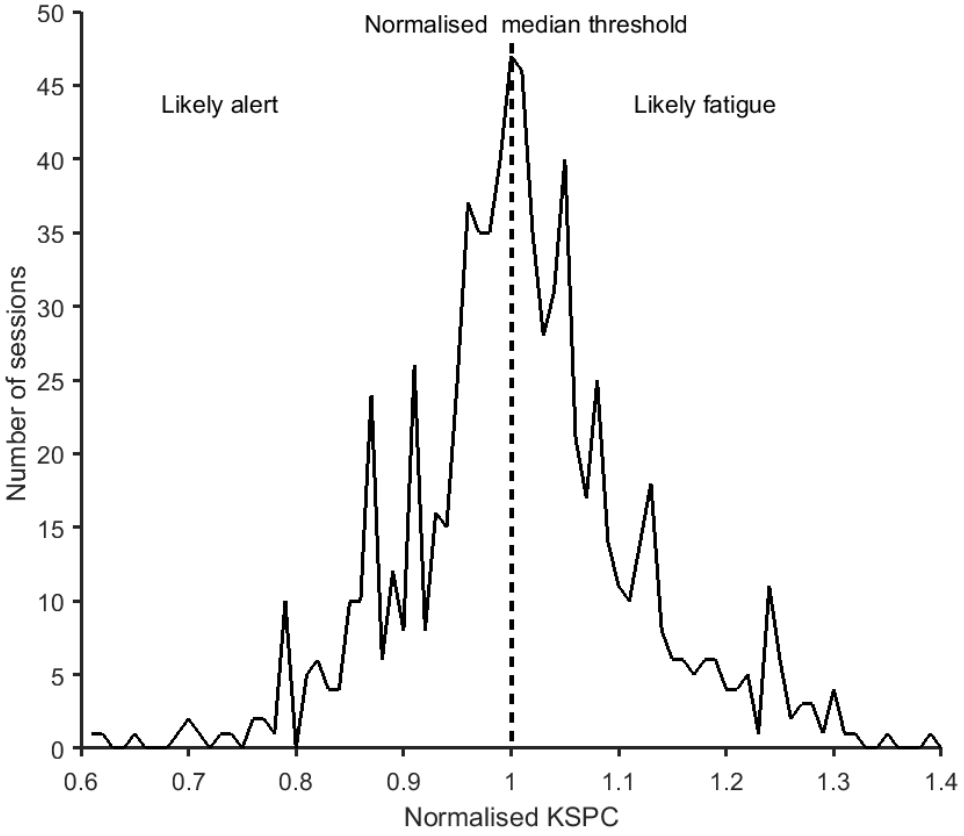


FIGURE 4.17: Threshold-setting for a 2-class labelling

ground truth dataset, SVM demonstrated accuracy of 62.3%, sensitivity of 60.1% and specificity of 64.7%

Figure 4.19 shows a confusion matrix of the BC with a test subset of data (25% out of the entire dataset). The performance metrics show an accuracy of 63.9%, a sensitivity of 65.5% and a specificity of 62.2%. In this system, the BC shows slight improvement over the SVM performance metrics in terms of accuracy and sensitivity while there is slight degradation in specificity metric due to the low performance metrics of ANN.

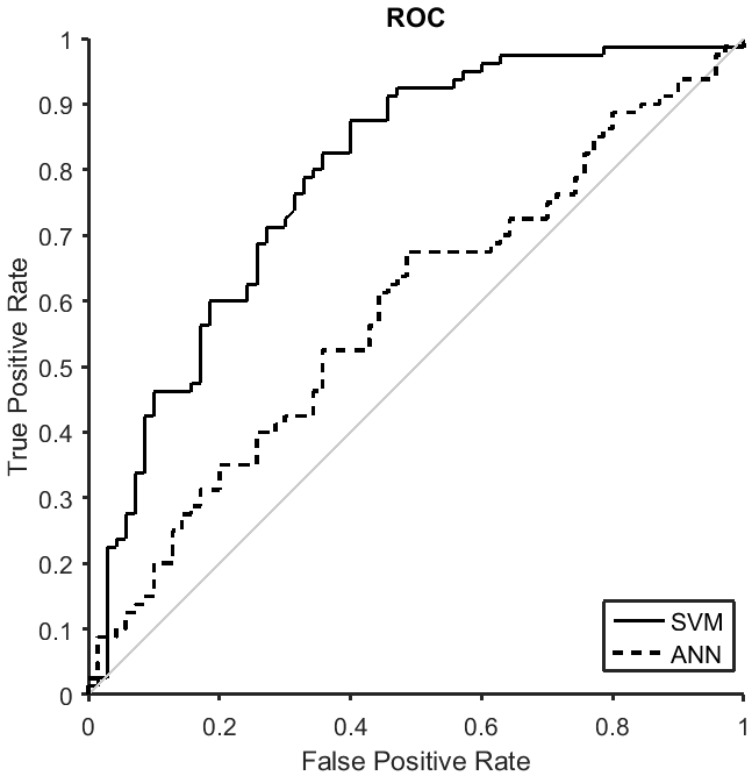


FIGURE 4.18: ANN and SVM classifiers performance (ROC metric) of smartphone dataset

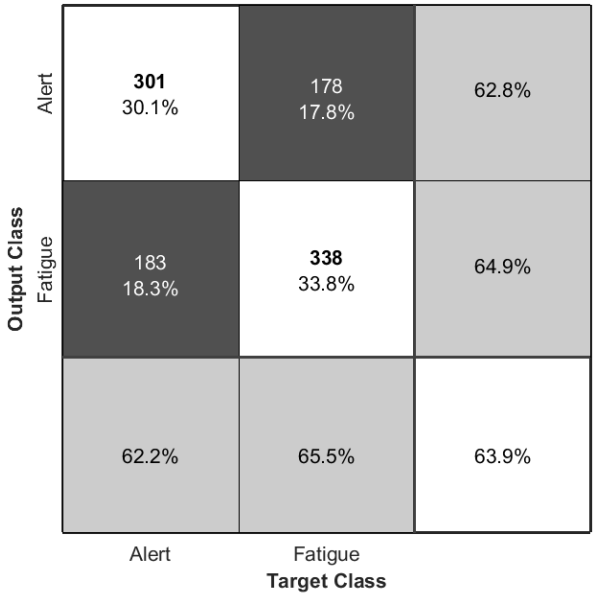


FIGURE 4.19: Bayesian combiner confusion matrix of smartphone dataset

Table 4.5: Summary of system performance metrics of smartphone dataset

Performance metrics	ANN(%)	SVM(%)	BC(%)
Accuracy	57.6	62.3	63.9
Sensitivity	63.2	60.1	65.5
Specificity	51.7	64.7	62.2

4.5 Discussion

Both Systems 1 and 2 have demonstrated the ability to use keyboard typing for operator fatigue detection. Although accuracies of classification performance vary between these experiments, all classifiers agree in the general trend. Table 4.6 shows promising performance values especially when taking in consideration that these experiments are conducted in realistic rather than laboratory environments.

Despite that performance of the results obtained from Dataset 1 is slightly lower than that of Dataset 2, its findings can be considered of a significant value since it depends on direct fatigue assessment scale. Dataset 2 demonstrated the best classification performance over the other datasets used in this study, as shown in Table 4.6. This is mainly due to the fact that the generation of Dataset 2 is based on collecting data from two extreme cases (rested and sleepy). The results generated from System 2 showed the least detection performance because the ground-truth was calculated based on performance degradation metrics and not assigned directly as in the case of other datasets.

Table 4.6: Summary of the three datasets detection performance metrics

Performance metrics	Dataset 1(%) Keyboard with KSS	Dataset 2(%) Keyboard (rested and sleepy)	Dataset 3(%) Smartphone
Accuracy	82.5	94.0	63.9
Sensitivity	75.1	94.6	65.5
Specificity	87.4	93.3	62.2

4.6 Conclusion

Keyboard typing task, the daily activity, is considered as behavioural measure based on operator typing style. The naturalistic environment of typing activity is the reason behind the selection of this task for this work. The implemented system is based on actual fatigue/alertness status collected from the users of Dataset 1. The obtained results demonstrated a promising performance taking into account the challenges associated with conducting this study in naturalistic environments.

Two systems and three datasets were presented and discussed in this chapter to demonstrate the ability of typing measures to capture operator fatigue status. A 2-stage fatigue detection method based on (i) ANN and SVM classifiers and (ii) Bayesian combiner were designed, developed and tested successfully. Keystroke dynamics metrics (i.e. hold time, digraph and release to release latency) with their features and time of day are used to to discriminate alertness status.

Dataset 2 helped development of an enhanced method for operator fatigue detection based on computer-keyboard typing style. This method has been implemented successfully using an existing dataset for psychomotor impairment

detection. As the operator fatigue status mainly depends on sleep needs, the developed system (System 1) captures the typing style changes that reflect changes in the fatigue level. Utilisation of the combiner in this study has proved to be effective as compared to the utilisation of the ANN and SVM classifiers alone. The obtained detection accuracy of Dataset 2 (94%) outperforms the accuracies of other two datasets and the state of the art results for this dataset.

Dataset 3 is utilised in this work to implement a smartphone texting performance degradation detection approach. Despite the relatively low detection accuracy associated with Dataset 3, this approach still shows the ability of detecting operator performance based on daily use of his/her smartphone.

The work in this chapter has fulfilled its main objectives, however, it also exhibited some limitations, including: (i) the type of participants was limited to university students due to time and funding restrictions and (ii) the size of data in the fatigue and alert classes was unbalanced. This was mainly due the fact that the participants were not willing to to carry out typing when they become tired or fatigued.

Chapter 5

Fatigue Detection - Naturalistic Driving

5.1 Introduction

The behavioural approach, which is of particular interest in this study, depends on the collection of symptoms of fatigued drivers indirectly through their driving style. Unlike the biological and visual approaches, this approach can be considered non-intrusive since it does not directly collect drivers data that may violate their privacy.

This chapter proposes a new method for the automatic fatigue detection algorithm using smartphone inertial sensors. The proposed method captures driving behaviour data in naturalistic environments through two inertial sensors (triaxial accelerometer and triaxial gyroscope). It should be mentioned here that all the experiments of this study are carried out on motorways since fatigue typically

develops in monotonous driving environments [169]. The proposed method (i) utilises inertial sensors rather than both inertial and magnetic sensors, (ii) overcomes the error caused by horizontal plane misalignment between the vehicle and the Earth in case of vehicle tilting, and (iii) implements data collected from real participants rather than simulated data. This chapter presents the framework and experiment design of the driver fatigue detection system including participant description and data collection method. It also includes details of a proposed new vehicle heading estimation based on smartphone sensors and details several sets of results including attitude estimation, vehicle heading and classification results with a discussion of the findings.

5.2 Smartphone Inertial Sensors and Quaternion Mathematics

Smartphones are quickly becoming the most popular and widespread form of personal communication worldwide. The increased usage of these devices over the world along with continuous advances in mobile technology and application development have led to handheld devices becoming a new way of data collection and management for a wide range of applications including fatigue detection [150], workflow management [170] and mobile health applications [171]. Some companies are already using mobile applications to deliver increased customer satisfaction and productivity as well as to help their employees to perform their work faster [172]

Newer smartphones are equipped with a range of sensors including inertial sensors. A wide range of important mobile applications rely on these inertial sensors,

which are embedded in smartphones or tablets to work properly. Figure 5.1 illustrates two inertial sensors, a triaxial accelerometer and a triaxial gyroscope. The smartphones accelerometer that measures the acceleration forces is used to calculate three key parameters: (i) the gravity vector, (ii) the heading and (iii) the horizontal plane. The gyroscope sensor measures the horizontal plane as well as the angular velocity from which the vehicles rotational angle is calculated. The horizontal plane is typically calculated using the smartphones magnetometer [173] that measures the Earth's magnetic field. It also measures the relative orientation of the smartphone to Earth's magnetic field with the aid of other sensors that are used to fix the uncertainty readings. In this study, however, only two inertial sensors (i.e. triaxial accelerometer and triaxial gyroscope) are considered adequate for estimation of the vehicle's attitude, thus eliminating the need for the triaxial magnetometer and its associated inaccuracy, as explained earlier.

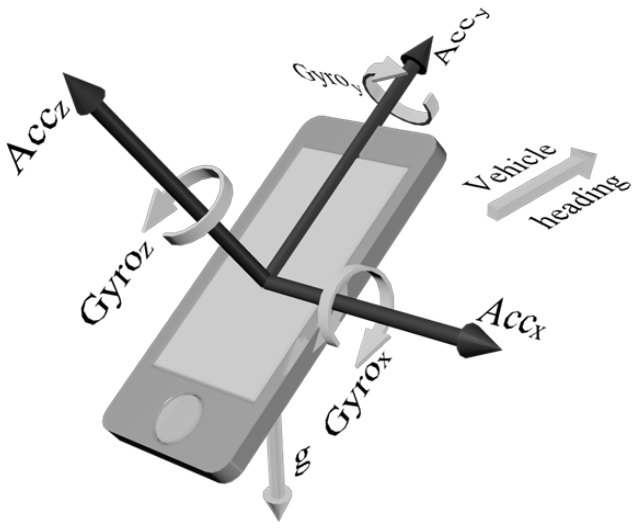


FIGURE 5.1: Accelerometer and gyroscope coordinates of the smartphone sensors

Utilisation of inertial/magnetic sensors existing in smartphones has been recently reported in several applications including detection of driving behaviour [85], drunk driving [174], aggressive driving [175] and others. Most of these applications

rely on the inertial as well as magnetic sensors, which are prone to disturbances caused by magnetic fields inside the vehicle [176]. In addition, these studies were based on simulation tests in which it was assumed that the horizontal plane of the vehicle is always aligned to the Earths horizontal plane and such an assumption may reduce the accuracy in orientation calculation at the event of tilting dynamic behaviour of the vehicle. To the best of our knowledge, utilisation of mobile inertial sensors in fatigue detection has not yet been reported in the literature.

A driver swerve, which is a fatigue-related feature [85], is an important metric that can be derived from smartphone inertial sensor data. Swerves can be quantified based on lateral car movement in the horizon plane, and thus inertial sensors can help with producing swerve related data. A driver usually keeps his/her smartphone either in his/her pocket or in any place close to him/her in the car and is unlikely to align the smartphone with the car axes. Estimation of smartphone orientation is a basic process to calculate the horizontal components of car swerves. Steering angle and steering angle rate measures are well-established research approaches that are among driver fatigue metrics [83]. The mathematics used in this work relies mainly on quaternion conversions to calculate the required fatigue-related features.

Several ways can be implemented to calculate three-dimensional object orientation such as rotation matrix and Euler angles methods [177]. To avoid the associated drawbacks such as non-linearity, gimbal lock effect and high computational cost [178], [179], the quaternion representation of rotation is used in this work. The quaternion is a mathematical number system that uses three-dimensional complex numbers. This system was developed initially by William Rowan Hamilton in 1843 and is mostly recruited in 3D space applications [180]. Mathematically, the

quaternion vector is given by

$$\mathbf{q} = q_1 + iq_2 + jq_3 + kq_4 = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} \cos(\frac{1}{2}\theta) \\ \sin(\frac{1}{2}\theta)\bar{v_x} \\ \sin(\frac{1}{2}\theta)\bar{v_y} \\ \sin(\frac{1}{2}\theta)\bar{v_z} \end{bmatrix} \tag{5.1}$$

where $i^2 = j^2 = k^2 = ijk = 1$; q_1 is the scalar part of the quaternion vector; q_2, q_3, q_4 are the three dimensions components of the quaternions; and θ is the rotation angle around the rotation vector $\mathbf{v}$. Although the smartphone sensors have a fixed alignment with the smartphone body, they can take any orientation globally. Three coordinate systems are used in this work, two of them are alternately. The first system is the global system represented with the symbol G, while the other, denoted by S, is related to the smartphone coordinates. The relation between the two representations can be described mathematically as a linear transformation as follows:

$$p^G = T^{G/S} p^S \tag{5.2}$$

where p^S is a point located in smartphone coordinate S, p^G is the same point as it is described in global coordinate G and $T^{(G/S)}$ is the transformation matrix of point from smartphone coordinate S to global coordinate G. The third is the vehicle coordinate system which is the targeted coordinate for features extraction and can be estimated related to the two other coordinates after it is identified from the inertial sensors' interaction. The quaternion derivative ($\dot{q}$) with respect to time represents the angular velocity [181]. The smartphone gyroscope sensor

can therefore provide the angular velocity that can mathematically be given by

$$\dot{q} = \frac{1}{2}T(Gyro) \cdot \mathbf{q} = \frac{1}{2}\begin{bmatrix} 0 & -Gyro_x & -Gyro_y & -Gyro_z \\ Gyro_x & 0 & Gyro_z & -Gyro_y \\ Gyro_y & -Gyro_z & 0 & Gyro_x \\ Gyro_z & Gyro_y & -Gyro_x & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} \tag{5.3}$$

Rearranging 5.3 , the quaternion derivative can also be given by

$$\dot{q} = \frac{1}{2}\bar{T}(q) \cdot \mathbf{Gyro} = \frac{1}{2}\begin{bmatrix} -q_2 & -q_3 & -q_4 \\ q_1 & -q_4 & q_3 \\ q_4 & q_1 & -q_2 \\ -q_3 & q_2 & q_1 \end{bmatrix} \begin{bmatrix} Gyro_x \\ Gyro_y \\ Gyro_z \end{bmatrix} \tag{5.4}$$

where $\mathbf{q}$ is the quaternion vector while T, $\bar{T}$ are the transformation matrices between quaternion and Euler rotation representations and **Gyro** is the matrix of three-axis smartphone gyroscope readings. The angular velocity matrix in 5.3 can be obtained from the collected gyroscope data.

5.3 System Overview

A new framework is proposed in this experiment to detect the fatigue status of the driver based on integrated sensors inside the smartphone only without the need for any extra installed devices or electronic sensors inside the vehicle. The methodology of this work consists of several stages: data collection, pre-processing,

attitude estimation, feature extraction and classification. These stages are shown in the block diagram of Figure 5.2 and are described as follows.

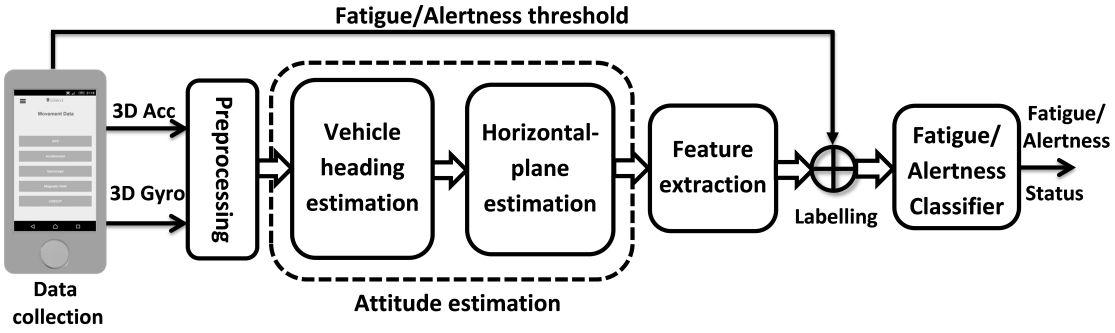


FIGURE 5.2: Block diagram of the work methodology stages

5.4 Data Collection and Pre-Processing

5.4.1 Data collection

In this experiment, 11 volunteer drivers aged 24 - 47 years participated.Each individual driver made two trips using a smartphone application inside the car to collect inertial data. All the volunteer drivers reported their fatigue level in a range of 1-6 out of 9. It is not very safe to ask the participants to drive beyond level 6.

A smartphone application is developed to manage the data collection for the proposed system. The architecture of the developed mobile application, shown in Figure 5.3(a), comprises several components, including:

a) Graphical user interface (GUI) – it is used to enter the user's login details, specify the sleepiness scale and initiate data transfer from the smartphone

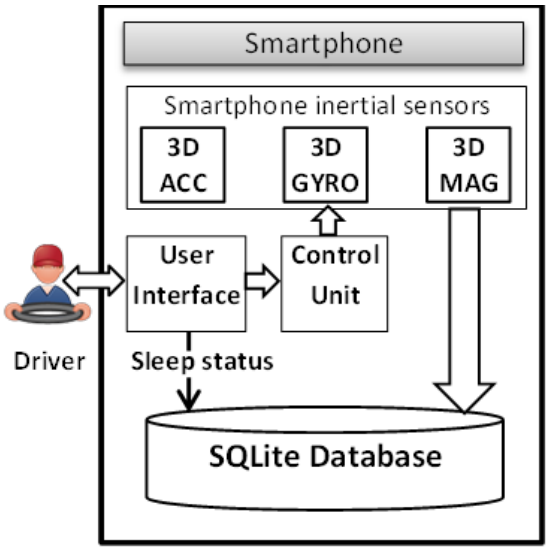
database to the processing device (i.e. PC, laptop, tablet, or a cloud server). An example screenshot for the developed GUI is shown in Figure 5.3(b) .

b) Control unit – this component controls the setting and the format of the collected data and configures the desired sampling rate of the data collection from the inertial sensors.

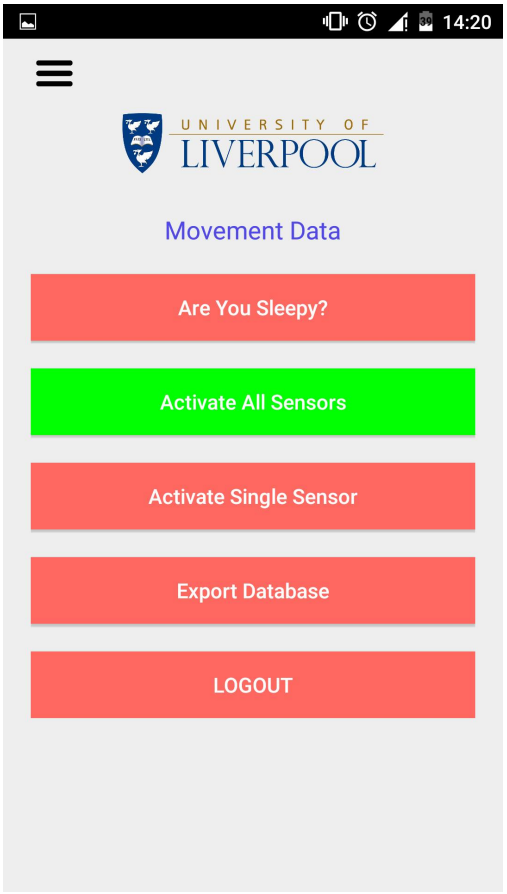
c) Inertial sensors – this part of the application deals directly with hardware smartphone sensors and manages to collect data in the setting of the control unit component.

d) Database – the collected data is stored in this component. It is build using the SQLite database and stores the raw data, which includes the timestamp and the three axes' data from the accelerometer and gyroscope. Moreover, it handles the format of the exported data.

The application is developed for an Android platform using Java programming language and SQLite as the database. Firstly, the recorded data is stored in the smartphone, then it is uploaded to the cloud when the Internet connection is available. There is an option with the application to store the dataset or delete it after uploading. The uploaded data is processed in a web server and used to detect the fatigue level of the driver. User interactivity with the developed system is performed through a GUI through which the user can configure various parameters such as sampling rate, sensor selection and storage time intervals.



(a) Application architecture



(b) Main-menu of the GUI

FIGURE 5.3: Example screenshot for the data-collection application interface

5.4.2 Pre-processing

Collected data from inertial sensors is usually contaminated with noise. Moving-average-window filter was used in this work to de-noise the data. A window of 20 samples was selected to remove random noise but still keep the monitored detailed movements of the vehicle. Missing slots of data were replaced either by averaging when these slots were less than 20 samples or by splitting the time series at the missing slot when ever one of the sensors reading is missing.

5.5 Attitude Estimation

Vehicle behaviour has a strong relation to driver alertness/fatigue status. Inertial sensors can capture the vehicle behaviour and monitor the change in driving style with driver fatigue evolution. The sensors frame was usually fixed to the vehicle frame and vector data is handled with previous knowledge of orientation. Since the data from smartphone-integrated sensors is collected without any prior knowledge of smartphone orientation relative to vehicle frame or heading, it is essential to estimate smartphone orientation relative to vehicle heading to extract effective measure of driver alertness/fatigue status.

5.5.1 Vehicle heading estimation

Fatigue-related features can be captured accurately from inertial/magnetic data if vehicle heading is estimated with acceptable accuracy. The main challenge in heading estimation approaches is the removal of gravity vector from the triaxial

accelerometer data. Several methods are used to purify the accelerometer data from gravitational components to obtain the proper acceleration [174, 182].

In this work, the acceleration and deceleration patterns of the vehicle are used as metrics for driver fatigue detection. Since the extracted features are related to the vehicle frame, all estimated coordinated are transferred to this frame. A new proposed method for heading estimation is adopted in the current work. The gravity is a stationary vector and it can be removed from the accelerometer data by dealing with time-changing data over a specific period of time. On the other hand, the standard deviation (SD) of acceleration data over a specific slot of time can capture the time varying data and ignore the stationary data. The idea of ignoring the stationary data, by calculating SD, is the key for the proposed method to find the vehicle heading. The right choice of the moving-window time, that gates the acceleration data, is an important factor in achieving reliable results because a trade-off is needed between removing the stationary gravity component and the targeted movement features.

The direction of SD of the triaxial acceleration data together with gyroscope data are used to estimate the vehicle heading. The acceleration SD has to pass two conditions to be valid as heading estimator. The first condition is to cross the acceleration threshold ($A_{cc}T$) to ensure that the vehicle is in motion. The second condition is used to guarantee the linear path of the vehicle with minimum curvature. The gyroscope data is used to fulfil the second condition when the magnitude of data decreases below the proper threshold ($GyroT$) as described in the following equation:

$$\overrightarrow{SD(A_{cc})}\,|_{Gyro<GyroT,\ A_{cc}>A_{cc}T} \triangleq \overrightarrow{H} = \text{vehicle heading} \tag{5.5}$$

where $GyroT$, a threshold for gyroscope data, is used to detect the non-rotational status of the vehicle. After that, the SD vector of acceleration data can point to the heading of the car since the rotational acceleration can be assumed to be zero. The vehicle normalised vector is considered as a pure quaternion as described in Equation 5.6 which is used as a reference to the smartphone quaternion.

$$\boldsymbol{q_v} = 0 + 0i + 1j + 0k \tag{5.6}$$

where $\boldsymbol{q_v}$ is the vehicle normalised quaternion aligned with the vehicle heading (vehicle y-axis).

The moving average window is applied over the accelerometer data to extract the gravity vector [183]. The gravity component of acceleration is calculated since it is the most stationary part of the acceleration data. The gravity vector $\boldsymbol{g}$ is removed from the triaxial accelerometer data to generate the linear acceleration vector $\boldsymbol{LA_{cc}}$. Equations 5.7 and 5.8 describe the mathematical representation of this process.

$$\boldsymbol{g} = W_A(\boldsymbol{A_{cc}}) \tag{5.7}$$

$$\boldsymbol{LA_{cc}} = \boldsymbol{A_{cc}} - \boldsymbol{g} \tag{5.8}$$

where W_A is the moving average window function.

5.5.2 Horizontal plane estimation

Vehicle attitude estimation methods typically estimate the Earth's horizontal plane but not the vehicle's horizontal plane [182], [85]. The difference the two horizontal planes comes from road topography and what is called super elevation [184]. To the best of the author's knowledge, obtaining the vehicle horizontal plane through a smartphone's inertial/magnetic sensors has not been reported yet in the literature. A thorough search of vehicle attitude estimation methods, based on smartphone inertial/magnetic sensors, has not found any method that deals with the vehicle horizontal plane. These methods are presented to estimate the Earth's (not the vehicle's) horizontal plane [182], [85].

Vehicle horizontal plane can be determined based on two perpendicular vectors, the heading and the lateral vectors. The heading vector has been calculated as mentioned earlier in subsection 4.4.1. The lateral vector is calculated using the proposed method. Smartphone gyroscope data includes three perpendicular rotational axes' data, and these data can capture the turns and manoeuvring of the vehicle. The gyroscope vector can be considered as the normal vector of the vehicle's horizontal plane in the event of the car turning. For the same event, the SD vector of accelerometer data can be considered as the lateral vector. Equations 5.9 and 5.10 show the conditions required to consider the accelerometer and gyroscope data as vehicle lateral vector and normal vector of the vehicle's horizontal plane, respectively.

$$\overrightarrow{SD(A_{cc})}\,|_{Gyro>GyroT,\ A_{cc}>A_{ccT}} \triangleq \overrightarrow{L} = \text{vehicle lateral vector} \tag{5.9}$$

$$\overrightarrow{Gyro}\,|_{Gyro>GyroT,\ A_{cc}>A_{cc}T} \triangleq \overrightarrow{N_H} = \text{normal of vehicle horizontal plane} \tag{5.10}$$

The cross product of the heading and the lateral vectors produces the normal vector which is perpendicular to both of them. This normal vector is the same as the normal vector calculated from gyroscope data. From a practical consideration, there is a deviation error ($\overrightarrow{ERR_N}$) between the two vectors. The lateral and horizontal plane values are then considered to be accurate when the deviation error is below acceptable limits as shown in EQ 5.11.

$$\overrightarrow{H} \times \overrightarrow{L} = \overrightarrow{N_H} + \overrightarrow{ERR_N} \tag{5.11}$$

5.6 Feature Extraction

Many research approaches have been developed to monitor driver behaviour [3], [81] and [185]. The most common symptoms used to detect fatigued drivers are shown as follows:

a) Longitudinal acceleration/deceleration patterns

A drowsy driver may face difficulty in maintaining road speed levels and typically make sudden and sharp changes in acceleration and abrupt braking. The longitudinal component of the accelerometer which is aligned with the vehicle heading can provide the driver's pattern of acceleration and deceleration.

b) Swerve, lane departure and weaving

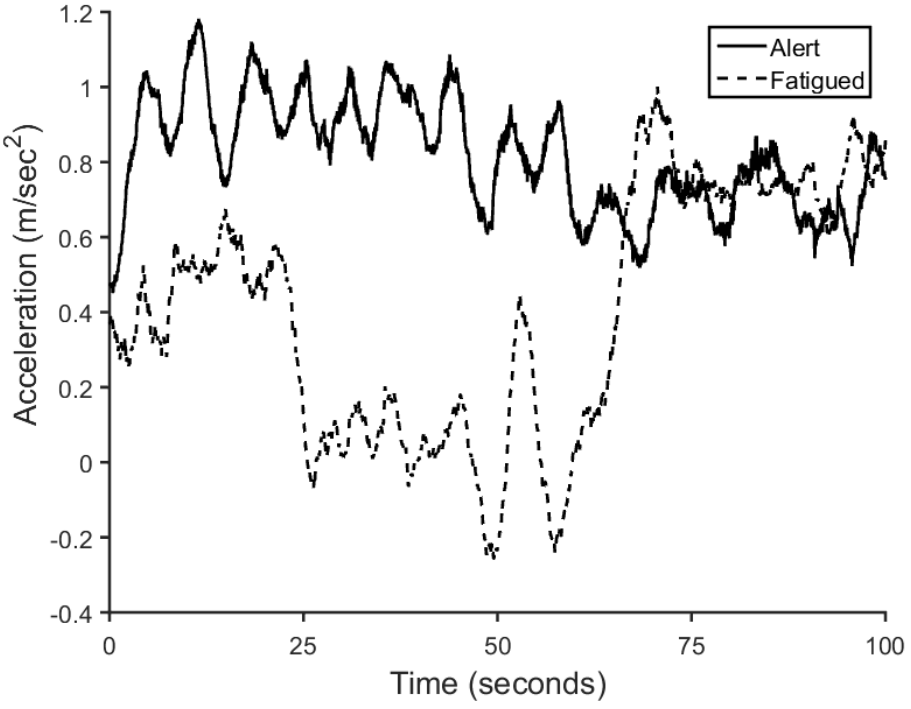
Swerve, lane departure and weaving of a vehicle have been strongly correlated to the driver alertness level [186]. The frequency and severity of these symptoms increase remarkably with the level of fatigue. The lateral movement of the vehicle is the best measure for the above symptoms and can be quantified based on accelerometer data.

c) Steering wheel movement

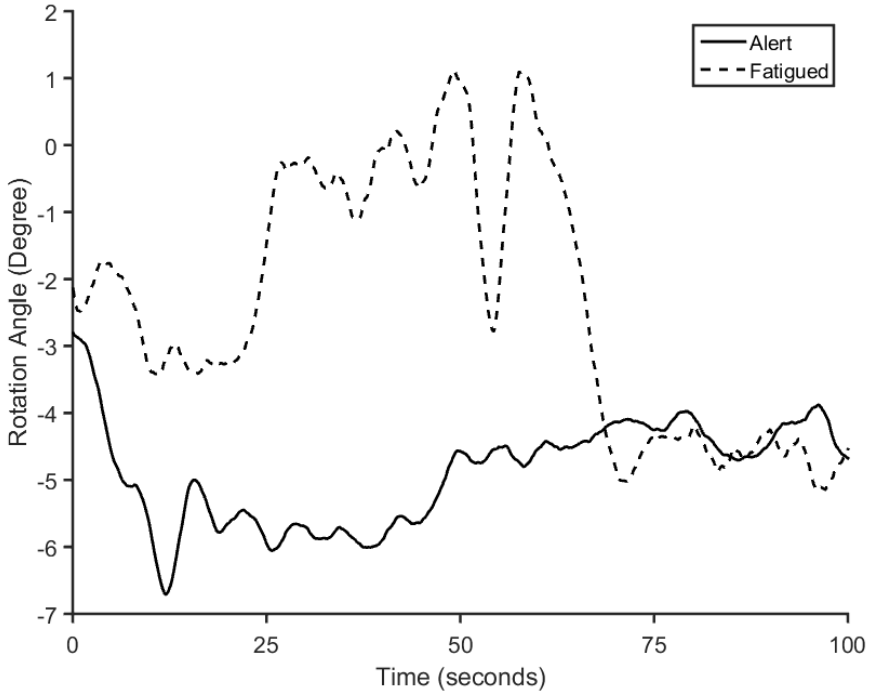
Fatigued drivers driving patterns differ from those of alert drivers. For example, micro corrections become less frequent as the driver's alertness decreases (i.e. fatigue increases). In addition, the larger steering angles become clearer with fatigue evolution [187]. Steering wheel movement (SWM) can be measured based on data collected from gyroscope, accelerometer or both.

Two sensors are chosen in the first example to show the difference in motion data between an alert anda fatigued driver. Accelerometer data, which is shown in Figure 5.4(a), illustrates many small waves drawn in a sold line for the alert driver with less frequent changes but of larger magnitude for the fatigued driver marked with a dashed line. Figure 5.4(b) shows the same difference between an alert and a fatigued driver's data, which was produced by the system rotation.

Combinations of the generated features are visualised to pre-check the best collection of features that can enhance the fatigue detection classification performance. Figure 5.5 shows two examples of statistical features derived from pre-processed data. The two graphs help the observer to distinguish between driver fatigue levels depending on gyroscope standard deviation value. The fatigued driver has lower gyroscope standard deviation value in general but has higher rotation and acceleration standard deviation values. Data visualisation, shown in



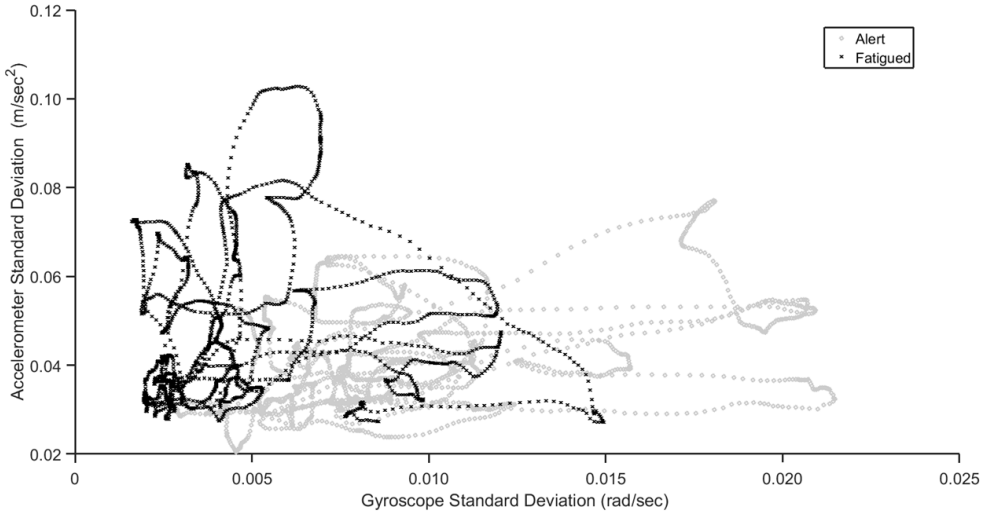
(a) Acceleration data



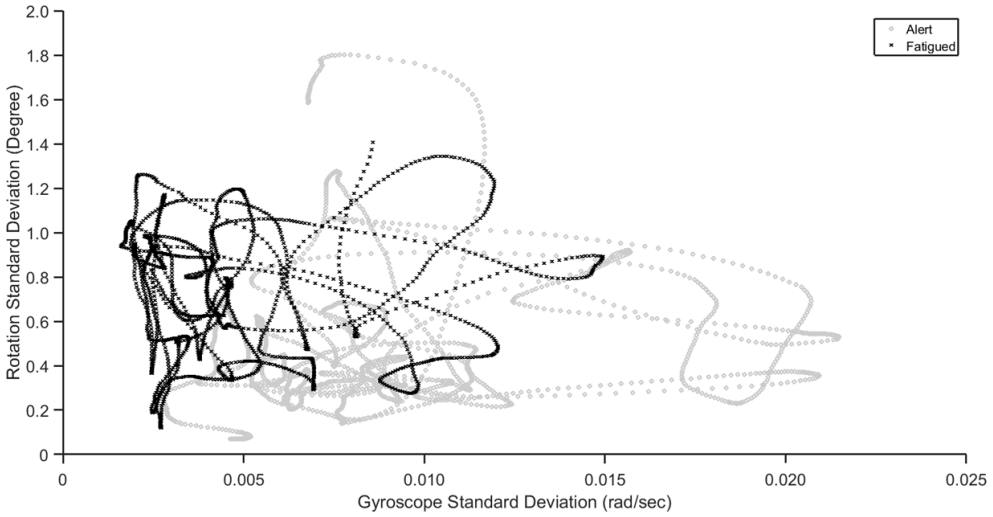
(b) Rotation data

FIGURE 5.4: Acceleration (a) and rotation (b) data for a driver during periods of alert and slightly fatigued driving

Figures 5.4 and 5.5, shows the differences between the alert and fatigued drivers. The earlier mentioned fatigue related symptoms (SWM, swerve etc.) are used as a core of the feature selection. For longitudinal acceleration/deceleration patterns, as an example, two features are selected to represent this symptom, and they are the longitudinal acceleration and its standard deviation.



(a) Accelerometer-Gyroscope data



(b) Rotation-Gyroscope data

FIGURE 5.5: Relation between standard deviation of (a) acceleration and gyroscope (b) rotation and gyroscope data for a driver during periods (100 seconds) of alert and slightly fatigued driving

A total of six features are used to detect the fatigue status of the driver as shown in Table 5.1. For comparison purposes (results shown in Section 5.8.2), the equivalent features that are extracted from data without applying the proposed attitude estimation approach, are listed in Table 5.1.

TABLE 5.1: Selected feature set with and without taking the proposed method

Behaviour	Features before using the proposed method	Features after using the proposed method
Longitudinal acceleration/ deceleration patterns	a) Y-axis acceleration. b)Y-axis acceleration SD.	a) Longitudinal acceleration. b) Longitudinal acceleration SD.
Swerve, lane departure and weaving	c) X-axis acceleration. d)X-axis acceleration SD.	c) Lateral acceleration. d) Lateral acceleration SD.
Steering wheel movement (SWM) rate	e) Rotational velocity. f) Rotational velocity SD.	e) Rotational velocity (aligned with vehicle horizontal plane). f) Rotational velocity SD (aligned with vehicle horizontal plane).

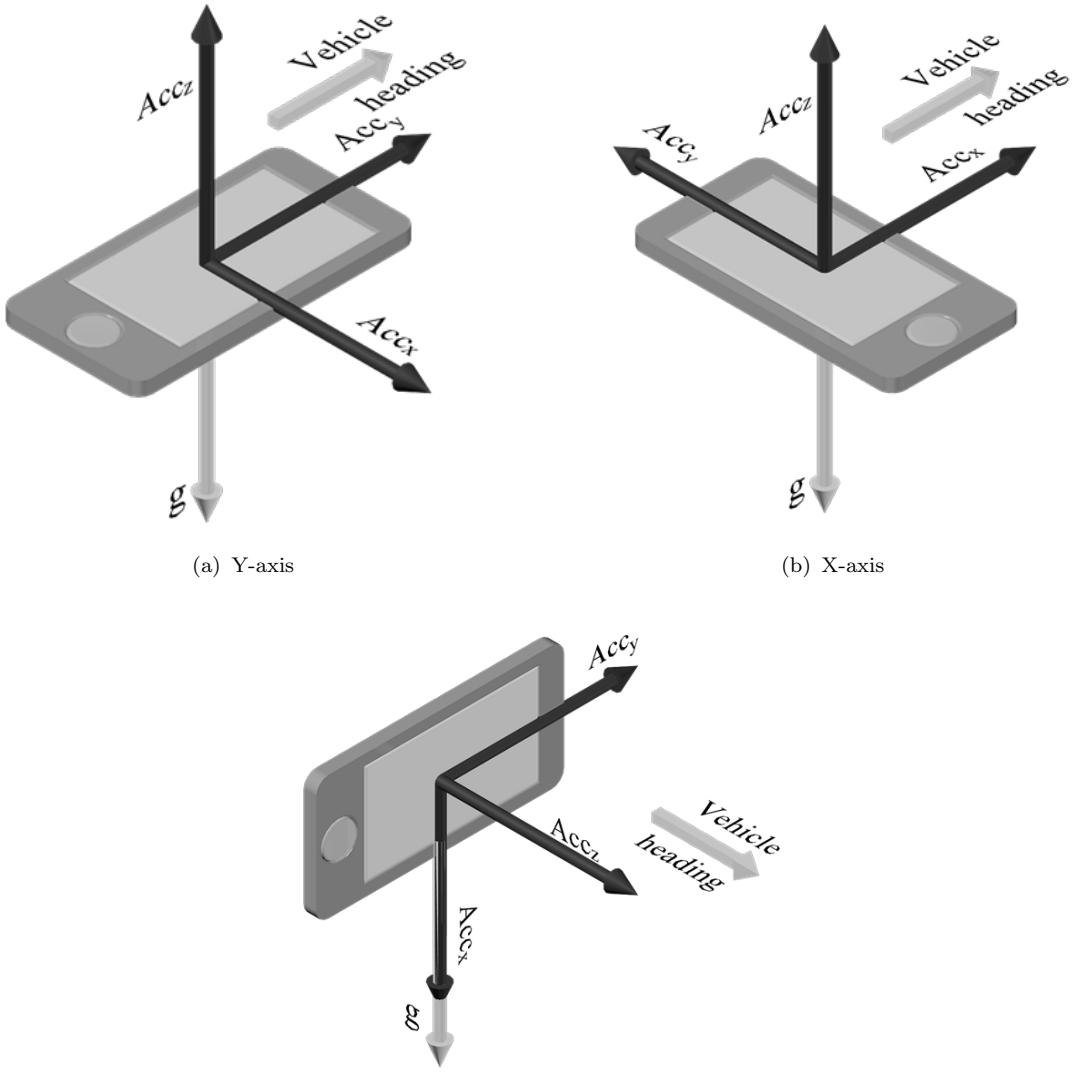
5.7 Classification

The collected dataset is labelled depending on fatigue level score (KSS). The records with scores of 1-3 are labelled as alert while the other records with scores of 4-6 are labelled as fatigued. A two-class neural network classifier was implemented to detect driver fatigue status. The developed ANN classifier consists of an input layer with six inputs, two hidden layers with 12 neurons each and one output layer. Eleven-fold cross-validation method is implemented to calculate the average performance of the classifier.

5.8 Results and Discussion

5.8.1 Attitude estimation

An experiment is conducted to evaluate the accuracy of the heading estimation algorithm with the aid of three drivers. Each driver is asked to conduct three trips, each lasts for 10-20 minutes. In each of these trips, the smartphone is aligned with one of the vehicle's axes (x y z), as shown Figure 5.6.



(a) Y-axis

(b) X-axis

(c) Z-axis

FIGURE 5.6: Smartphone heading alignment

Three sets of data are used to check the performance of the proposed heading estimation. Each set relates to one of the three Cartesian coordinates. The performance of this algorithm is calculated based on the deviation of the estimated coordinate from the positioned one. The projection of the misaligned coordinate over the positioned one, which is the effective component of the feature $Proj$, is demonstrated as an example for x direction as follows.

$$Proj_x = \frac{Acc_x}{\sqrt{Acc_x^2+Acc_y^2+Acc_z^2}} = cos\phi \tag{5.12}$$

where ϕ is the angle between the two heading vectors, the measured and correctly positioned heading, and the percentage error is calculated as follows.

$$Error\% = (1 - Proj_x) \times 100\% \tag{5.13}$$

An example of attitude estimation of the proposed method is illustrated with data shown in Figure 5.7. The first event is the heading estimation, which is calculated once two conditions have occurred. These conditions are the accelerometer data passing over the $Acct$ threshold and the gyroscope data going below the $GyroT$ threshold. The two conditions ensure that the vehicle is accelerating or decelerating in a straight line with minimum curvature action. This event, which is bounded with two vertical dotted lines, is shown in Figure 5.7 with a slot of time before and around the 50 seconds time tick.

The second event is about the horizontal plane and lateral vector estimations. A time slot, which is bounded with the dotted verticals lines, borders the horizontal estimation event. Again, two conditions should be satisfied before the calculation

starts. Both the accelerometer and the gyroscope data have to cross over the *Acct* and *GyroT* limits. These conditions are applied to check the vehicle turning status with proper angular acceleration

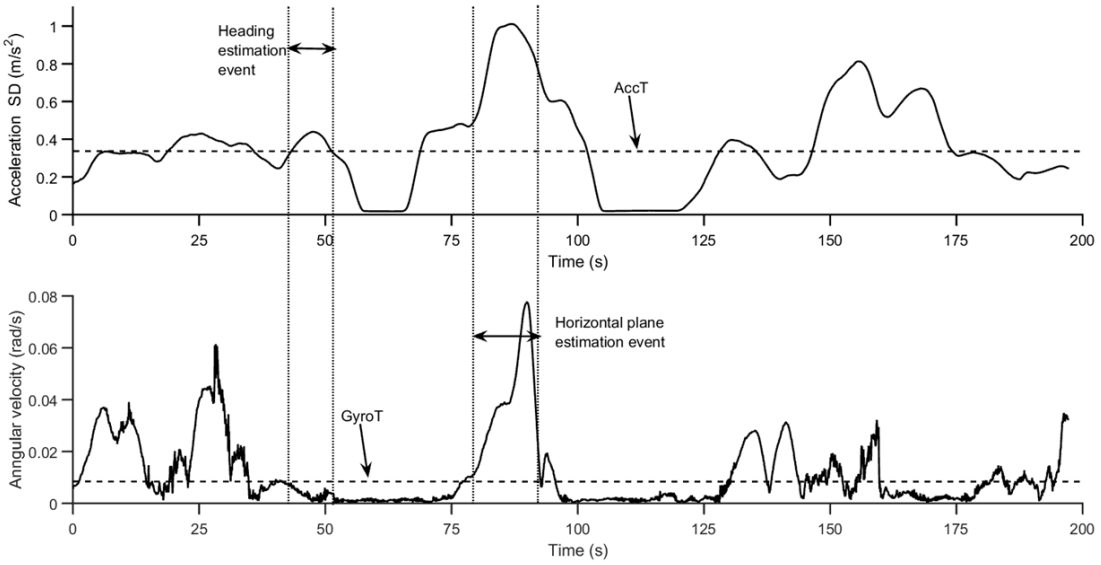


FIGURE 5.7: Attitude estimation illustration

The errors, shown in Table 2, are expected and generated by imperfect smartphone position inside the vehicle sensor accuracy, noise and vibration, and algorithm inaccuracy. For the fatigue detection system, the accuracy of this algorithm is acceptable because the fatigue scoring level is limited to a resolution of 1/9 (KSS scale includes 9 scales).

TABLE 5.2: Percentage errors of the proposed method

Error Type	X axis	Y axis	Z axis
Heading Error %	2.36	5.46	2.80
Lateral Error %	7.34	4.56	8.23

5.8.2 Classification performance

The results listed in Table 3 show that the success of the system in detecting the fatigue status of the driver depends only on the inertial sensors' data (accelerometer and gyroscope) without using magnetometer data. The proposed orientation correction algorithm enhanced the detection accuracy from 81.2% to 87.3%. This accuracy comes from picking the right fatigue-related features even when the smartphone is not aligned with the vehicle frame and not oriented with the driving heading. This algorithm overcomes the disadvantages of using a magnetometer and gravity sensor in calculating the heading and horizontal plane, like the disturbances come from magnetic fields inside the vehicle and the tilting and sloping of the vehicle on the road.

Table 5.3: Classification accuracy improvement with proposed method

Output status	Before alignment	After alignment
Accuracy%	81.2	87.3
Sensitivity%	79.3	84.4
Specificity%	82.8	90.1

5.9 Conclusions

A new approach for automatic detection of driver fatigue using smartphone inertial sensors has been presented and discussed in this chapter. To the best of the author's knowledge this approach has not yet been thoroughly investigated in a realistic environment. Unlike existing approaches that are mainly based on physiological data, driver behaviour and others, the proposed approach utilises

non-intrusive data collected from the built-in inertial sensors of a smartphone. Moreover, this approach can estimate the smartphone orientation relative to the vehicle frame orientation no matter how it is oriented inside the vehicle. In addition, the proposed method distinguishes between the vehicle and the Earths horizontal planes, which improves the algorithm's accuracy. This algorithm uses only inertial sensors data without the need for magnetic sensor data which causes a large part of the orientation estimation errors especially with magnetic disturbances inside vehicles. No further instruments are therefore needed to detect the fatigue/alertness status other than the existing smartphones. In addition, the proposed approach is based on collecting data from inertial sensors, which are not relevant to the drivers personal data, Thus this approach preserves the user's privacy.

The results show a good improvement of classifier performance after using the proposed algorithm. Real-life driving data and the non-intrusive collection method help to obtain reliable and realistic results. Usinga smartphone, the ubiquitous tool, in this approach, improves road safety by increasing the availability of a fatigue assessment solution for drivers.

Limitations of the experiments carried out in this chapter can be outlined as follows; (i) the top three levels (7 - 9) of the self-reported fatigue assessment scale are deemed too dangerous to carry out in naturalistic driving, thus no data were reported on these levels and (ii) the rules of the work regulations and labour union limited the number of available participants.

Chapter 6

Modular Structure: Design, Implementation and Test

6.1 Introduction

The need for a robust, reliable and flexible operator fatigue detection system is an essential requirement for work safety which can be achieved using a modular structure. Another reason for using a modular structure is the independence of each individual module from other modules. This also allows heterogeneous modules to be gathered in a single system without the need to agree on the design details of individual modules. Different types of fatigue-related modules can be combined together such as physiological-based module (explained in Chapter 3) or behavioural-based modules (explained in Chapter 4 and Chapter 5).

In this chapter, a new scalable modular design approach is proposed, implemented and tested successfully using a Bayesian combiner and particle swarm optimiser

(PSO). A new modification is proposed to the Bayesian combiner to enable the particle swarm optimiser algorithm to work with the combiner. This enables utilisation of input-modules depending on their availability and thus improves flexibility and robustness against losing one or more data sources. This chapter describes the modular structure of the proposed fatigue detection system which comprises three main stages: fatigue detection modules, a combiner and an optimiser. The obtained test results are also presented and discussed in this chapter along with concluding remarks.

6.2 System Overview

System structure is based on the modular design approach. As shown in the block diagram of Figure 6.1, it comprises mainly three stages. The first is the classification and labelling stage which consists of four classifications modules with their labelling component. The second stage uses the output of the classifiers to combine them based on a Bayesian combiner. Finally, the last stage is the particle swarm optimiser which is used to enhance the accuracy of the fatigue detection system by assigning an optimal set of weight combinations for the modules.

A block diagram of the proposed system is shown in Figure 6.1. As illustrated, the system incorporates three stages. The first stage includes four modules. These modules are designed and implemented using an existing dataset offered by SHRP2NDS [188]. The second stage uses the Bayesian combiner to fuse outputs of the first stage modules. The last stage is added to enhance system performance by assigning different weights for modules using the PSO. The modular design of

the proposed system is adopted to gain the robustness, flexibility and practical implementation advantages.

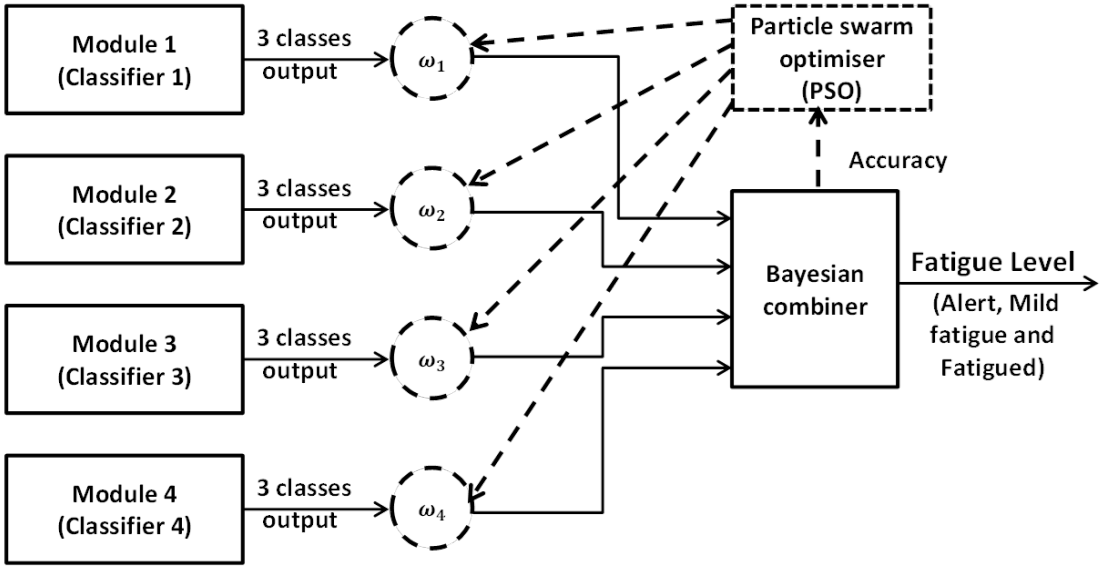


FIGURE 6.1: Modular architecture of the proposed system

6.3 Theoretical Background

This section presents a theoretical background for two key aspects of the proposed work, importance of modularity and particle swarm optimiser (PSO).

6.3.1 Importance of modularity

The need for a system that can combine heterogeneous subsystems (modules) efficiently is behind the idea of *"interdependence within and independence across modules"* [189], which is one of the definitions of modularity. The modular design is preferred over the integral design due to its benefits including [190, 191]:

a) Reliability and robustness: The modular structure enhances the robustness and improves the reliability of certain systems if they are designed properly to maintain their functional mission even if they lose one or more of their modules.

b) Flexibility: The need for a flexible system is an essential practical requirement [192]. For a fatigue detection system where a variety of detection methods are available, it is very important to design the system to be flexible from practically and applicability views.

c) Comprehensibility: In complex systems such as fatigue detection systems, a modular structure makes them more understandable and easier to handle on a module level.

d) Independence: The main motivation of using a modular structure is the independence of each individual module from other modules. This feature is very helpful for modular systems to gather heterogeneous modules together with no need for them to agree in any interior details and in the abstract output protocol.

e) Abstraction level: This level is responsible for the interfacing between modules. The abstraction and independence of the modules help to make the system more practical and flexible.

The modular structure system needs a frame to be plugged in to produce the combined final output. Several combination algorithms have been proposed to combine heterogeneous sets of modules (e.g. majority voting, weighted majority voting or Bayesian combiner) [138], [139]. The Bayesian combiner is ideally suited for problems when the output of the modules is independent even when the number of modules is dropped to two [140].

Only a few research papers have identified the use of optimisation techniques to enhance the performance level of the modular organisation [193].

6.3.2 Particle swarm optimisation

Several optimisation algorithms have been used in the operator fatigue detection area to improve classification performance, such as genetic algorithm [194], swarm optimisation [195] and others. Different swarm optimisation (e.g. ant colony optimisation, bacterial foraging optimisation and particle swarm optimisation (PSO) have been proposed in the transportation sector to find the optimal traffic network situation [196]. However, few studies have used optimisation algorithms for operator fatigue and mostly they have adopted PSO as an optimiser [197]. PSO is used in this work to improve the performance of the Bayesian combiner.

Particle swarm optimisation is a meta-heuristic algorithm devised by Kennedy and Eberhart in 1995. PSO is inspired by the social behaviour of birds. Several versions of PSO were derived later to cover new applications or to address some limits and challenges discovered with the original version [198].

PSO can find the optimal solution for an optimisation problem in a D-dimensional hyperspace. A swarm of N particles is recruited to find the best position according to individual perspective (*Pbest*) and the overall perspective (*Gbest*) [199]. Each particle tries to update its position (solution) to achieve best fitness value (minimise cost function). The update stochastic function is based on three parts; inertia part, self-knowledge part and team-work part. The update rule is determined as follows:

$$v_i^{k+1} = wv_i^k + c_1 R_1 \times \left(Pbest_i^k - x_i^k\right) + c_2 R_2 \times \left(Gbest_i^k - x_i^k\right) \tag{6.1}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{6.2}$$

where w is the inertia (habitual behaviour) weight, c_1 is an acceleration constant of the self-knowledge (memory) component , c_2 is an acceleration constant of the team-work component, and R_1 and R_2 are random numbers.

In this work, the PSO is used for its appropriate characteristics including: speed of implementation, low computational cost and no need for derivative function [200]. These characteristics meet the requirements for developing an online system based on handheld devices especially the computational efficiency, which improves the devices' battery consumption.

6.4 Module Design

In the proposed modular structure, the modules could be any fatigue-related decision maker such as classifier and/or operator assessment. In this chapter, an existing dataset [188] is adopted to build the modules. Similarly, fatigue detection methods which were previously discussed in chapters 3, 4 and 5 can form separate modules in this structure.

The adopted data is collected for several trips, each of which last for 90 minutes. This period can be considered adequate for fatigue development over time. The most common four metrics relevant to fatigue detection are considered in this study: 3-axes accelerometer, gyroscope, head position and head rotation. The

first two metrics are collected with a sampling rate of 10 Hz, while the other two are dynamically calculated relative to a baseline with a sampling rate of 2-15 Hz. These metrics are pre-processed from missing slots and noises by averaging using a moving average with a 15 sample window. The methodology of this study compromises three main stages: classification and labelling, Bayesian combiner and particle swarm optimiser. These stages are described as follows.

Four fatigue-related modules are built for this work to demonstrate the modularity structure as shown in Figure 6.1. The modules are: (i) accelerometer data module, (ii) gyroscope data module, (iii) head position data module and (iv) head rotation data module. The driver and the driven vehicle data are fed into four ternary neural networks (ANN) classifiers, each of which generates an alert, mild fatigue or fatigued status as shown in Figure 6.2. Ground truth of these classifiers is built upon the fact that the blink rate increases with fatigue growth [33], [201] and [202]. In this work, the blink-rate is counted manually for randomly selected times in minutes over a certain trip period and is used for classification labelling.

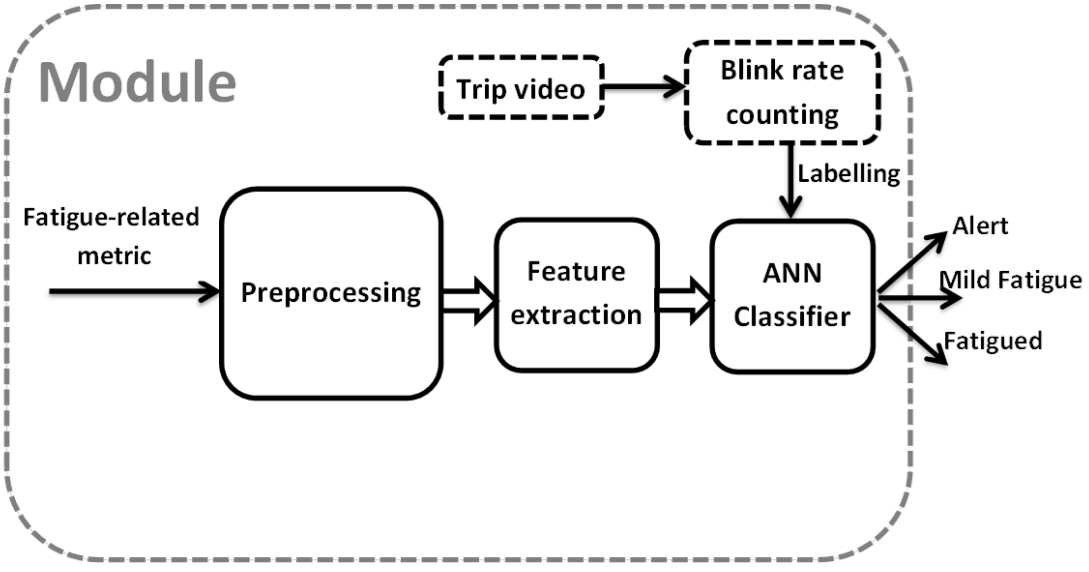


FIGURE 6.2: An example of the module

Three classes are used to label the three driver statuses. Six features are extracted and fed to each classifier: mean and standard deviation of each axis of the three-dimensional values of the fatigue-related metrics (i.e., acceleration, gyroscope, head-position and head-rotation). Different numbers of hidden layers and numbers of neurons in these layers are used to construct the four ANN classifiers. This variation in ANN architecture is purposely adopted in addition to the different seeds of random generation of ANN weight initialisation in order to maintain the independence of the module's decision.

6.5 Bayesian Combiner

Outputs of the ANN classifiers are fed to the Bayesian combiner to improve system performance irrespective of the number of input modules. Bayes' rule can be further refined using Eqs. 3.13 and 3.14, which produces the following.

$$= \frac{P\left(h_j\right) \prod_{i=1}^{L} P\left(b_i \mid h_j\right)}{P\left(\boldsymbol{b}\right)}, \qquad j = 1, ..., 3 \tag{6.3}$$

In this work, three classes' modules (classifiers) are implemented (i.e., $c = 3$) and different combination sizes are experimented with (i.e., $L = 4$, $L = 3$ and $L = 2$). The confusion matrix of each classifier is assumed to be an approximation of posterior probability. The Bayesian combiner can predict the combined output based on Eq. 6.3 with extra modification. Since the combined output depends on the outputs of all the modules, the probabilities of the modules' outputs can be ignored and the probability of prediction $\beta(\boldsymbol{x})$ for a certain input to be classified as a specific class will be proportion to:

$$\beta_j(\boldsymbol{x}) \propto P\left(h_j\right) \prod_{i=1}^{L} P\left(b_i \mid h_j\right) \tag{6.4}$$

The confusion matrix CM^i of each classifier is calculated using the training set and a matrix of 3×3 (based on number of classes) is produced. The prior probability of the class h_j can be considered as an estimate of N_j/N where N_j is the number of outputs classified as the class h_j and N is the total number of outputs, while cm^i_{j,b_i} is an estimation of the probability $P\left(b_i \mid h_j\right)$. This will result in the following equation:

$$\beta_j(\boldsymbol{x}) \propto \frac{N_j}{N} \prod_{i=1}^{4} \frac{cm^i_{j,b_i}}{N_j} \tag{6.5}$$

Finally,the class that gains maximum probability, will be chosen as the predicted label. Moreover, to satisfy the required condition for Bayes' theory [139], the generated outputs of the classifiers are considered independent. Such an assumption is maintained through utilisation of different sets of features and different seeds for the ANN's initial weights. Since the Bayesian combiner that deals with abstracted output labels does not depend on the classification process, the number of input modules can vary depending on the available sources of driver and vehicle data, thus improving flexibility and modularity attributes of the system.

6.6 Particle Swarm Optimiser

The fatigue-related metrics differ in their responses to fatigue, thus each module has a different weight in fatigue identification. Although the combiner usually enhances the performance of the ensemble classifiers, a further step is needed to weight each classifier with its share of contribution in the combiner's performance. A particle swarm optimiser (PSO) is adopted in this work to find the sets of optimal weights $\boldsymbol{\omega}$ of each metrics/module. The PSO is used in this work to improve the performance of the Bayesian combiner. The cost function of the optimiser is a function of the combiner accuracy, while the best solution will be represented in the best set of weights related to each module (it is not related to the inertia weight of the optimiser update rule in Eq. 6.1).

A new modification is proposed in this work to weight the classifiers' output by adding some changes to the combiner method. This rule is implemented by strengthening the true output class and weakening the false one for the module which is expected to have the best response to fatigue, as follows.

$$\widehat{cm^i_{j,b_i}} = \begin{cases} \omega_i \times cm^i_{j,b_i} & j = b_i \\ \frac{1}{\omega_i} \times cm^i_{j,b_i} & j \neq b_i \end{cases} \tag{6.6}$$

where $\widehat{cm^i_{j,b_i}}$ is the weighted confusion matrix element.

The optimiser keeps updating the weights until the cost function ε is minimised below a certain value or the iteration of the PSO algorithm exceeds a set value. The cost function is related to the accuracy metric of the combiner, as follows.

$$\varepsilon(\boldsymbol{\omega}) = 1 - accuracy(\boldsymbol{\omega}) \tag{6.7}$$

6.7 Results and Discussion

The findings demonstrated the ability to detect fatigue status from the dataset used in this study along with the trend of fatigue growth over time. The findings are detailed in the following subsection.

6.7.1 Module results

Each module predicts one of the three driver fatigue using a trained classifier that is labelled based on blink rate level. Blink rate is calculated manually from the available sample videos by dividing the video into time slots of 15 minutes, randomly selecting five minutes from each 15-minute slot, and then they are averaged to come up with one number to represent the blink rate of the 15-minute slot.

Three levels of blink rate are created to label three classes for the classifiers as shown in Figure 6.3. The trend of three trips is very clear to be growing with time of trip and in turn with fatigue. It is worth pointing out that the three drivers have different trip durations and they and their blink rate growth respond differently according to their fatigue level. The selection of the three levels with the choosing of boundaries needs more investigation and this will be achievable when more datasets and videos are available.

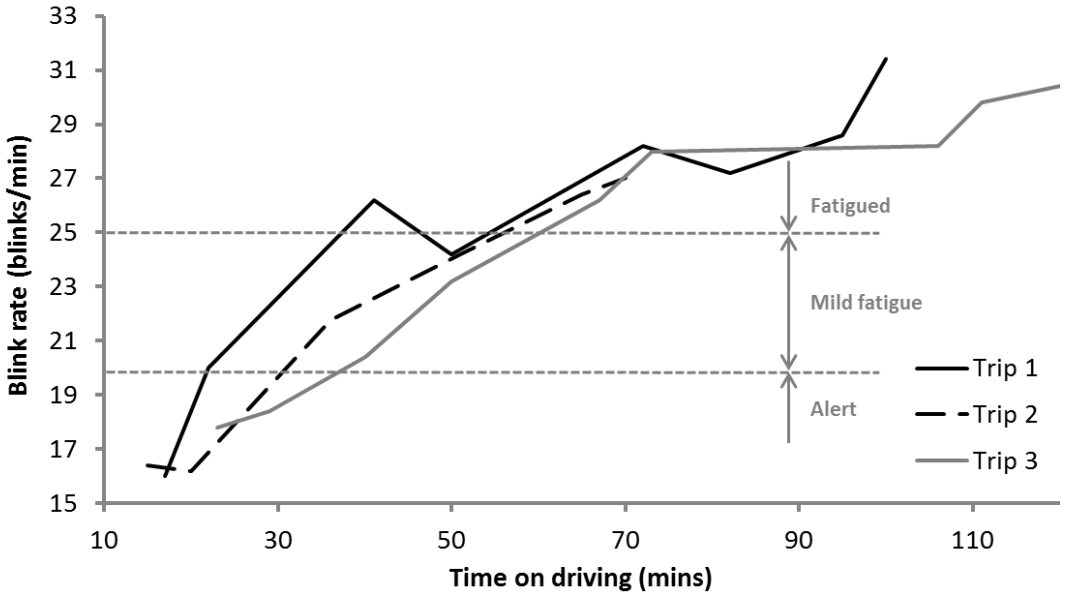


FIGURE 6.3: Blink rate changes with driving time

The four feed-forward ANN classifiers (modules) are build with different architecture (number of hidden layers and number of neurons in each hidden layer) to sustain the independent decision which is a requirement for the Bayesian combiner. Each module has six inputs to fit with the six features (i.e., mean and standard deviation of the 3-axes' sensor data) and one ternary output for the three classes. A back-propagation algorithm is used to train the modules with 70% of the dataset while the rest (30%) is used to test the system and calculate the performance of the classifiers.

Performance of the ANN classifiers was assessed in terms of accuracy, sensitivity and specificity. These metrics that are calculated for each class of the ternary classifiers are shown in Table 6.1. The overall metrics are then computed by averaging the individual metric over the three classes. It can be noticed that the performance of the visual data (i.e., head position and rotation) is superior to that of the inertia data (i.e., accelerometer and gyroscope)

Table 6.1: Performance of individual classification modules

Performance Metrics	ANN classifiers			
	Acceleration	Gyroscope	Head-Position	Head-Rotation
Accuracy (%)	74.8	70.6	81.9	79.6
Sensitivity (%)	72.5	77.1	82.9	79.1
Specificity (%)	74.7	79.0	82.1	79.6

6.7.2 Combiner results

The Bayesian combiner utilises the output labels from the classifiers stage to generate the final ternary decision. Possible modules (classifiers) combinations were experimented with to validate the modularity. Three combination sets were tested, starting with the presence of all four modules as one combination, followed by three and two modules per set, as shown in Table 6.2, in which performance metrics of the Bayesian combiner for these combinations are summarised. It should be noted here that the values given in this table for the performance metrics are averaged to obtain one value per metric.

Table 6.1 demonstrated validity and effectiveness of the proposed modular approach in capturing changes of fatigue-related features taking into account the growth of fatigue due to time spent driving. The visual metrics used in this study show a better performance than the inertial sensors' metrics. It is also demonstrated in Table 6.2 that utilisation of the Bayesian combiner has a significant positive impact on the fatigue detection performance as compared to that of the individual classifiers. As expected, the best performance (accuracy: 90.4%, sensitivity: 92.6% and specificity: 90.7%) was obtained with utilisation of the four input modules and the poorest performance (accuracy: 71.1%, sensitivity:

82.9% and specificity: 77.5%) was obtained with two modules. It is also observed that the highest performance (accuracy: 80.5%, sensitivity: 85.1% and specificity: 80.9 %) achieved with two-module combinations was with the existence of visual sensors (i.e., modules 3 and 4).

TABLE 6.2: Performance of combined classification modules

No. of Modules	Combinations of Classifier Modules	Bayesian Combiner performance		
		Accuracy (%)	Sensitivity (%)	Specificity (%)
4	1, 2, 3, 4	90.4	92.6	90.7
3	1, 2, 3	88.5	91.5	88.9
	1, 2, 4	85.7	88.8	86.2
	1, 3, 4	85.9	88.9	86.9
	2, 3, 4	86.3	90.0	88.0
2	1, 2	77.4	83.6	77.9
	1, 3	76.3	83.1	76.7
	1, 4	76.3	82.7	76.7
	2, 3	77.4	83.4	77.8
	2, 4	77.1	82.9	77.5
	3, 4	80.5	85.1	80.9

6.7.3 PSO results

Figure 6.4 shows a comparative view for accuracy metric changes of the Bayesian combiner before and after weight optimisation. The accuracy is improved slightly after selection of the optimal weights. The first cluster of bars represents the four-module group, which consists of one combination of modules, and this is the reason that it has no standard deviation bar. The second cluster shows

the three-module group which consists of four combinations with their standard deviation bars. Finally, the last cluster shows the two-module group, which includes six combinations. Figure 6.4 also illustrates the slight improvement in accuracy after selection of the optimal weights. It is worth mentioning that the highest improvement from optimisation is gained by the two-module group.

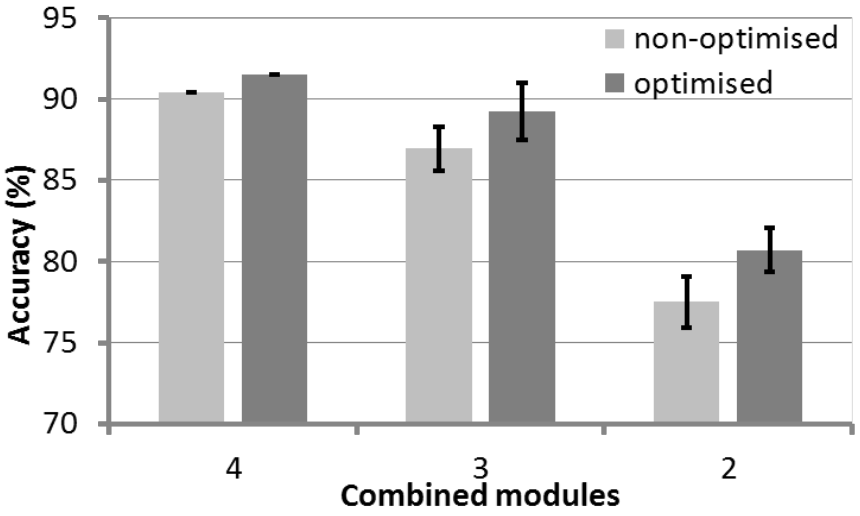


FIGURE 6.4: Bayesian combiner accuracy changes before and after optimisation

6.8 Conclusions

A prototype of the proposed fatigue detection system has been designed, implemented and tested successfully using combinations of four different fatigue detection modules. The obtained results are found to be in agreement with the previously reported findings but with significant system design flexibility and robustness against losing one or more data sources. Performance assessment of the developed prototype showed that accuracy, sensitivity and specificity of the fatigue detection are only slightly affected by the partial loss of input data.

To the best of the author's knowledge, a modular structure of the multi-source fatigue detection approach has not been thoroughly explored in the literature yet. This design approach allows for integrating different heterogeneous modules in a single system, thus obtaining a more efficient ad robust fatigue detection and assessment system.

The findings reported in this chapter were limited by: (i) unavailability of adequate funding to obtain a larger-size dataset which can improve the overall system performance and (ii) the use of an indirect fatigue-assessment method (blink rate) due to unavailability of a direct method in the dataset used in this study. The obtained findings can therefore be improved by using a more informative dataset.

Chapter 7

Conclusions and Future Work

7.1 Introduction

The need for a robust fatigue detection and assessment system has been addressed in this thesis. In particular three main objectives have been met: (i) working in naturalistic environments, (ii) building in a modular structure and (iii) capturing individual differences between operators. This is achieved through conducting numerous experiments in a real-life environment and a customised bio-mathematical model has been developed.

A functional pilot system prototype for fatigue detection and prediction that is based on a multi-sensors modular design approach was then developed, implemented and tested successfully. The main concluding remarks derived from this work as well as some challenges/limitations and suggestions for future work are given in the next sections.

7.2 Conclusions

1) A new multi-sensor fatigue detection system has been proposed and implemented successfully. The developed prototype demonstrated the feasibility of utilising low-cost wearable devices to detect operators' fatigue in real-life environments. The obtained test results showed a relatively high level of accuracy which outperforms state-of-the-art results in terms of classification performance and number of detection classes.

2) Based on Bayesian algorithm, a combiner is suggested to improve overall classification accuracy of the developed system and create the modularity structure. The obtained results demonstrated a 0-10% improvement when compared to those obtained from individual modules.

3) Fatigue-sensitivity of the keystroke-dynamic metrics and the naturalistic environments of typing activity were the reasons behind selection of keyboard typing to assess the performance of the developed fatigue detection system. The developed prototype is also based on an actual fatigue/alertness scale that is subjectively assessed by the participants. Despite challenges associated with conducting this study in a naturalistic environment, the obtained results demonstrated a promising performance when compared to the state-of-the-art findings. Furthermore, two other existing datasets are used to test the system with different real-life environments including a smartphone-based typing dataset.

4) A new approach for automatic detection of driver fatigue using smartphone's inertial sensors has also been presented and discussed. This approach utilises non-intrusive data collected from the built-in inertial sensors of a smartphone.

It can also estimate the smartphone's orientation relative to the vehicle frame orientation, no matter how it is oriented inside the vehicle. The proposed algorithm is also capable of distinguishing between the vehicle's and Earth's horizontal planes, thus improving accuracy of the orientation algorithm.

5) This developed orientation algorithm excludes the need for magnetic sensor data, which represents a main source of error for orientation estimation. No other instruments than the existing inertial sensors in the mobile phone are required. The real-life driving data and utilising of the non-intrusive data collection method used in this study have therefore led to the attaining of reliable and realistic results.

6) The proposed modular structure has been implemented and tested successfully. Different module combinations are experimented with and the obtained results supported the anticipated improvement of maintaining robustness and flexibility. This was clearly reflected in the obtained results and experimental observations of the modular design in the implemented system. A new modification is applied to the Bayesian combiner to enable the particle swarm optimiser algorithm to work together with the combiner for further classification improvements.

7) A customised circadian model has been built and tested using the heart-rate data. Developed of this model is based on theory relating to the shape invariant model. The developed model has not only demonstrated the circadian rhythm of participants successfully but also captured the individual differences in their circadian rhythms depending upon their bio-data.

8) The success with modelling the fatigue effect on the daily typing tasks of operators forms a solid base for building a non-intrusive prediction and

monitoring fatigue system. An adaptive control of thoughts-rational (ACT-R) architecture has been explored with the aim of adapting this architecture for fatigue impact. This is achieved through instantiating the fatigue effect in keyboard typing. Some preliminary results have been obtained and presented. These results demonstrate a strong correlation between the experimental and model-based data.

7.3 Challenges and Limitations

Although this work has met the planned objectives, some challenges and limitations were faced and their impact on the conducted research was mitigated through practical intervals. Of these, the following are the most important:

1) Fatigue is a complex feeling and thus needs close collaboration with domain-knowledge experts, which is not the case in this study. This challenge was faced through establishing private contact with experts in sleep research and psychologists from the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, and Carnegie Mellon University, Pittsburgh, Pennsylvania. . Their inputs and comments on various aspects relevant to this study were valuable.

2) Data collection in real-life environments requires special devices and approaches to keep the naturalistic behaviour of the operator. Also, restrictions of funding, work regulation and labour union rules limited the number of available participants. Hence, mobile, non-intrusive, noise-protected and accurate

devices were carefully selected and used. Also, a wide range of experiments was conducted to compensate for the relatively small number of participants.

7.4 Future Work

The work presented in this thesis can be further developed through conducting further studies which build on the developed fatigue detection system and the obtained research findings. Potential areas of extensions can be summarised as follows:

1) The proliferation of smartphones and other mobile devices along with continuous advances in mobile technology and application development have opened new ways of data collection and management for a wide range of applications including fatigue detection in real environments. For example, the smartphone can be used as a user terminal to collect and exchange data with other remote fatigue detection and assessment hubs which host a central database and business logic of distributed fatigue-management scenarios.

2) In the present study, the circadian customisation and the modular structure of the developed system have been dealt with separately. Further improvement can be made through combining both in an integrated framework. This is expected to combine the benefits of both frameworks.

3) The success with fatigue instantiation in modelling the daily typing tasks of operators, which are briefly demonstrated in the next section, will offer a good opportunity for building a non-intrusive prediction and monitoring fatigue system. The adaptive control of though-rational (ACT-R) cognitive

architecture was adopted in this study because it relies on well-established psychological background theories. The next section describes briefly the use of an ACT-R model to simulate the fatigue effect of operator performance while typing and preliminary results are shown.

7.5 Future Work: Cognitive Modelling and Fatigue Prediction

Fatigue prediction that needs robust mathematical modelling has not been validated in the literature. This is mainly due to unavailability of adequate real data that can be used for modelling and validation purposes. Fatigue effect on the operator performance has been mainly investigated using a bio-mathematical modelling which essentially depends on a two-process model of sleep regulation. Recently, one group of researchers adopted new approach in fatigue prediction cognitive modelling. It depends on a hypothesis model called adaptive control of though-rational (ACT-R) architecture. This approach is followed based on a fatigue-sensitive task.

7.5.1 Fatigue instantiation in ACT-R Architecture

Although ACT-R is a programming architecture, it is opposite to most programming languages in several major concepts. The first difference is that ACT-R is the low-level script used to manage a brain (processor) with high-level abstracted capabilities, while the other programming languages are a high-level set of instructions written to run a low-level processor for execution. The second

difference is the order of execution the commands (productions) with ACT-R architecture does not follow the order these commands like most programming languages. These commands (productions) are executed (fired) based on their matches with the current state of the buffers and modules. Finally, and maybe the most important difference between ACT-R architecture and programming languages is that ACT-R is trying to simulate human behaviour which is not always is optimal if it is viewed from programmer perspective [203].

Fatigue effect on operators, using the ACT-R frame, was previously reported in [204], [205], [206]. In these studies, a model was developed to simulate the behaviour changing of participants due to sleep deprivation or type of task [207]. Most of the experiments that were conducted to generate data were based on psychomotor vigilance test (PVT) due to (i) a strong correlation between PVT metrics and operator performance and (ii) easiness of conducting this test in the ACT-R frame [206]. In this work, for the fist time, fatigue effect is simulated is after naturalistic typing task account is instantiated in ACT-R environments.

7.5.2 ACT-R Based Fatigue Modelling and Implementation

In this study, several fatiguerelated parameters (i.e. which include production utility function, noise and cycle time) [205] are adjusted to instantiate fatigue impact on the operator behaviour. These parameters and their relation to fatigue will be described briefly as follows.

1) *Utility function:* It calculates the likelihood of the production to be fired (U) from [205]

$$U = PG - C + \varepsilon \tag{7.1}$$

where P is the likelihood of achieving the chosen goal by the production, G is the goal value, C is the cost value to achieve the mentioned goal, and ε is a random noise related to variation between cycles.

2) *Production time:* It is the time needs by the central procedural unit to fire the current production after it matches the current state. The suggested default number with ACT-R model is 50 ms [208]. Changing of this parameter value is control the model processing speed. The shorter production time is the faster task completion and vice versa.

An experiment was conducted to show the harm effect of sleep deprivation on psychomotor vigilance [75]. The experiment recruited 14 participants (seven males and seven females) to conduct four typing sessions for each participant. Two of these typing sessions were done under rest status and the others were done under sleep inertia status. The existing dataset that collected from the explained experience is used in this study to develop an ACT-R model for typing task.

The model is tested by typing a paragraph with the same number of clicks of each participant's transcription. Then, deference between digraph mean of the experimental and model-based data is used to optimise the fatigue-sensitive parameters. Figure 7.1 illustrate the procedure of adjusting fatigue-related parameters in ACT-R typing model to have best fit with experimental digraph data of the 14th participants.

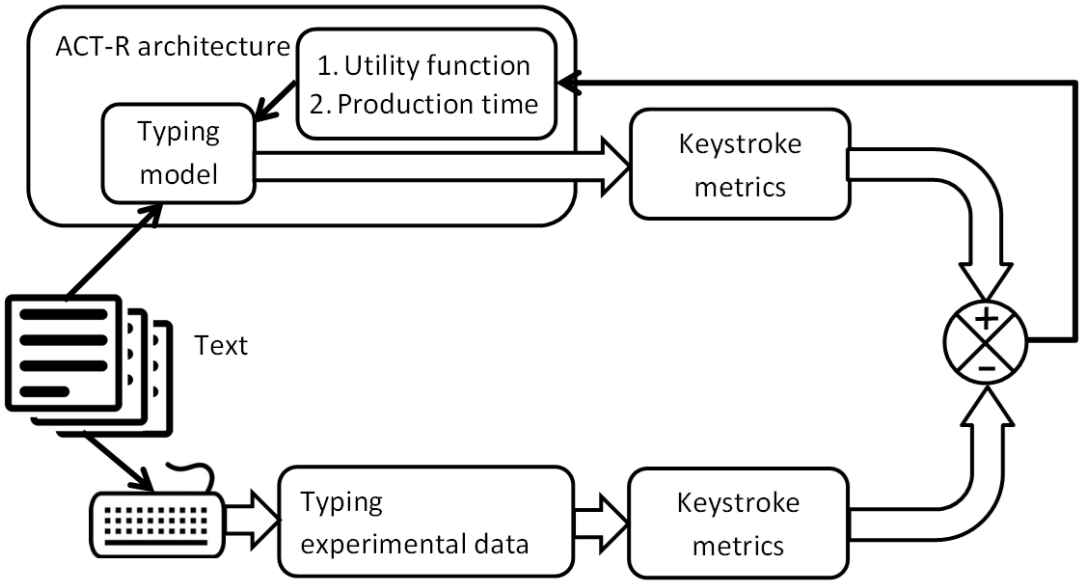
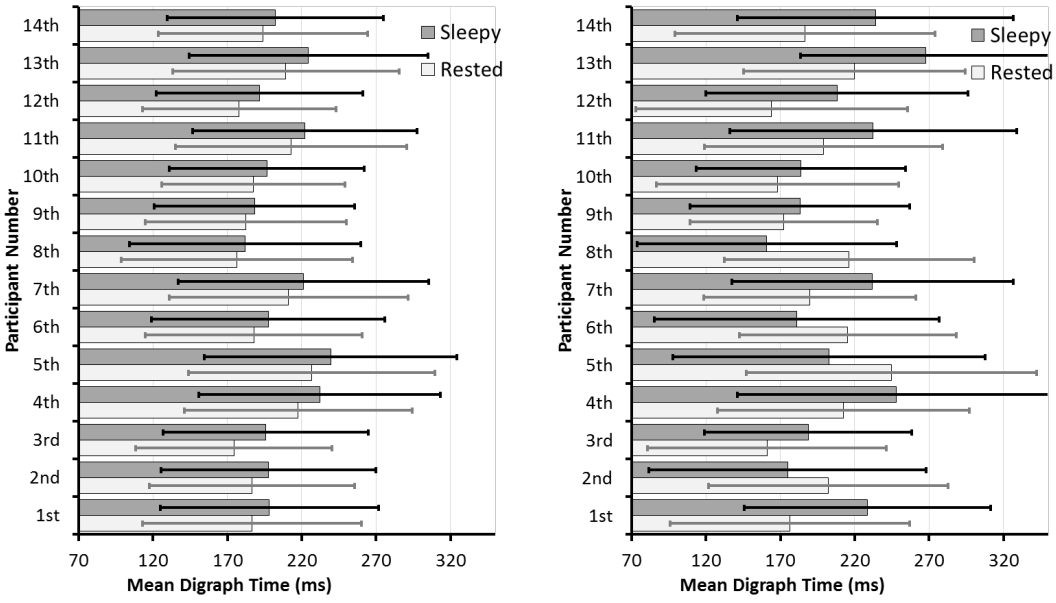


FIGURE 7.1: Model-based performance parameter tuning

7.5.3 Preliminary results

The statistical analysis of the experimental data shows a clear effect of fatigue on typing performance. Digraph (time between two consecutive key-presses) is one of the basic keystroke metrics. The digraph-time variation between two states of alertness in same participant is shown in Figure 7.2(a). This figure demonstrates the effect of fatigue on digraph-time. All participant showed an increase in the average digraph-time when participants experience fatigue. Moreover, the standard deviation of the digraph-times follows the same trends of the average that reflects more variation when the participant experiences more fatigue. However, for model-based generated data, almost same trend illustrated in Figure 7.2(b) except few participants (such as 2nd, 5th) show unexpected results where digraph-time of rested is higher than sleepy participant. This unexpected observation may rise from randomness effect of utility function as well as differences between typed texts.



(a) Experimental

(b) Model-base

FIGURE 7.2: Digraph-time variations between rested and sleepy states for experimental and model based data

Bibliography

[1] Valerie J Gawron, Jonathan French, and Doug Funke. An overview of fatigue. In *Stress, workload, and fatigue*, pages 581–595. Lawrence Erlbaum Associates Publishers, 2001.

[2] Matt T Bianchi. *Sleep Deprivation and Disease: Effects on the Body, Brain and Behavior*. Springer Science & Business Media, 2013.

[3] H. Abbood, W. Al-Nuaimy, A. Al-Ataby, S. A. Salem, and H. S. AlZubi. Prediction of driver fatigue: Approaches and open challenges. In *2014 14th UK Workshop on Computational Intelligence (UKCI), Bradford, England, 8 - 10 September 2014*, pages 1–6. IEEE.

[4] Jeanne F Duffy, Kirsi-Marja Zitting, and Charles A Czeisler. The case for addressing operator fatigue. *Reviews of human factors and ergonomics*, 10(1):29–78, 2015. Sage Publications Sage CA: Los Angeles, CA.

[5] M Stork, J Skala, P Weissar, R Holota, and Z Kubik. Various approaches to driver fatigue detection: A review. In *International Conference on Applied Electronics (AE), 2015*, pages 239–244. IEEE, 2015.

[6] Hilal Al-Libawy, Waleed Al-Nuaimy, Ali Al-Ataby, and Majid A. Al-Taee. Estimation of driver alertness using low-cost wearable devices. In *Conference*

on Applied Electrical Engineering and Computing Technologies (AEECT), Jordan, pages 1–5. IEEE, 2015.

[7] Hilal Al-Libawy, Ali Al-Ataby, Waleed Al-Nuaimy, and Majed A. Al-Taee. HRV-based operator fatigue analysis and classification using wearable sensors. In *13th International Multi-Conference on Systems, Signals & Devices (SSD), March 21 - 24, 2016 - Leipzig, Germany*, pages 268 – 273. IEEE.

[8] Centre for Road Safety. NSW road fatalities report. Technical report, Dec 2016. `http://roadsafety.transport.nsw.gov.au/downloads/road-fatalities-2106.pdf` "[Online; accessed 21-May-2014]".

[9] Thomas J Balkin, William J Horrey, R Curtis Graeber, Charles A Czeisler, and David F Dinges. The challenges and opportunities of technological approaches to fatigue management. *Accident Analysis & Prevention*, 43(2):565–572, 2011. Elsevier.

[10] Andrew Vakulin, Jillian Dorrian, Tristan R Duggan, Carla A Litchfield, Katherine J Cobb, Nick A Antic, Doug McEvoy, Stuart D Baulk, and Peter G Catcheside. Behavioural observation as a means of assessing sleepiness related driving impairment in obstructive sleep apnea. *Eat, Sleep, Work*, 1(1), 2016. Open Journal Systems.

[11] Zuojin Li, Shengbo Eben Li, Renjie Li, Bo Cheng, and Jinliang Shi. Online detection of driver fatigue using steering wheel angles for real driving conditions. *Sensors*, 17(3):495, 2017. Multidisciplinary Digital Publishing Institute.

[12] Wei Sun, Xiaorui Zhang, Srinivas Peeta, Xiaozheng He, Yongfu Li, and Senlai Zhu. A self-adaptive dynamic recognition model for fatigue driving based on multi-source information and two levels of fusion. *Sensors*, 15(9):24191–24213, 2015. Multidisciplinary Digital Publishing Institute.

[13] Arun Sahayadhas, Kenneth Sundaraj, and Murugappan Murugappan. Detecting driver drowsiness based on sensors: a review. *Sensors*, 12(12):16937–16953, 2012. Multidisciplinary Digital Publishing Institute.

[14] Alec Levenson. Workplace fatigue is a systems problem. *Consulting Psychology Journal: Practice and Research*, 69(2):130, 2017. Educational Publishing Foundation.

[15] Karl E Friedl, Melissa M Mallis, Stephen T Ahlers, Stephen M Popkin, and Willard Larkin. Research requirements for operational decision-making using models of fatigue and performance. *Aviation, space, and environmental medicine*, 75(Supplement 1):A192–A199, 2004.

[16] Melissa M Mallis, Sig Mejdal, Tammy T Nguyen, and David F Dinges. Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviation, space, and environmental medicine*, 75(3):A4–A14, 2004. Aerospace Medical Association.

[17] Lin Lu, Fadel M. Megahed, Richard F. Sesek, and Lora A. Cavuoto. A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among u.s. manufacturing workers. *Applied Ergonomics*, 65:139 – 151, 2017. PubMed.

[18] Civil Aviation Safety Authority. Biomathematical fatigue models, 2014. `https://www.casa.gov.au/sites/g/files/net351/f/_assets/main/aoc/fatigue/fatigue_modelling.pdf` [Online; accessed 21-Oct-2017].

[19] Steven R Hursh, Daniel P Redmond, Michael L Johnson, David R Thorne, Gregory Belenky, Thomas J Balkin, William F Storm, James C Miller, and Douglas R Eddy. Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine*, 75(Supplement 1):A44–A53, 2004. Aerospace Medical Association.

[20] Tu Mushenko. Getting real about biomathematical fatigue models, 2014. `hhttps://www.interdynamics.com/download/articles/GettingRealAboutBMMs.pdf` [Online; accessed 21-Oct-2017].

[21] Jeanne F Duffy, Kirsi-Marja Zitting, and Charles A Czeisler. The case for addressing operator fatigue. *Reviews of human factors and ergonomics*, 10(1):29–78, 2015. Sage Publications Sage CA: Los Angeles, CA.

[22] Drew Dawson, Janine Chapman, and Matthew JW Thomas. Fatigue-proofing: a new approach to reducing fatigue-related risk using the principles of error management. *Sleep medicine reviews*, 16(2):167–175, 2012. Elsevier.

[23] Khosro Sadeghniiat-Haghighi and Zohreh Yazdi. Fatigue management in the workplace. *Industrial psychiatry journal*, 24(1):12, 2015. Medknow Publications.

[24] Lori Mooren, Raphael Grzebieta, Ann Williamson, Jake Olivier, and Rena Friswell. Safety management for heavy vehicle transport: A review of the literature. *Safety Science*, 62:79 – 89, 2014. Elsevier.

[25] Caterpillar Global Mining. Viewpoint perspectives on modern mining. `https://mining.cat.com/cda/files/2785330/7/Viewpoint_2007_issue2.pdf`, 2007. [Online; accessed 21-May-2014].

[26] P. Jackson, C. Hilditch, A. Holmes, N. Reed, N. Merat, and L. Smith. Fatigue and road safety:a critical analysis of recent evidence. Technical report, London: Department for Transport, 02 2011.

[27] T. Akerstedt, R. Mollard, A. Samel, M. Simons, and M. Spencer. Meeting to discuss the role of EU FTL legislation in reducing cumulative fatigue in civil aviation. `http://archive.etsc.eu/documents/pre_19feb03.pdf`, 02 2003. [Online; accessed 21-May-2014].

[28] Ross O Phillips. A review of definitions of fatigue–and a step towards a whole definition. *Transportation research part F: traffic psychology and behaviour*, 29:48–56, 2015. Elsevier.

[29] Michael Ingre, Wessel Van Leeuwen, Tomas Klemets, Christer Ullvetter, Stephen Hough, Göran Kecklund, David Karlsson, and Torbjörn Åkerstedt. Validating and extending the three process model of alertness in airline operations. *PLoS ONE*, 9:e108679, 10 2014. Public Library of Science.

[30] RandolphW. Hall. Transportation science. In RandolphW. Hall, editor, *Handbook of Transportation Science*, volume 56 of *International Series in Operations Research and Management Science*, pages 1–5. Springer US, 2003.

[31] Yantao Qiao, Kai Zeng, Lina Xu, and Xiaoyu Yin. A smartphone-based driver fatigue detection using fusion of multiple real-time facial features. In

Consumer Communications & Networking Conference (CCNC), 2016 13th IEEE Annual, pages 230–235. IEEE, 2016.

[32] Lizong Lin, Chao Huang, Xiaopeng Ni, Jiawen Wang, Hao Zhang, Xiao Li, and Zhiqin Qian. Driver fatigue detection based on eye state. *Technology and health care*, 23(s2):S453–S463, 2015. IOS Press.

[33] Wasan Tansakul and Poj Tangamchit. Fatigue driver detection system using a combination of blinking rate and driving inactivity. *Journal of Automation and Control Engineering*, 4(1), 2016.

[34] G Sun, Y Jin, Z Li, F Zhang, and L Jia. A vision-based head status judging algorithm for driving fatigue detection system. *Advances in Transportation Studies*, (37), 2015.

[35] Nawal Alioua, Aouatif Amine, and Mohammed Rziza. Drivers fatigue detection based on yawning extraction. *International journal of vehicular technology*, 2014, 2014. Hindawi Publishing Corporation.

[36] Valerie J. Gawron. Overview of self-reported measures of fatigue. *The International Journal of Aviation Psychology*, 26(3-4):120–131, 2016. Taylor & Francis.

[37] JS Dixon and HA Bird. Reproducibility along a 10 cm vertical visual analogue scale. *Annals of the Rheumatic Diseases*, 40(1):87–89, 1981. BMJ Publishing Group Ltd.

[38] Helen J Michielsen, Jolanda De Vries, and Guus L Van Heck. Psychometric qualities of a brief self-rated fatigue measure: The fatigue assessment scale. *Journal of psychosomatic research*, 54(4):345–352, 2003. Elsevier.

[39] Anna Brndal, Marie Eriksson, Per Wester, and Lillemor Lundin-Olsson. Reliability and validity of the swedish fatigue assessment scale when self-administrered by persons with mild to moderate stroke. *Topics in Stroke Rehabilitation*, 23(2):90–97, 2016. Taylor & Francis.

[40] Torbjörn Åkerstedt and Gillberg Mats. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2):29–37, 1990. Informa UK Ltd UK.

[41] Chin-Shun Hsieh and Cheng-Chi Tai. A fatigue state evaluation system based on the band energy of electroencephalography signals. *Sensors and Materials*, 25(9):697–706, 2013.

[42] Maarten A Hogervorst, Anne-Marie Brouwer, and Jan BF van Erp. Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, 8, 2014. Frontiers Media SA.

[43] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[44] Michal Teplan. Fundamentals of EEG measurement. *Measurement science review*, 2(2), 2002. De Gruyter.

[45] Leonard J Trejo, Karla Kubitz, Roman Rosipal, Rebekah L Kochavi, and Leslie D Montgomery. Eeg-based estimation and classification of mental fatigue. *Psychology*, 6(05):572, 2015. Scientific Research Publishing.

[46] Sylvie Charbonnier, Raphaelle N. Roy, Stephane Bonnet, and Aurelie Campagne. EEG index for control operators mental fatigue monitoring using

interactions between brain regions. *Expert Systems with Applications*, 52:91 – 98, 2016. Elsevier.

[47] Simon L. Kappel, David Looney, Danilo P. Mandic, and Preben Kidmose. Physiological artifacts in scalp eeg and ear-eeg. *BioMedical Engineering OnLine*, 16(1):103, Aug 2017. BMC.

[48] A Blazquez, A Martinez-Nicolas, FJ Salazar, MA Rol, and JA Madrid. Wrist skin temperature, motor activity, and body position as determinants of the circadian pattern of blood pressure. *Chronobiology international*, 29(6):747–756, 2012. Taylor & Francis.

[49] Kurt Kräuchi, Christian Cajochen, and Anna Wirz-Justice. Thermophysiologic aspects of the three-process-model of sleepiness regulation. *Clinics in sports medicine*, 24(2):287–300, 2005. Elsevier.

[50] Ashrant Aryal, Ali Ghahramani, and Burcin Becerik-Gerber. Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction*, pages –, 2017. Elsevier.

[51] Bart Roelands, Jeroen Van Cutsem, Samuele Marcora, and Romain Meeusen. Does mental fatigue alter core and skin temperature in the heat? *Medicine & Science in Sports & Exercise*, 48(5 Suppl 1):123, 2016. Lippincott Williams and Wilkins.

[52] Malini Subramanyam, Prajwal Muralidhara, and M Pooja. Mental workload and cognitive fatigue: A study. *IUP Journal of Management Research*, 12(2):29, 2013. IUP Publications.

[53] Elena Miro, MC Cano-Lozano, and Gualberto Buela-Casal. Electrodermal activity during total sleep deprivation and its relationship with other

activation and performance measures. *Journal of Sleep Research*, 11(2):105–112, 2002. Wiley Online Library.

[54] Edward W. Geldreich. Skin conductance changes occurring during mental fatigue. *Transactions of the Kansas Academy of Science (1903-)*, 42:393–395, 1939. Kansas Academy of Science.

[55] M. K. Sharma and M. M. Bundele. Design and analysis of performance of k-means algorithm for cognitive fatigue detection in vehicular drivers using skin conductance signal. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 707–712, March 2015. New Delhi, India.

[56] M. M. Bundele and R. Banerjee. An svm classifier for fatigue-detection using skin conductance for use in the bits-lifeguard wearable computing system. In *2009 Second International Conference on Emerging Trends in Engineering Technology*, pages 934–939, Dec 2009.

[57] C Heinze, U Trutschel, T Schnupp, D Sommer, A Schenka, J Krajewski, and M Golz. Operator fatigue estimation using heart rate measures. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 930–933. Springer, 2010.

[58] Faizul Rizal Ismail, Nor Kamaliana Khamis, Mohd Zaki Nuawi, Dieter Schramm, and Benjamin Hesse. Measurement of heart rate to determine car drivers performance impairment in simulated driving: An overview. *Jurnal Teknologi*, 78(6-9):15–23, 2016. Penerbit Universiti Teknologi Malaysia.

[59] Saroj K.L. Lal and Ashley Craig. A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, 55(3):173 – 194, 2001. Elsevier.

[60] Mohammed Aqel and Majid Al-Taee. Mobile-based interpreter of arterial blood gases using knowledge-based expert system. *J. of Advances in Modelling (B) - Signal Processing and Pattern Recognition*, 45(4):61–71, 2002. AMSE.

[61] Saif Ahmad, Miodrag Bolic, Hilmi Dajani, Voicu Groza, Izmail Batkin, and Sreeraman Rajan. Measurement of heart rate variability using an oscillometric blood pressure monitor. *IEEE Transactions on Instrumentation and Measurement*, 59(10):2575–2590, 2010. IEEE.

[62] James McNames and Mateo Aboy. Reliability and accuracy of heart rate variability metrics versus ecg segment duration. *Medical and Biological Engineering and Computing*, 44(9):747–756, 2006. Springer.

[63] G Dorfman Furman, A Baharav, C Cahan, and S Akselrod. Early detection of falling asleep at the wheel: A heart rate variability approach. In *Computers in Cardiology, 2008*, pages 1109–1112. IEEE, 2008.

[64] Ross Owen Phillips. What Is Fatigue and How Does It Affect the Safety Performance of Human Transport Operators? Fatigue in Transport Report I. `https://www.toi.no/getfile.php?mmfileid=38953`, 2014. [Online; accessed 21-Oct-2017].

[65] Sofia Isabel Ribeiro Pereira, Felipe Beijamini, Roberta Almeida Vincenzi, and Fernando Mazzilli Louzada. Re-examining sleeps effect on motor skills: How to access performance on the finger tapping task? *Sleep Science*, 8(1):4 – 8, 2015. Elsevier.

[66] Kanav Kahol, Mario J Leyba, Mary Deka, Vikram Deka, Stephanie Mayes, Marshall Smith, John J Ferrara, and Sethuraman Panchanathan. Effect

of fatigue on psychomotor and cognitive skills. *The American Journal of Surgery*, 195(2):195–204, 2008. Elsevier.

[67] Samuel G Charlton and Peter H Baas. Fatigue, work-rest cycles, and psychomotor performance of new zealand truck drivers. *New Zealand Journal of Psychology*, 30(1):32, 06 2001. `https://search.proquest.com/docview/212442453?accountid=12117`.

[68] Nicole Lamond and Drew Dawson. Quantifying the performance impairment associated with fatigue. *Journal of sleep research*, 8(4):255–262, 1999. Wiley Online Library.

[69] Steven Lerman, Daniel Mollicone, and Sara Coats. Use of the psychomotor vigilance test in fitness for work assessments. *Journal of Occupational and Environmental Medicine*, 2017. LWW.

[70] Marilyn H Oermann. Psychomotor skill development. *The Journal of Continuing Education in Nursing*, 21(5):202–204, 1990. SLACK Incorporated.

[71] In-Soo Lee, Wayne A Bardwell, Sonia Ancoli-Israel, and Joel E Dimsdale. Number of lapses during the psychomotor vigilance task as an objective measure of fatigue. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 6(2):163, 2010. American Academy of Sleep Medicine.

[72] Ehsan B Khosroshahi, Dario D Salvucci, Bella Z Veksler, and Glenn Gunzelmann. Capturing the effects of moderate fatigue on driver performance. In *Proceedings of the 14th International Conference on Cognitive Modeling*, pages 163–168. Pennsylvania, USA.

[73] Matti Koivisto and Andrew Urbaczewski. Accuracy metrics in mobile text entry. In *Proc. 1st IASTED Conf. on Human-Computer Interaction*, 2005.

[74] Pin Shen Teh, Andrew Beng Jin Teoh, and Shigang Yue. A survey of keystroke dynamics biometrics. *The Scientific World Journal*, 2013, 2013. Hindawi Publishing Corporation.

[75] L Giancardo, A Sánchez-Ferro, I Butterworth, CS Mendoza, and JM Hooker. Psychomotor impairment detection via finger interactions with a computer keyboard during natural typing. *Scientific reports*, 5:9678, 2015. Nature Publishing Group.

[76] R William Soukoreff and I Scott MacKenzie. Metrics for text entry research: an evaluation of msd and kspc, and a new unified error metric. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120. ACM, 2003.

[77] Ahmed Sabbir Arif and Wolfgang Stuerzlinger. Analysis of text entry performance metrics. In *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, pages 100–105. IEEE, 2009.

[78] Zdeňka Sitová, Jaroslav Šeděnka, Qing Yang, Ge Peng, Gang Zhou, Paolo Gasti, and Kiran S Balagani. Hmog: New behavioral biometric features for continuous authentication of smartphone users. *IEEE Transactions on Information Forensics and Security*, 11(5):877–892, 2016. IEEE.

[79] André Pimenta, Davide Carneiro, Paulo Novais, and José Neves. Analysis of human performance as a measure of mental fatigue. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 389–401. Springer, 2014.

[80] Gys Albertus Marthinus Meiring and Hermanus Carel Myburgh. A review of intelligent driving style analysis systems and related artificial intelligence algorithms. *Sensors*, 15(12):30653–30682, 2015. MDPI.

[81] Pradnya N Bhujbal and Sandipann P Narote. Lane departure warning system based on hough transform and euclidean distance. In *2015 Third International Conference on Image Information Processing (ICIIP)*, pages 370–373. IEEE, 2015.

[82] Kwang-Hee Lee, Jeong-Hyeon Bak, and Chul-Hee Lee. Study on active steering control of vehicle for safe driving in highway with gps information. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 554–557. IEEE, 2012.

[83] Samuel Lawoyin, Ding-Yu Fei, and Ou Bai. Accelerometer-based steering-wheel movement monitoring for drowsy-driving detection. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of automobile engineering*, 229(2):163–173, 2015. Sage Publications Sage UK: London, England.

[84] Samuel A Lawoyin, Ding-Yu Fei, Ou Bai, et al. A novel application of inertial measurement units (imus) as vehicular technologies for drowsy driving detection via steering wheel movement. *Open Journal of Safety Science and Technology*, 4(04):166, 2014. Scientific Research Publishing.

[85] Jarret Engelbrecht, Marthinus Johannes Booysen, Gert-Jan van Rooyen, and Frederick Johannes Bruwer. Survey of smartphone-based sensing in vehicles for intelligent transportation system applications. *IET Intelligent Transport Systems*, 9(10):924–935, 2015. IET.

[86] A Koenig, T Rehg, and R Rasshofer. Statistical sensor fusion of ecg data using automotive-grade sensors. *Advances in Radio Science: ARS*, 13:197, 2015. Copernicus GmbH.

[87] H. Yin, Y. Su, Y. Liu, and D. Zhao. A driver fatigue detection method based on multi-sensor signals. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–7, March 2016.

[88] Céline Craye, Abdullah Rashwan, Mohamed S. Kamel, and Fakhri Karray. A multi-modal driver fatigue and distraction assessment system. *International Journal of Intelligent Transportation Systems Research*, 14(3):173–194, Sep 2016. Springer.

[89] Rongrong Fu, Hong Wang, and Wenbo Zhao. Dynamic driver fatigue detection using hidden markov model in real driving condition. *Expert Systems with Applications*, 63:397 – 411, 2016. Elsevier.

[90] Zuojin Li, Liukui Chen, Jun Peng, and Ying Wu. Automatic detection of driver fatigue using driving operation information for transportation safety. *Sensors*, 17(6):1212, 2017. Multidisciplinary Digital Publishing Institute.

[91] Khan Richard Baykaner, Mark Huckvale, Iya Whiteley, Svetlana Andreeva, and Oleg Ryumin. Predicting fatigue and psychophysiological test performance from speech for safety-critical environments. *Frontiers in bioengineering and biotechnology*, 3, 2015. Frontiers Media SA.

[92] Alexander A Borbély, Serge Daan, Anna Wirz-Justice, and Tom Deboer. The two-process model of sleep regulation: a reappraisal. *Journal of sleep research*, 2016. Wiley Online Library.

[93] Joshua B Gross, Glenn Gunzelmann, Kevin A Gluck, Hans PA Van Dongen, and David F Dinges. Computational modeling of the combined effects of circadian rhythm and sleep deprivation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 297–302, 2006.

[94] Drew Dawson, Y. Ian Noy, Mikko Hrm, Torbjorn kerstedt, and Gregory Belenky. Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, 43(2):549 – 564, 2011. Advancing Fatigue and Safety Research.

[95] Alexander A Borbély. A two process model of sleep regulation. *Human Neurobiology*, 1:195–204, 1982. Springer-Verlag.

[96] Namni Goel, Mathias Basner, Hengyi Rao, and David F Dinges. Circadian rhythms, sleep deprivation, and human performance. *Progress in Molecular Biology and Translational Science*, 119:155–190, 2013. NIH Public Access.

[97] Anne C Skeldon, Derk-Jan Dijk, and Gianne Derks. Mathematical models for sleep-wake dynamics: comparison of the two-process model and a mutual inhibition neuronal model. *PloS one*, 9(8):e103877, 2014. Public Library of Science.

[98] T Akerstedt and S Folkard. A model of human sleepiness. *Horne JA. Sleep*, 90:310–313, 1990. Pontenagel Press.

[99] Timothy H Monk. The post-lunch dip in performance. *Clinics in Sports Medicine*, 24(2):e15–e23, 2005. Elsevier.

[100] John A List and Steven D Levitt. What do laboratory experiments tell us about the real world. *NBER working paper*, pages 1–60, 2005.

[101] Gregory Mitchell. Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, 7(2):109–117, 2012. Sage Publications Sage CA: Los Angeles, CA.

[102] David B Miller. Roles of naturalistic observation in comparative psychology. *American Psychologist*, 32(3):211, 1977. American Psychological Association.

[103] Pooja Rajdev, David Thorsley, Srinivasan Rajaraman, Tracy L Rupp, Nancy J Wesensten, Thomas J Balkin, and Jaques Reifman. A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *Journal of Theoretical Biology*, 331:66–77, 2013. Elsevier.

[104] Pål O Westermark and Hanspeter Herzel. Mechanism for 12 hr rhythm generation by the circadian clock. *Cell reports*, 3(4):1228–1238, 2013. Elsevier.

[105] Robert Wang, Gordon Blackburn, Milind Desai, Dermot Phelan, Lauren Gillinov, Penny Houghtaling, and Marc Gillinov. Accuracy of wrist-worn heart rate monitors. *Jama cardiology*, 2(1):104–106, 2017. American Medical Association.

[106] PEAK - Fitness and Sleeep Tracker Watch. `http://www.mybasis.com`. [Online; accessed 21-Oct-2015].

[107] Polar H7 heart rate sensor strap. `http://www.polar.com/uk-en/products/accessories/H7_heart_rate_sensor`. [Online; accessed 21-Apr-2017].

[108] Fluxtream: open source non-profit personal data visualization framework. `https://www.fluxtream.org/`. [Online; accessed 21-Apr-2017].

[109] Pablo Valdez, Candelaria Ramírez, and Aída García. Circadian rhythms in cognitive performance: implications for neuropsychological assessment. *ChronoPhysiology & Therapy*, 2:81–92, 2012. Dove Medical Press Ltd.

[110] Magali Cariou, Edith Galy, and Claudine Melan. Differential 24-hour variation of alertness and subjective tension in process controllers: Investigation of the relationship with body temperature and heart rate. *Chronobiology international*, 25(4):597–609, 2008. Taylor & Francis.

[111] Mahesh M Bundele and Rahul Banerjee. Detection of fatigue of vehicular driver using skin conductance and oximetry pulse: a neural network approach. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, pages 739–744. ACM, 2009.

[112] E Miró, MC Cano-Lozano, and Gualberto Buela-Casal. Electrodermal activity during total sleep deprivation and its relationship with other activation and performance measures. *Journal of Sleep Research*, 11(2):105–112, 2002. Wiley Online Library.

[113] Hans PA Van Dongen and David F Dinges. Circadian rhythms in fatigue, alertness, and performance. *Principles and practice of sleep medicine*, 20:391–9, 2000. Saunders Philadelphia.

[114] Natalia Toporikova, Megan Hastings Hagenauer, Paige Ferguson, and Victoria Booth. A two-process model for circadian and sleep-dependent

modulation of pain sensitivity. *bioRxiv*, page 098319, 2017. Cold Spring Harbor Labs Journals.

[115] Paul S Albert and Sally Hunsberger. On analyzing circadian rhythms data using nonlinear mixed models with harmonic terms. *Biometrics*, 61(4):1115–1120, 2005. Wiley Online Library.

[116] Yuedong Wang, Chunlei Ke, and Morton B Brown. Shape-invariant modeling of circadian rhythms with random effects and smoothing spline anova decompositions. *Biometrics*, 59(4):804–812, 2003. Wiley Online Library.

[117] Semhar Ogbagaber, Paul S Albert, Daniel Lewin, and Ronald J Iannotti. Summer activity patterns among teenage girls: harmonic shape invariant modeling to estimate circadian cycles. *Journal of circadian rhythms*, 10(1):2, 2012. BioMed Central.

[118] WH Lawton, EA Sylvestre, and MS Maggio. Self modeling nonlinear regression. *Technometrics*, 14(3):513–532, 1972. Taylor & Francis.

[119] Yuedong Wang and Morton B Brown. A flexible model for human circadian rhythms. *Biometrics*, pages 588–596, 1996. JSTOR.

[120] Noa Avni, Isaac Avni, Erez Barenboim, Bella Azaria, David Zadok, REUVEN KOHEN-RAZ, and Yair Morad. Brief posturographic test as an indicator of fatigue. *Psychiatry and clinical neurosciences*, 60(3):340–346, 2006. Wiley Online Library.

[121] Nermine Munla, Mohamad Khalil, Ahmad Shahin, and Azzam Mourad. Driver stress level detection using HRV analysis. In *2015 International*

Conference on Advances in Biomedical Engineering (ICABME), pages 61–64, Sept 2015.

[122] M Patel, SKL Lal, Diarmuid Kavanagh, and Peter Rossiter. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert systems with Applications*, 38(6):7235–7242, 2011. Elsevier.

[123] Zvi Shinar, Solange Akselrod, Yaron Dagan, and Armanda Baharav. Autonomic changes during wake–sleep transition: A heart rate variability based approach. *Autonomic Neuroscience*, 130(1):17–27, 2006. Elsevier.

[124] A Noda, M Miyaji, Y Wakuda, Y Hara, Toshio Yasuma, F andFukuda, Kunihiro Iwamoto, and Norio Ozaki. Simultaneous measurement of heart rate variability and blinking duration to predict sleep onset and drowsiness in drivers. *Journal of Sleep Disorders & Therapy*, 4(5), 2015. OMICS International.,.

[125] Farhad Abtahi, Anna Anund, Carina Fors, Fernando Seoane, and Kaj Lindecrantz. Association of drivers sleepiness with heart rate variability: A pilot study with drivers on real roads. In *EMBEC & NBC 2017*, pages 149–152. Springer, 2017. Springer.

[126] Hui-Min Wang and Sheng-Chieh Huang. SDNN/RMSSD as a surrogate forLF/HF: a revised investigation. *Modelling and Simulation in Engineering*, 2012:16, 2012. Hindawi Publishing Corp.

[127] JT Ramshur. Design, evaluation, and application of heart rate variability software (HRVAS). Master's thesis, The University of Memphis, Memphis, TN, 2010.

[128] Juha-Pekka Niskanen, Mika P Tarvainen, Perttu O Ranta-Aho, and Pasi A Karjalainen. Software for advanced hrv analysis. *Computer methods and programs in biomedicine*, 76(1):73–81, 2004. Elsevier.

[129] Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[130] Mika P Tarvainen, Perttu O Ranta-Aho, and Pasi A Karjalainen. An advanced detrending method with application to hrv analysis. *IEEE Transactions on Biomedical Engineering*, 49(2):172–175, 2002. IEEE.

[131] Melody Y Kiang. A comparative assessment of classification methods. *Decision Support Systems*, 35(4):441–454, 2003. Elsevier.

[132] Jake Lever, Martin Krzywinski, and Naomi Altman. Points of significance: Classification evaluation. *Nature Methods*, 13(8):603–604, 2016. Nature Research.

[133] Jianfeng Hu. Comparison of different features and classifiers for driver fatigue detection based on a single EEG channel. *Computational and mathematical methods in medicine*, 2017, 2017. Hindawi Publishing Corporation.

[134] Zhendong Mu, Jianfeng Hu, and Jianliang Min. Driver fatigue detection system using electroencephalography signals based on combined entropy features. *Applied Sciences*, 7(2):150, 2017. Multidisciplinary Digital Publishing Institute.

[135] Jack V Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11):1225–1231, 1996. Elsevier.

[136] Salim Chitroub. Classifier combination and score level fusion: concepts and practical aspects. *International Journal of Image and Data Fusion*, 1(2):113–135, 2010. Taylor & Francis.

[137] International society of information fusion (isif). `http://isif.org/`. Accessed: 2017-04-20.

[138] Dennis Bahler and Laura Navarro. Combining heterogeneous sets of classifiers: Theoretical and experimental comparison of methods. In *17th National Conference on Artificial Intelligence (AAAI 2000), Workshop on New Research Problems for Machine Learning*, 2000.

[139] Ludmila I Kuncheva. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.

[140] Hyun-Chul Kim and Zoubin Ghahramani. Bayesian classifier combination. In *Artificial Intelligence and Statistics*, pages 619–627, 2012.

[141] Fabio Roli. Multiple classifier systems. *Encyclopedia of Biometrics*, pages 1142–1147, 2015.

[142] Behnaz Bigdeli, Farhad Samadzadegan, and Peter Reinartz. A decision fusion method based on multiple support vector machine system for fusion of hyperspectral and lidar data. *International Journal of Image and Data Fusion*, 5(3):196–209, 2014. Taylor & Francis.

[143] Vladimir Naumovich Vapnik and Vlamimir Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.

[144] Bertrand Le Saux and Horst Bunke. Combining svm and graph matching in a bayesian multiple classifier system for image content recognition.

In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 696–704. Springer, 2006.

[145] Claudio De Stefano, Ciro D'Elia, Angelo Marcelli, and Alessandra Scotto di Freca. Using bayesian network for combining classifiers. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 73–80. IEEE, 2007.

[146] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009. Elsevier.

[147] Jianfeng Hu and Ping Wang. Noise robustness analysis of performance for eeg-based driver fatigue detection using different entropy feature sets. *Entropy*, 19(8):385, 2017. Multidisciplinary Digital Publishing Institute.

[148] James D Evans. *Straightforward statistics for the behavioral sciences*. Brooks/Cole, 1996.

[149] José Vicente, Pablo Laguna, Ariadna Bartra, and Raquel Bailón. Drowsiness detection using heart rate variability. *Medical & biological engineering & computing*, 54(6):927–937, 2016. Springer.

[150] Hilal Al-Libawy, Ali Al-Ataby, Waleed Al-Nuaimy, and Majed A. Al-Taee. Fatigue detection method based on smartphone text entry performance metrics. In *Proc. 9th International Conference on Developments in eSystems Engineering (DeSE 2016), Liverpool & Leeds, England, 31st August-2nd September*, 2016.

[151] André Pimenta, Davide Carneiro, Paulo Novais, and José Neves. Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In *Hybrid Artificial Intelligent Systems*, pages 222–231. Springer, 2013. Springer.

[152] Salil P Banerjee and Damon L Woodard. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, 7(1):116–139, 2012.

[153] Syed Zulkarnain Syed Idrus, Estelle Cherrier, Christophe Rosenberger, Soumik Mondal, and Patrick Bours. Keystroke dynamics performance enhancement with soft biometrics. In *2015 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–7. IEEE, 2015.

[154] Yu Zhong, Yunbin Deng, and Anil K Jain. Keystroke dynamics for user authentication. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 117–123. IEEE, 2012.

[155] Clayton Epp, Michael Lippold, and Regan L Mandryk. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 715–724. ACM, 2011.

[156] Roman V Yampolskiy and Venu Govindaraju. Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, 1(1):81–113, 2008. Inderscience Publishers.

[157] R William Soukoreff and I Scott MacKenzie. Measuring errors in text entry tasks: an application of the levenshtein string distance statistic. In *CHI'01*

extended abstracts on Human factors in computing systems, pages 319–320. ACM, 2001.

[158] Fred J Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3):171–176, 1964. ACM.

[159] Zhen Liang, Qiang Fu, and Zheru Chi. Eye typing of chinese characters. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 237–240. ACM, 2012.

[160] Elsbeth de Korte, Lottie Kuijt-Evers, and Peter Vink. Effects of the office environment on health and productivity 1: auditory and visual distraction. *Ergonomics and Health Aspects of Work with Computers*, pages 26–33, 2007. Springer.

[161] Ahmad Fuad Rezaur Rahman and Michael C Fairhurst. Multiple classifier decision combination strategies for character recognition: A review. *International Journal on Document Analysis and Recognition*, 5(4):166–194, 2003. Springer.

[162] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[163] Shigeo Abe. *Support vector machines for pattern classification*, volume 53. Springer, 2005. Springer.

[164] Azmeh Shahid, Jianhua Shen, and Colin M Shapiro. Measurements of sleepiness and fatigue. *Journal of psychosomatic research*, 69(1):81–89, 2010. Elsevier.

[165] Jianhua Shen, Joseph Barbera, and Colin M Shapiro. Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep medicine reviews*, 10(1):63–76, 2006. Elsevier.

[166] HMOG dataset. `http://www.cs.wm.edu/~qyang/hmog.htm`. [accessed 1-Apr-2017].

[167] Qing Yang, Ge Peng, David T Nguyen, Xin Qi, Gang Zhou, Zdeňka Sitová, Paolo Gasti, and Kiran S Balagani. A multimodal data set for evaluating continuous authentication performance in smartphones. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, pages 358–359. ACM, 2014.

[168] Anil Jain, Karthik Nandakumar, and Arun Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270 – 2285, 2005.

[169] Zuojin Li, Liukui Chen, Jun Peng, and Ying Wu. Automatic detection of driver fatigue using driving operation information for transportation safety. *Sensors*, 17(6):1212, 2017.

[170] Z. J. Muhsin, A. Al-Taee, M. A. Al-Taee, W. Al-Nuaimy, and A. Al-Ataby. Mobile workflow management system based on the internet of things. In *2016 13th International Multi-Conference on Systems, Signals Devices (SSD)*, pages 436–441, March 2016.

[171] AM Al-Taee, A Al-Taee, ZJ Muhsin, MA Al-Taee, and Waleed Al-Nuaimy. Towards developing online compliance index for self-monitoring of blood glucose in diabetes management. In *Proc. 9th International Conference on*

Developments in eSystems Engineering (DeSE 2016), Liverpool & Leeds, England, 31st August-2nd September, 2016.

[172] Yvette E Gelogo and H Kim. Development of mobile enterprise inventory management system application with cbd. *Int. J. of Software Engineering and Its Applications*, 8(1):385–396, 2014. Citeseer.

[173] Fatemeh Abyarjoo, Armando Barreto, Jonathan Cofino, and Francisco R Ortega. Implementing a sensor fusion algorithm for 3d orientation detection with inertial/magnetic sensors. In *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*, pages 305–310. Springer, 2015.

[174] Jiangpeng Dai, Jin Teng, Xiaole Bai, Zhaohui Shen, and Dong Xuan. Mobile phone based drunk driving detection. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS*, pages 1–8. IEEE, 2010.

[175] Derick A Johnson and Mohan M Trivedi. Driving style recognition using a smartphone as a sensor platform. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1609–1615. IEEE, 2011.

[176] Roberto G Valenti, Ivan Dryanovski, and Jizhong Xiao. Keeping a good attitude: A quaternion-based orientation filter for imus and margs. *Sensors*, 15(8):19302–19330, 2015. Multidisciplinary Digital Publishing Institute.

[177] Aleš Janota, Vojtech Šimák, Dušan Nemec, and Jozef Hrbček. Improving the precision and speed of euler angles computation from low-cost rotation sensor

data. *Sensors*, 15(3):7016–7039, 2015. Multidisciplinary Digital Publishing Institute.

[178] Joan Solà Ortega. Quaternion kinematics for the error-state KF. 2016. `http://healforce.com`, [Online; accessed 21-Oct-2017].

[179] Marco D Tundo, Edward Lemaire, and Natalie Baddour. Correcting smartphone orientation for accelerometer-based analysis. In *Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on*, pages 58–62. IEEE, 2013.

[180] William Rowan Hamilton. Ii. on quaternions; or on a new system of imaginaries in algebra. *Philosophical Magazine Series 3*, 25(163):10–13, 1844. Taylor & Francis.

[181] Basile Graf. Quaternions and dynamics. *arXiv preprint arXiv:0811.2889*, 2008. `https://arxiv.org/pdf/0811.2889.pdf`, [Online; accessed 21-Oct-2017].

[182] Michael B Del Rosario, Nigel H Lovell, and Stephen J Redmond. Quaternion-based complementary filter for attitude determination of a smartphone. *IEEE Sensors Journal*, 16(15):6008–6017, 2016. IEEE.

[183] FJ Bruwer and Marthinus J Booysen. Vehicle acceleration estimation using smartphone-based sensors. 2015. `https://repository.up.ac.za/bitstream/handle/2263/57729/Bruwer_Vehicle_2015.pdf?sequence=1`, [Online; accessed 21-Oct-2017].

[184] American Association of State Highway and Transportation Officials. *A Policy on Geometric Design of Highways and Streets, 2011*. AASHTO, 2011.

[185] Johannes Paefgen, Flavius Kehr, Yudan Zhai, and Florian Michahelles. Driving behavior analysis with smartphones: insights from a controlled field study. In *Proceedings of the 11th International Conference on mobile and ubiquitous multimedia*, page 36. ACM, 2012.

[186] Charles C Liu, Simon G Hosking, and Michael G Lenné. Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of safety research*, 40(4):239–245, 2009. Elsevier.

[187] Jarek Krajewski, David Sommer, Udo Trutschel, Dave Edwards, and Martin Golz. Steering wheel behavior based estimation of fatigue. In *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 118–124, 2009.

[188] Naturalistic Driving Study. `https://insight.shrp2nds.us/`. [accessed 1-Apr-2017].

[189] Nile W Hatch. Design rules, volume 1: The power of modularity. *Academy of Management Review*, 26(1):130–133, 2001. Academy of Management.

[190] Jeremy Avigad. Modularity in mathematics. `http://www.andrew.cmu.edu/user/avigad/Papers/modularity.pdf`, 2016.

[191] Neeraj Pradhan, Subinay Dasgupta, and Sitabhra Sinha. Modular organization enhances the robustness of attractor network dynamics. *EPL (Europhysics Letters)*, 94(3):38004, 2011. IOP Publishing.

[192] Ron Sanchez and Joseph T Mahoney. Modularity, flexibility, and knowledge management in product and organization design. *Strategic management journal*, 17(S2):63–76, 1996. Wiley Online Library.

[193] Orlando Durán, Luis Pérez, and Antonio Batocchio. Optimization of modular structures using particle swarm optimization. *Expert Systems with Applications*, 39(3):3507–3515, 2012. Elsevier.

[194] Iván García Daza, Luis Miguel Bergasa, Sebastián Bronte, Jose Javier Yebes, Javier Almazán, and Roberto Arroyo. Fusion of optimized indicators from advanced driver assistance systems (adas) for driver drowsiness detection. *Sensors*, 14(1):1106–1131, 2014. Multidisciplinary Digital Publishing Institute.

[195] Raofen Wang, Jianhua Zhang, Yu Zhang, and Xingyu Wang. Assessment of human operator functional state using a novel differential evolution optimization based adaptive fuzzy model. *Biomedical Signal Processing and Control*, 7(5):490–498, 2012. Elsevier.

[196] Ramnik Sharma and Anita Kumari. A review on traffic route optimizing by using different swarm intelligence algorithm. *IJCSMC*, 4(5):271–277, 2015. IJCSMC Publishing Team.

[197] D. Sandberg and M. Wahde. Particle swarm optimization of feedforward neural networks for the detection of drowsy driving. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 788–793, June 2008.

[198] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm intelligence*, 1(1):33–57, 2007. Springer.

[199] Aleksandar Lazinica. *Particle swarm optimization*. InTech Kirchengasse, 2009.

[200] Rania Hassan, Babak Cohanim, Olivier De Weck, and Gerhard Venter. A comparison of particle swarm optimization and the genetic algorithm. In *Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference*, volume 18, page e21, 2005.

[201] John A Stern, Donna Boyer, and David Schroeder. Blink rate: a possible measure of fatigue. *Human factors*, 36(2):285–297, 1994. Sage Publications Sage CA: Los Angeles, CA.

[202] Z. A. Haq and Z. Hasan. Eye-blink rate detection for fatigue determination. In *2016 1st India International Conference on Information Processing (IICIP)*, pages 1–5, Aug 2016.

[203] Dan Bothell. Act-r 7.0- toturial 1-act-r model writing. `http://act-r.psy.cmu.edu/wordpress/wp-content/themes/ACT-R/actr7/units.zip`, 2017. [Online; accessed 21-May-2017].

[204] Glenn Gunzelmann and Kevin A Gluck. Approaches to modeling the effects of fatigue on cognitive performance. In *Proceedings of the seventeenth conference on behavior representation in modeling and simulation*, pages 136–145. Citeseer, 2008.

[205] Glenn Gunzelmann, Joshua B Gross, Kevin A Gluck, and David F Dinges. Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Science*, 33(5):880–910, 2009. Wiley Online Library.

[206] GMG Jongman. How to fatigue act-r. In *Proceedings of the second European conference on cognitive modelling, Nottingham, UK*, pages 52–57, April, 1998.

[207] Leslie M Blaha, Christopher R Fisher, Matthew M Walsh, Bella Z Veksler, and Glenn Gunzelmann. Real-time fatigue monitoring with computational cognitive models. In *International Conference on Augmented Cognition*, pages 299–310. Springer, 2016.

[208] John R Anderson, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological review*, 111(4):1036, 2004. American Psychological Association.

Appendix A:

Certificates of Ethical Approval

UNIVERSITY OF LIVERPOOL

Faculty of Science and Engineering Committee on Research Ethics

11 June 2017

Dear Dr Al Ataby,

I am pleased to inform you that your application for research ethics approval has been approved. Details and conditions of the approval can be found below:

Reference:	1935
Project Title:	operator fatigue modelling
Principal Investigator/Supervisor:	Dr Ali Al Ataby
Co-Investigator(s):	Mr Hilal Al Libawy, Dr Waleed Al-Nuaimy
Lead Student Investigator:	-
Department:	Electrical Engineering and Electronics
Reviewers:	Dr Harm Van Zalinge, Dr Peter Green
Approval Date:	11/06/2017
Approval Expiry Date:	Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions

- All serious adverse events must be reported via the Research Integrity and Ethics Team (ethics@liverpool.ac.uk) within 24 hours of their occurrence.
- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the research, please create and submit an amendment form using the research ethics system.
- If the named Principal Investigator or Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form using the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Faculty of Science and Engineering Committee on Research Ethics

foseeth@liverpool.ac.uk

0151 795 0649

VIRGINIATECHVIRGINIATECHVIRGINIATECHVIRGINIATECH
Certificate of Completion
This certifies that
Hilal M. Libaw
Has completed
Training in Human Subjects Protection
On the following topics:
Historical Basis for Regulating Human Subjects Research
The Belmont Report
Federal and Virginia Tech Regulatory Entities, Policies and Procedures
on
July 21, 2016
David Moore, IRB Chair

Appendix B:

Wearable Devices Specifications

Basis Peak Watch

The Basis Peak fitness watch [106] as shown in Figure 3 is adopted in this study to collect the following bio-data from the participated subjects; heart rate, body temperature, and skin conductance. This watch comprises the following set of sensors:

(a) Optical blood flow sensor

(b) 3D accelerometer

(c) Body temperature reading

(d) Ambient temperature reading

(e) Conductivity sensor for galvanic skin response

This fitness watch stores bio-data in its internal memory. The stored data is exported to a mobile phone application which is in turn upload it to a remote server. The specification and price of this watch is listed in Table 1

FIGURE 3: Fitness tracker watch [106]

TABLE 1: Technical Specifications of Basis Peak Watch

Feature	Details
Battery Technology	Lithium
Battery life	4 days
Body placement	wrist
Waterproofing	Yes, 50 m
User Interface Type	Yes
Screen size	4.52 cm
Screen Type	LCD
Display resolution	144 x 168 pixels
Data Transfer type	Bluetooth 4.0
Weigh	51 g
Z-Height	3.49 cm
Width	4.13 cm
Depth	1.27 cm
Price	£140 (at time of purchase: October, 2015)

Polar H7 Chest Strap

Polar H7 is a chest strap hear monitor which uses an ECG (electrocardiogram) heart-rate sensor. It measures the electrical activity of the heart to deliver continuous and resting heart rate data. The H7 (shown in Figure 4) uses Bluetooth Smart to continuously transfer data to smartphone. The specification and price of this strap is listed in Table 2

FIGURE 4: Fitness heart-rate sensor Polar (H7) [107]

Table 2: Technical Specifications of Polar H7 Chest Strap

Feature	Details
Battery Type	CR 2025
Operation time	200 h
Body placement	Chest
HR measurement method	ECG, chest strap
Waterproofing	Yes, 30 m
Operating temperature	-10 C to +50 C / 14 F to 122 F
Strap material	38% Polyamide, 29% Polyurethane, 20% Elastane, 13% Polyester, Silicone prints
Data Transfer type	Bluetooth low energy (5khz)
Weigh (sensor/strap)	118 g/ 100 g
Height	1 cm
Width	2 cm
Length	3 cm
Price	£32 (at time of purchase: October, 2015)