Towards Real-time Tracking of Persons in Distress Phase Situations using Emotional Physiological Signals

By

Abdultaofeek Abayomi

(21451441)

Submitted in fulfilment of the requirements of the

Degree of Doctor of Philosophy in Information Technology

in the

Department of Information Technology

in the

Faculty of Accounting and Informatics

at the

Durban University of Technology

JANUARY 2019

Declaration

I, ABAYOMI A., hereby declare that this dissertation is my own work and has not been previously submitted in any form to any other university or institution of higher learning by other persons or myself. I further declare that all the sources of information used in this dissertation have been acknowledged and a list of references is provided.

[redacted]

A. Abayomi

Approved for final submission

Promoter: [redacted]

Professor Oludayo O. Olugbara (PhD)

Co-Promoter: [redacted]

Dr. Delene Heukelman (DTECH: IT) Date

Dedication

This thesis is dedicated to ALLAH (swt) - the First and the Last, the Most High, the Most Intimate; and He is the Knower of all things.

Acknowledgements

“We will attend to you, O prominent beings. So which of the favours of your Lord would you deny?” - Quran Chapter 55 vs 31-32.

All thanks, adoration and praises belong to ALLAH - the cherisher, the sustainer; who has made the successful completion of this doctoral programme.

First and foremost, I would like to specifically greatly appreciate and thank my supervisor, Prof. Oludayo Olugbara for his support, instructions, advices, priceless guidance, constructive criticisms and encouragements throughout my study without which it would have been extremely difficult to progress. Thank you Sir.

I am also sincerely indebted to Dr. Delene Heukelman - my co-supervisor, for all her efforts and contributions as well as the kindness, calmness and patience that she brought on board throughout this study even during difficult situations. Thanks Dr.

I would like to express my deepest appreciation to my ever-loving spouse – Nafisat, for being strong at the home front, her unwavering support, encouragement, commitment, care and patience while also assuming the roles of mother and father concurrently during my absence from home. Love you loads. My angels - Hameeda and Aisha, thank you very much for your maximum cooperation, obedience and understanding towards the successful running of the home by your darling mom.

I also specifically thank all members of the immediate and extended Abayomi’s and Salawu’s families for their prayers and support. May ALLAH reward all of you abundantly and grant the late Alhaji Hussein Atilola Abayomi an eternal rest.

Finally, in DUT, I really appreciate Prof Moyo Sibusiso as well as all students and staff of the IT Department. I also appreciate the assistance of the staff members of the Accounting and Informatics Faculty, the Research and Postgraduate Support Directorate and the Faculty’s postgraduate research support. A big thank you to my fellow postgraduate students in the laboratory, the ICT and Society Research niche, all

friends, well-wishers and everyone too numerous to mention who have positively touched my life and contributed in whatever little way towards the successful completion of this doctoral programme. God bless you all.

Table of Contents

Page

List of Figures

List of Tables

List of Equations

List of Abbreviations

ADL	Activities of Daily Living
ANEW	Affective Norms for English Words
ANN	Artificial Neural Networks
ANS	Autonomic Nervous System
AS	Affective Slider
BVP	Blood Volume Pulse
CCTV	Closed Circuit Television
COCOM	Contextual Control Model
CPU	Central Processing Unit
CWT	Continuous Wavelet Transform
DEAP	Database for Emotion Analysis using Physiological Signals
DRL	Driven Right Leg
DTW	Dynamic Time Warping
ECG	Electrocardiograph
EDA	Electrodermal Activity
EEG	Electroencephalograph
EMD	Empirical Model Decomposition
EMDB	Emotional Movie Database
EMG	Electromyography
EOG	Electro-oculogram

EPW	Ear Pulse Waves
FACS	Facial Action Codings
FAP	Facial Animation Parameter
FDP	Facial Detection Parameter
FFT	Fast Fourier Transformation
GAPED	Geneva Affective Picture Database
GEW	Geneva Emotion Wheel
GIS	Geographic Information System
GPRS	General Packet Radio Services
GPS	Global Positioning System
GSM	Global System for Mobile Communication
GSR	Galvanic Skin Response
HAR	Human Activity Recognition
HER	Human Emotion Recognition
HHT	Hilbert Huang Transform
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradient
HRV	Heart Rate Variability
IADS	International Affective Digitized Sounds
IAPS	International Affective Picture System
IBN	Intelligent Base Node

ICT	Information Communication Technology
ID	Integration Device
IFT	Inverse Fisher Transform
JAFFE	Japanese Female Facial Expression
JCS	Joint Cognitive System
k-NN	k-Nearest Neighbor
LBP	Local Binary Patterns
LDS	Linear Discriminant Analysis
LED	Light Emitting Diodes
LPC	Linear Prediction Coefficients
LPCC	Linear Prediction Cepstral Coefficient
LSM	Least Square Method
MAP	Maximum a Posteriori
MFCC	Mel Frequency Cepstral Coefficient
MLP	Multilayer Perceptron
MSE	Mean Square Error
NPV	Normalized Pulse Volume
OASIS	Open Affective Standardized Image Set
PAD	Pleasure-Arousal-Dominance
PANAS	Positive and Negative Affect Schedule
PCA	Principal Component Analysis

PCG	Phonocardiography
PLP	Perceptual Linear Prediction
PPG	Photoplethysmography
PRV	Pulse Rate Variability
PSD	Power Spectral Density
RAM	Random Access Memory
RBF	Radial Basis Function
RMS	Root Mean Square
RSST	Robust Singular Spectrum Transform
SAM	Self-Assessment Manikin
SH	Smart Home
SMS	Short Message Service
SNR	Signal to Noise Ratio
SpO_2	Blood Oxygen Saturation
SVM	Support Vector Machine
TCP/IP	Internet Protocol
UDP	User Datagram Protocol
uLBP	Uniform Local Binary Pattern
WSN	Wearable Sensor Network

Abstract

This research work investigates physiological signals based human emotion and its incorporation in an affective system architecture for real-time tracking of persons in distress phase situations to prevent the occurrence of casualties. In a casualty situation, a mishap has already occurred leading to life, limb and valuables being in a state of peril. However, in a distress phase situation, there is a high likelihood that a tragedy is about to occur unless an immediate assistance is rendered. The distress phase situations include the spate of kidnapping, human trafficking and terrorism related crimes that could lead to casualty such as loss of lives, properties, finances and destruction of infrastructure. These situations are of global concern and worldwide phenomenon that necessitate a system that could mitigate the alarming trend of social crimes. The novel idea of deploying a combination of data and knowledge driven approaches using wearable sensor devices supported by machine learning methods could prove useful as a preventive mechanism in a distress phase situation. Such a system could be achieved through modelling human emotion recognition, including the harvesting and recognising human emotion physiological signals. Different methods have been applied in emotion recognition domain because the extraction of relevant discriminating features has been identified as an unresolved and one of the most daunting aspects of physiological signals based human emotion recognition system. In this thesis, emotion physiological signals, image processing technique and shallow learning based on radial basis function neural network were used to construct a system for real-time tracking of persons in distress phase situations. The system was tested using the Database for Emotion Analysis using Physiological Signal (DEAP) to ascertain the recognition performance that could be achieved. Emotion representations such as Arousal, Valence, Dominance and Liking have been creatively mapped to different conditions of human safety and survival state like happy phase, distress phase and casualty phase in a real-time system for tracking of persons. The constructed system can practically benefit security agencies, emergency services, rescue teams

and restore confidence to both the potential victims and their family by proactively providing assistance in an emergency event of a distress phase situation. Moreover, the system would prove beneficial in stemming the tide of the identified societal crimes and tragedies by thwarting the successful progress of a distress phase situation through an application of information communication technology to address critical societal challenges. The performance of the recognition algorithmic component of the constructed system gives accuracy that outperforms the state of the art results based on deep learning techniques.

CHAPTER ONE

Introduction

The spate of abductions or kidnappings, human trafficking and terrorism related crimes leading to loss of lives, properties, financial loss and grounding of infrastructure is of global concern as these crimes are now a worldwide phenomenon (Harrendorf, Heiskanen and Malby 2010; Maldwyn 2012; Curry and Hughes 2014; Vannini, Detotto and Mccannon 2015; NYA 2016; Zen 2018), thus necessitating a system that could mitigate this alarming trend.

This research work proposes a system towards a real-time tracking of persons in distress phase situations based on human emotional state recognition of physiological signals. In a distressed and casualty situation, a mishap has already occurred leading to life, limb, valuables or properties being in a state of peril. However, in a distress phase situation, there is a high chance that an emotional state might lead to the occurrence of a tragedy, including accident (vehicle or fire), person slumping, suicide, murder, kidnapping or abducting a person scenario, if not thwarted by offering immediate and useful assistance to the potential emotional imbalance victim, might result in a distressed or casualty situation. Thus, human emotion recognition, human activity recognition and location information are three broad domains that could be explored separately or holistically in order to provide an emergency assistance service to people in challenging and critical situations.

The novel idea of deploying a combination of data and knowledge driven approaches to help address this situation, could prove useful in a distress phase situation. This could be achieved through the modelling of human emotion and activity recognitions, including the harvesting of physiological data, ambulatory data, location and environmental state information of a person in a distress phase and offering a real-time triggering of rescue teams to pro-actively thwart a potential event that could lead to distress.

The Database for Emotion Analysis using Physiological Signal (DEAP) (Koelstra et al. 2012) emotional dataset is used for the experimentations conducted in this study while

applying our proposed methodology. The DEAP dataset was developed by Koelstra et al. (2012) using video clips stimuli to elicit human emotions from 32 subjects (16 females) and their physiological data such as the Electroencephalogram (EEG), Electro-oculogram (EOG), Electromyogram (EMG), Galvanic Skin Response (GSR), Respiration (RESP), Blood Volume Pulse (BVP) and Temperature (TEMP) were concurrently collected as they watched 40 one–minute extracts of music video clips. These clips are capable of eliciting the target or reported felt emotions of anger, contempt, disgust, elation, envy, fear, guilt, hope, interest, joy, pride, relief, sadness, satisfaction, shame and surprise. The physiological signals as well as frontal face videos of 22 subjects were acquired using various sensors and active electrodes with the Biosemi Active II system.

Thus, the goal of this research is to apply digital image processing based techniques to extract features from emotion physiological signals of DEAP (Koelstra et al. 2012) data set and compare performance with other techniques and extracted features in the literature as feature extraction is still an open issue in pattern recognition studies (Nweke et al. 2018). This is because; emotion recognition performance obtained relies heavily on features extracted. Thus, the quality of the discriminating features extracted and recognition performance obtained will further enhance efficient communication and decision making using critical body signals collected from a person in a distress phase situation and a prototype affective system.

The proposed system can practically benefit security agencies, emergency services and rescue teams, and restore confidence to both the potential distressed victims and their families by significantly reducing casualties in an event of distress.

1.1 Background

Some of the real life challenges, including the context within which this research study can be proactively situated and applied, are snappishly discussed below bearing in mind that the required emotion physiological signals to be harvested for human emotion recognition are especially of the victims of an act capable of causing accident and tragedies that could trigger a distress phase or casualty situation.

Keeping illegal custody of, taking away or holding, a person against his/her will is a serious crime anywhere in the world. The practice of kidnapping/abduction, terrorism, human trafficking and piracy, which are all direct violations of a victim's human rights, is however gaining tremendous ground across the globe. Nigeria and South Africa are indeed emerging as countries with the highest prevalence in Africa, according to Moor (2008) with the rise of kidnappings in South Africa stemming from the explosive growth in crimes since the 1990s. Nigeria and West Africa ranked number four among the top ten kidnapping countries in the world (Maldwyn 2012). The abduction of more than 270 school girls in Nigeria (Curry and Hughes 2014) readily comes to mind here, while the figures from the South African Police indicate that the country has a critical contact and other contact related crime challenges (Harrendorf, Heiskanen and Malby 2010; SAPS 2017). Contact crimes are committed against a person and it involves direct physical contact between the victim(s) and the perpetrator(s). Abduction is one of such examples.

From the Middle East - Afghanistan, Iraq, Pakistan; to Asia - China, India; Africa - Nigeria, South Africa; and America - Mexico, Argentina; kidnapping or abduction and terrorism crimes are now a global phenomenon (Moor 2008; Vannini, Detotto and Mccannon 2015; NYA 2016; Zen 2018). This is challenging the freedom, security and safety of people, infrastructures, political stability and service delivery across the globe with the tendencies of the crimes spreading across various nations' borders coupled with the attendant negative and disruptive impacts, thereby necessitating cooperation among nations.

These crimes are occurring as they are being investigated, and it was estimated in 1999 by Hiscox Group, an American based insurance company that 92% of abduction crimes takes place in just ten countries of the world. The group also recorded in December 2013, a 13.1% growth in kidnappings and ransom premiums to 6% of the group's total control income for the year 2013 coming from kidnapping and ransom insurance covers (Hiscox 2014).

Many factors, ranging from financial, political and religious motives, are responsible for the increasing rates of these crimes and the direct negative impact on the

society is huge, challenging the security, safety and lifestyles of people who are the main drivers of the world activities such as sports, agriculture, economy, information and communication, education, and transportation among others. These crimes diminish peoples' sense of security and could result in a brain drain arising from the victims relocating abroad for perceived safety reasons in order to avoid living through the terrifying experience. Kidnapping for Ransom (KFR) is a fall out of the financial aspect of these crimes, as it is an avenue for a quick, lucrative and a low risk source of income for the perpetrators while also serving as a source of terrorism funding across the globe.

KFR is now considered the most significant terrorist financing threat today (Cohen 2012) and according to Hiscox Group – a main purveyor of KFR insurance, about $310 million per year as estimated, is taken out as KFR insurance coverage worldwide (Hiscox 2014). Countries that are significantly susceptible to crises and liable to internal and external shocks or find it extremely difficult performing its basic security and developmental functions exhibit the characteristics of a fragile/failed state. Such (fragile/failed) states witnessing social unrests, disorganized legal systems and government instability suffer a high risk of occurrence of these crimes, as the political objectives of the crimes emanate from kidnappings or abductions of foreigners as a weapon of putting pressure on foreign governments for a specific demand to be met.

Other avoidable tragedies of global proportion and attention are vehicular, flood and fire related accidents. As stated by the World Health Organization (WHO 2015) every year, about 1.3 million people die in road crashes with an average of 3,287 deaths per day. From the records, someone is being killed every 30 seconds while 19 others are injured on roadways worldwide (WHO 2015).

An additional 20 million are disabled or injured per year, while more than 90% of all road traffic fatalities occur in middle and low-income countries, including Nigeria and South Africa. These deaths are, however preventable. It was further stated that, unless urgent action is taken to prevent and reduce road traffic deaths through injuries sustained, by prompt rescue and emergency services efforts to people in the identified distress phase, it was predicted that road traffic crashes would become the fifth leading cause of

death by the year 2030. Of significant interest are lone vehicular accidents not witnessed and where no information is known of its occurrence by anyone and as such emergency services could not be prompted in this distress phase scenario to save lives.

Fire outbreak in residences and workplaces is also a global health problem. As stated by the World Health Organization (WHO 2014) fact sheet, burns resulting from such outbreaks account for about 265,000 global deaths every year. Low and middle-income countries, including South Africa and Nigeria, constitute where the majority of these deaths occur while about half ensue in the South-East Asia Region according to WHO. In addition, nonfatal burns cause deaths. Prolonged hospitalization treatment, disability, scars and impairments, often resulting in stigma and societal rejection characterisation. It is noted that most of the victims are sometimes trapped indoors when the accident occurs and due to a lack of detailed information regarding the number of occupiers, as a result of the state of confusion with everyone scampering to his or her own safety first, the absence of a well ordered and coordinated procedural rescue effort, might often result in more casualties than expected. Proactively harvesting physiological signal information vis–a-vis the emotional state as well as the location and activity information of a trapped victim could prevent these deaths.

In South Africa, for instance, about US$ 26 million is estimated to be spent annually for the care of burns sustained from kerosene (paraffin) cook stove related mishaps according to World Health Organization (WHO 2014) report. Indirect financial costs of lost man-hour wages, protracted care for disfigurements and emotional trauma, and spending of meagre financial resources of the families is also among the socioeconomic impacts of fire accidents.

Examining distress state concept and how it is related to a process flow involving human emotions is therefore essential in order for a system to determine how casualties can be averted. The distress is a state of extreme necessity, the condition of being in need of immediate assistance. It causes pain, strain, anxiety, suffering, trouble and severe danger. A person in distress state is therefore in need of immediate rescue to avoid the severe unpleasant experience and feelings capable of leading to loss of lives,

limbs, properties and valuables. More often than not, the type, manner and urgency with which the assistance is sought and rendered could determine the magnitude of the damage caused or averted.

However, before a distress phase situation occurs, it is usually preceded by series of spontaneous human emotions and actions/activities that could trigger the seeking of assistance in order to avert the danger. Such human emotions could include fear, anger, contempt, disgust and sadness. Human actions in the form of ambulatory activities might include walking, shouting, running, pulling, pushing, struggling and falling, which are all critical signs of an impending disaster in a distress situation scenario. All these are occurring within a distress phase situation and if assistance is rendered, probably by way of rescue or emergency efforts, then the distress chain shown in Figure 1.1 as a linear transition in a distress phase process flow is aborted from progressing to casualty phase such that lives, properties and valuables are protected from the looming threat.

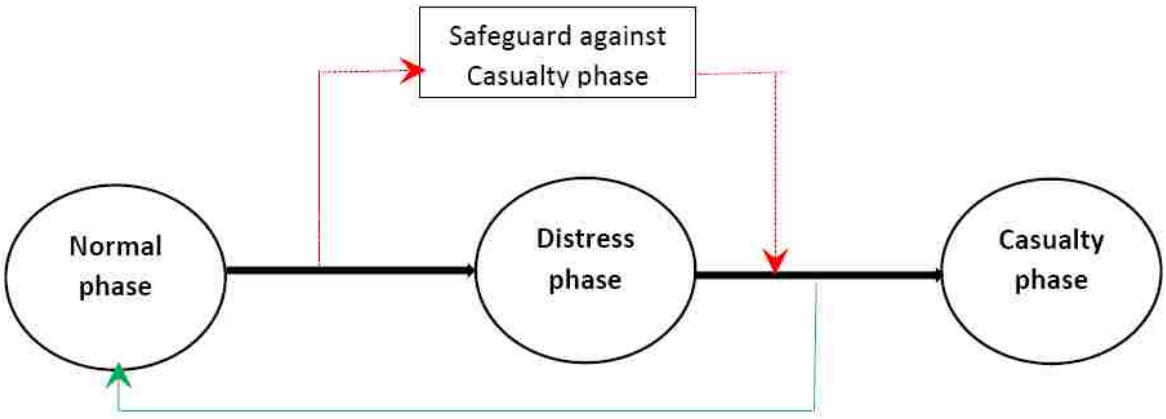


Figure 1. 1: Linear transition process flow from normal/happy to distress, to casualty phase.

A distress phase, as defined in this study, implies situations where a person, who might fall into an imminent danger as a result of his/her own emotional state or activity or is threatened by an external imminent or severe danger which will also definitely impact his emotional state or human activity and thus requires immediate assistance. The tuple – critical/severe, imminent danger and immediate assistance characterizes a distress phase situation.

1.2 Research Problem

New emerging technologies such as mobile devices, wearable sensors, social media, emails, location aware devices and services, digital images, smart phones, records retention and the internet, as offered by Information Communication Technologies (ICT), are now an integral part of everyday living and shaping our lifestyles. These technologies have made it feasible for collection and distribution of very large data set including physiological signals for machine learning. Several pattern recognition problems, including human emotion recognition, human activity recognition, image classification, location, and object detection have therefore emerged over the past few decades. Pattern recognitions and machine learning solutions particularly to human emotion recognition problems thus have direct impact on real-life applications including security/surveillance, Smart Home (SH), computer games, intelligent tutoring, telemedicine and psychopathology, depression detector and management, multimedia sector, social and emotion development research, safe driving, mother-infant interaction, customer services, call centre s, fitness tracking, affective computing and human computer interaction applications (Wagner, Jonghwa and Andre 2005; Fried 1976; Cohn and Tronick 1988; Ekman, Matsumoto and Friesen 1997; Larson and Rodriguez 1999; Roisman, Tsai and Chiang 2004; Ekman and Rosenberg 2005).

Emotion is a complex state of human mind. It is influenced by body physiological changes and interdependent external events thus making an automatic recognition of emotional state a challenging task. In the literature, human emotions have been measured and recognised using physiological signals, audiovisual methods of facial expression, gestures and speech (Wagner, Jonghwa and Andre 2005; Chanel et al. 2006; Koelstra et al. 2012; Noppadon, Setha and Pasin 2013a). Though, in the recent past, the physiological signals based human emotion recognition has received relatively less attention in comparison to the other stated methods, however, researchers are now trending in the physiological signals based human emotion recognition domain because of the emerging technologies earlier stated above and the availability of mobile devices,

affordable wearable sensors as well as equipment that enable the seamless collection of physiological signals. Meanwhile, the audiovisual methods have been criticized due to some of their intrinsic drawbacks. They are capable of being easily faked and the subject needs to be within a perimeter defined by the camera or must always listen to an audio signal. Conversely, physiological signals evolve automatically and spontaneously. They are human reactions over which they have less controls and are less influenced by social, language and cultural differences (Chanel et al. 2006; Eun-Hye et al. 2012; Heng et al. 2013; Noppadon, Setha and Pasin 2013a). These factors, among others informed the decision to adopt the physiological signals based emotion recognition system for the research reported in this dissertation.

A number of recognition methods have been applied in recent years to recognize human emotion as the trend towards making significant contributions now lies around features extracted as well as the methods applied. However, while noticeable, but varied recognition accuracy results were recorded with these methods (Wagner, Jonghwa and Andre 2005; Koelstra et al. 2012; Maaoui and Pruski 2010; Noppadon, Setha and Pasin 2013a), part of the unresolved issues in mobile and wearable sensor-based pattern recognition domain is the extraction of relevant discriminating features (Nweke et al. 2018) which has also been identified as one of the most daunting aspects of physiological signals based human emotion recognition system (Maaoui and Pruski 2010; Noppadon, Setha and Pasin 2013a). In addition, opinion still varies about which set of emotion features and recognition methods would give the best result of an emotion recognition system while it also remains an open challenge in affective computing research to fix an agreed recognition accuracy for an affective recognition system (Jerritta et al. 2011; Noppadon, Setha and Pasin 2013a). This is because of the wide disparities in the number of emotions to be recognised, the number and types of bio-signals measured, data set used and its quality, number of subjects sampled, emotional stimulus, modality considered, emotion models employed, pattern recognizers used and features extracted from physiological signals among others (Chanel et al. 2009; Jerritta et al. 2011; Heng et al. 2013; Noppadon, Setha and Pasin 2013a).

Therefore, this study attempt to experimentally discover discriminating emotion features on which a recognition method can be applied, using the physiological signals

harvested from individuals while experiencing an emotion as contained in the DEAP (Koelstra et al. 2012) dataset, to build an efficient emotion recognizer model that can be embedded in an affective system towards a real-time tracking of an individual in distress phase situations.

The techniques of deep learning neural networks were jettisoned for the experimentations conducted in this study because during extensive literature review of classification methods and techniques, the results obtained by (Yin et al. 2017; Wang and Shang 2013; Li et al. 2015; Jirayucharoensak, Pan-Ngum and Israsena 2014) with deep learning approaches using the DEAP (Koelstra et al. 2012) data set that was also utilised in this study can definitely be improved upon. This is despite the deep learning approach's strengths in automatically extracting features from raw data as well as learning from labelled or unlabeled data (Nweke et al. 2018). This confirms the weak quality of discriminating features engineered by the deep learning approaches reported in (Yin et al. 2017; Wang and Shang 2013; Li et al. 2015; Jirayucharoensak, Pan-Ngum and Israsena 2014) and the need for new discriminatory features discovery as carried out in this study.

In addition, deep learning methods can also be computationally intensive because of the required high parameter initialization, tuning and update (Nweke et al. 2018) as well as a higher number of layers which can run to tens or hundreds of successive layers of representation (Chollet and Allaire 2017) rather than just the only one or two layers of representation available in other classical shallow machine learning and pattern recognition algorithms such as Support Vector Machine (SVM), k-nearest neighbour and variants of Multilayer Perceptron Artificial Neural Networks (MLP-ANN) and the Radial Basis Function Neural Networks (RBFNN). Since, the layered representation in deep learning via models that are called neural networks (Chollet and Allaire 2017), this also guided this study's choice of a neural network pattern recognition algorithm called the RBFNN, which is a shallow learning approach with an input layer, a single hidden layer and an output layer but with the capability of yielding good performance.

1.3 Study Aim and Objectives

The major aim of this research is to apply a machine learning approach to realize an intelligent system that can recognize human emotional state, which could be used to track a person in an emotionally induced distress phase situation.

The under-listed research objectives are structured in order to achieve the aim of this research study:

(i) To discover a set of physiological properties that are suitable for identifying an individual's emotional state.
(ii) To explore existing methods that could help to develop an intelligent system to identify an individual's emotional state.
(iii) To enhance detection of and matching between body signals and an individual's emotional state.
(iv) To test the reliability of the automated identification of an individual's emotional state by the intelligent system using data from an existing database.

1.4 Research Questions

Flowing from the aim and objectives as well as the problem statement of this study, the following research questions were investigated:

i. What physiological properties are suitable for identifying an individual's emotional state?
ii. How are existing methods used to develop an intelligent system to identify an individual's emotional state?
iii. How can detection of and matching between body signals and an individual's emotional state be enhanced?

iv. How reliable in terms of recognition performance achieved, is an automated system of identification of an individual's emotional state using data from existing databases?

The relevant distress phase scenarios where an intelligent system like this could be deployed include abductions, fire and vehicular accidents, earthquakes and other natural disasters, (violent) rape, collapsing of buildings and missing person scenarios.

1.5 Rationale for the Research

As a result of the ever-growing world population and the strain on resources to cope with this growth, there is a need to have emergency plans in place to mitigate the effect of disasters (fire, flood, abduction, traffic accident, terrorism) on people and their valuables. A proposed system relying on human emotional states to determine an individual in a distress phase scenario such that emergency services could be triggered is therefore crucial for prompting a coordinated, effective and efficient real-time response to avert the occurrence of casualties or salvage the disaster when it occurs. The real-life purpose in application areas of such systems is primarily to save lives, valuables and persons suffering.

This research work is therefore of tremendous significance in view of the need to enhance safety by providing automated surveillance/tracking of individuals even in public places (Hassan et al. 2018) using emotion physiological signals as well as detecting emotional states capable of threatening the safety and behaviour of the individual or other people. This can be used to either detect the occurrence or predict crimes that could be happening in the near future (Hassan et al. 2018). It is also essential to note the increasing global crime rates leading to loss of lives, revenues and valuables among others. The continuous devising of means by the perpetrators to stop anyone that wants to prevent them from committing crimes through raising their expertise while taking note of the important roles that data and signals play in crime prevention, detection and investigation hence the focus of this study on physiological signals and emotions. This is because

physiological signals are described as human reactions that evolve automatically and are less influenced by social, language and cultural differences (Chanel et al. 2006; Eun-Hye et al. 2012; Heng et al. 2013; Noppadon, Setha and Pasin 2013a) while emotions are also present in every individual (Al-Shawaf et al. 2016).

The other importance of the research include enhancing real time communications via data signals collection between an individual – a potential perpetrator or victim and the proposed remote tracking system. In addition, it includes providing logistics and support information to security agencies and rescue teams as well as the victims' family toward drastically reducing the scourge of some identified crimes. In addition, it involves promoting individual's confidence in terms of safety and reducing brain drain because of accidents and crimes while also managing the emotional stress of the victim's family members caught in the distress phase and casualty phase scenarios with a broad view of preventing casualties.

The other motivation for this research is the open and unresolved issues of feature engineering in pattern recognition and machine learning problems especially as the trending deep learning approaches in machine learning classification problems which are designed to automatically extract salient discriminatory features from raw data without relying on strenuously handcrafted features (Nweke et al. 2018) but are yielding results (Yin et al. 2017; Wang and Shang 2013; Li et al. 2015; Jirayucharoensak, Pan-Ngum and Israsena 2014) that could be improved upon, even with the DEAP(Koelstra et al. 2012) data set that is utilised in this study.

1.6 Study Contributions

An intensive literature review of Human Emotion Recognition (HER) through physiological signals is performed in this study. After this exercise, the distinctive contributions of this study are enumerated as follows:

(a) The first apparent unique contribution of this study in comparison to the previous research is the representation of human emotion physiological signals as hyperspectral images through inverse Fisher transformation. The hyperspectral

imaging takes dozens to several contiguous narrow wave band images and have capabilities for vast quantities of data because of these high numbers of bands that are simultaneously imaged (Yuen and Richardson 2010). The hyperspectral image pixels in this context are vectors representing the spectral characteristics of the physiological signals in each channel. Because of its strong discriminatory ability between materials, though hyperspectral imaging technique was originally developed for geological and mining applications, it has since been extended to other areas including material identification, anomaly detection, target detection and recognition of target patterns (Yuen and Richardson 2010). The hyperspectral imaging techniques have also been used for remote sensing and detection of physical and emotional stress using facial expressions (Yuen and Richardson 2010). This is indeed a novel idea for counterterrorism operations, a notion that was leveraged on in this study by using physiological signals because they are auto-evolving and less influenced by humans, cultures and languages rather than face (audiovisual) expressions because of their inherent faking drawback (Chanel et al. 2006; Eun-Hye et al. 2012; Heng et al. 2013; Noppadon, Setha and Pasin 2013a). It is also of critical importance to note that terrorism and other related crimes are capable of being detected through human behaviours, activity, gestures and facial expressions (Yuen and Richardson 2010) while physiological signals can also be employed to assess the intent of an individual including perpetrator and the victim of a crime.

(b) The second unique contribution of this study is the novel discovery of highly intra-class image similarity and inter-class image dissimilarity features based on state of the art inverse Fisher transform (IFT), histogram of oriented gradient descriptors, local binary pattern descriptor, histogram of images and Eigen vector decomposition. Feature discovery remains an important task in image processing, computer vision and machine learning researches (Nweke et al. 2018; Wang et al. 2018). Through this novelty, nine new features have been discovered for human emotion recognition. These features are listed under each modality which include; (i) Peripheral physiological signal modality: Histogram of Oriented Gradient

PEripheral Physiological Signal (HOGPEPS), Local Binary Pattern PEripheral Physiological Signal (LBPPEPS), Histogram of Images PEripheral Physiological Signal (HIMPEPS); (ii) Electroencephalogram (EEG) modality: Histogram of Oriented Gradient Physiological Signal (HOGPS), Local Binary Pattern Physiological Signal (LBPPS), Histogram of Images Physiological Signal (HIMPS); and (iii) Fused (EEG+Peripherals) modality: HOG Human Emotion Signal (HOGHES), LBP Human Emotion Signal (LBPHES) and Histogram of Images Human Emotion Signal (HIMHES) respectively.

(c) The application of the newly discovered human emotion physiological signals to recognize human emotion with a binary output variable based on two-class classification problems mapped under the Valence, Arousal, Dominance and Liking emotion representations in the DEAP (Koelstra et al. 2012) corpus using the Radial Basis Function Neural Network (RBFNN) machine learning algorithm. A three-class classification problem based on a novel modelling of emotion to suit the stated distress phase state in terms of the above mentioned four representations of emotions was also designed.

(d) A qualitative and quantitative evaluation of the inverse Fisher transforms algorithm with which the discriminating features were extracted was carried out. The image representations of the different channels and modalities after applying the feature descriptors were presented to qualitatively analyse and compare the inherent differences and similarities. The quantitative evaluation is realised with the performance metrics of accuracy and mean square error obtained for each of the features and the three modalities considered.

This study builds on the use of an existing pattern recognition algorithm - the RFBNN and provides answers to how effective the algorithm could be in recognition of human emotion in a distress phase state with a view to specifically develop an intelligent system towards a real-life tracking of an individual. The research also explores and investigate the use of emerging technologies for deploying a pervasive emotion recognition/tracking and activity monitoring service.

1.7 Research outputs

This research study has led to the publications of the articles listed below:

A. Abayomi, O.O. Olugbara, D. Heukelman and E. Adetiba. 2019. Physiological Signals Based Automobile Drivers' Stress Levels Detection Using Shape and Texture Feature Descriptors: An Experimental Study. *In: J. Mizera-Pietraszko et al. (eds.), Lecture Notes in Real-Time Intelligent Systems: RTIS 2017, AISC 756, 436-447.*

A. Abayomi, O.O. Olugbara and D. Heukelman. 2018. An Architecture Utilizing Human Emotions and Activities Recognition for Remote Monitoring. *In IEEE Xplore, 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD, August 2018)*.

A. Abayomi, O.O. Olugbara, Delene Heukelman. 2018. Prevention of Road Traffic Accidents using Physiological Signals to Detect Automobile Drivers' Emotion. Abstract proceedings of the 3rd Interdisciplinary Research and Innovation Conference (IRIC 2018), Durban. South Africa.

A. Abayomi, O.O. Olugbara, E. Adetiba and D. Heukelman. 2016. Training Pattern Classifiers with Physiological Cepstral Features to Recognize Human Emotion. *In: Pillay N., Engelbrecht A., Abraham A., du Plessis M., Snášel V., Muda A. (eds.), Advances in Nature and Biologically Inspired Computing. Advances in Intelligent Systems and Computing, vol. 419. Springer, Cham.*

The following journal manuscripts are being edited for publication:

A. Abayomi, O.O. Olugbara and D. Heukelman. 2019. Recognition of Human Emotion using Radial Basis Function Neural Networks of Inverse Fisher Transformed Physiological Signals.

A. Abayomi, O.O. Olugbara and D. Heukelman. 2019. Recognition of Distress Phase Situation using Inverse Fisher Transformed Human Physiological Emotion Signals with Radial Basis Function Neural Networks.

1.8 Synopsis

The outline of this dissertation is as follows.

Chapter 1 of this thesis covers the introduction, background of the study as well as the research problems to be solved, including some previous works, what have been addressed and what is left unresolved in the field of human emotion recognition, especially as it relates monitoring persons in distress phase while its link to human activity recognition and location information is also explored.

In Chapter 2, relevant literature and related works about this study: Towards real-time tracking of persons in distress phase situations is presented. Emotion recognition studies and how it can be integrated with Human Activity Recognition (HAR) as well as location information and sensors were thoroughly explored.

The components of emergency response systems and management is presented in Chapter 3. An architecture for combining human emotions and human activities recognition is proposed. Various signals that could be measured for human emotions and activities are explored.

The mathematical models and design of the proposed feature extraction algorithm based on the inverse Fisher transform is presented in Chapter 4.

In Chapter 5, the human emotion recognition methodology and designs of experimentations conducted as well as the application software coding are discussed. Several feature extractions techniques such as the histogram of oriented gradient, local binary patterns and histogram of image features of physiological signals relying on the designed inverse Fisher based data transformation algorithm were applied to the DEAP (Koelstra et al. 2012) emotion physiological signals for analysis. The Gaussian radial basis function artificial neural network pattern recognition algorithm was used for the various experiments conducted and reported in this thesis.

The experimental results, evaluation metric, interpretations of results as well as discussions are presented in Chapter 6 of this thesis while the study's summary, conclusions, recommendations, limitations and future works are presented in Chapter 7.

1.9 Research Scope

This study only covers the use of emotion laced physiological signals harvested from individuals as contained in the DEAP (Koelstra et al. 2012) data set to detect human emotion. This data set which is renowned in human emotion recognition and affective computing domains was used to detect and classify emotions into two/three classes to reflect the designed distress phase emotion model using the four emotion representations of valence, arousal, dominance and liking. No emotion physiological signals were directly collected in this study and human activities recognition experiments were not carried out. Only an architecture that could combine human emotion, human activities and location information using various sensors was presented for effective emotional tracking of an experiment subject especially in a distress phase situation.

CHAPTER TWO

Related Works

This chapter offers a synopsis of the latest topics, avant-garde practices, challenges and issues in emotion recognition studies using wearable sensors with a view to tracking an individual in a distress phase situation. Researches in activity recognition and location detection domains are also discussed with their capabilities to be combined with human emotions for tracking of people in distress phase scenarios.

2.1 Emotions Context

Emotion is a very complex, somewhat intangible human state and a challenging task to define as a lot of definitions have been proposed. This is supported by the famous coinage by Beverley and Russell (1984) that “everyone knows what an emotion is until asked to give a definition. Then, it seems no one knows”. Consensus on the various definitions due to debates among contemporary theoreticians and researchers on the best way to theorize emotion as well as interpreting its roles in human life, is therefore not currently to be found.

Efforts to provide definitions of emotion started as far back as the era of the early philosophers Plato and Aristotle as well as psychologists McDougall, Wundt and James. Emotion has been considered as a mental event (Wundt 1924), behaviour (Watson 1919), physiological activities triggered by the autonomic nervous system (Wenger 1950) and also as a group of muscular and glandular reactions or facial behaviours (Tomkins 1980).

However, many psychologists have dismissed the idea of stringently confining emotion definition to only mental, only behavioural or only physiological activities. They argued the possibility of occurrence of certain emotions without distinct behavioural or physiological signs while these signs can also occur in other activities not linked to emotion such as in the workout and acting. This necessitated the merging by Izard (1972) of mental, physiological and behavioural events into a single definition of emotion and proposed as “a complex process that has neurophysiological, motor-expressive and phenomenological aspects”.

It has been stated that emotion involves behaviour, feelings, physiological change as well as cognitions and always occurs within a particular context, which powers it. Its major function is to offer support and information for decision making to the individual with respect to interaction with the ecosphere. The behavioural aspect of this definition includes vocal/verbal, gestures, facial and postural responses of the individual. On the other hand, the feeling component comprises the glaring changes observed in an emotionally aroused person and are readily associated with an emotional drive, which stirs the physiological change component (Heng et al. 2013). The nervous system is activated in the physiological change component as electrochemical activities are triggered for a fight or flight response.

Emotion is part and parcel of human's everyday living and constitute an essential part of his/her survival and existence (Izard 1972; Izard 1977). Emotional feelings and expressions are very germane to the enhancement and regulation of human interpersonal relationships. This is because (verbal, facial and gesture) communications between people are often laced with anticipated changes in the emotional state of the parties either to express agreements or disagreements with opinions or to clarify spoken words or gestures. Human beings, therefore continuously express and perceive emotions and do not switch off their minds as revealed by the brainwave activities obtained from the scalp.

The terms "affect", "emotions" and "moods" have often been misused interchangeably by people. Affect is a general term representing diverse feelings that people experience and it incorporates both emotions and moods. Emotions, on the other hand consists of passionate feelings aimed at an object or individual. Moods are less passionate feelings than emotions and most frequently devoid a contextual stimulus. According to Ekman and Davidson (1994), moods and emotions are distinct from each other in terms of their duration of occurrence or time course. Moods usually last longer than emotions and could take hours or days while emotions can be very brief, occurring within a second or at most minutes (Ekman 1984). It has also been posited by Ekman and Davidson (1994) that a mood state that stands for weeks or months is an affective disorder and no longer a mood.

Emotions (such as happy, fear and disgust) are associated with unique facial expressions while neither affective disorder, moods, irritability nor emotional traits have facial expressions mapped to them.

2.2 Theories of Emotion

Different theories of emotion can be appositely distinguished by their choice of emotion definition and interpretations concerning the cognitive, expressive, feelings, physiological and motivational components of emotions. Some of the known theories of emotion are described below.

2.2.1 Evolutionary Theories

The evolutionary theories describe the historical background relating to the evolution of human emotions by considering natural selection that occurred in the past and its nexus with the manifestation of emotions in humans today (Johnson 2009). Changes in traits across generational time and space are considered; as genetic drift, natural selection and chance are possible candidates of the traits' changes. Natural selection triggered traits are called adaptation and Richardson (1996) opined that adaptation is traits' prevalence that conferred a greater fitness, while others have argued that fitness could be conferred by a trait without necessarily making the trait an adaptation. Despite this argument, many theories have emerged that describe emotions as adaptations. Keltner, Haidt and Shiota (2006) stated that emotions are well-organized and synchronized human reactions that aid reproduction, avoid physical threats and embrace survival, offer protection to young ones and maintain cooperative associations with others, thus conferring the seals of adaptation.

2.2.2 Social theories

This approach views emotion as social constructs, which are derived from human cultures, traditions and societies and are acquired by people through experience.

Some of the ideas that inspired this approach include;

(i) Different words are used for emotions across cultures and languages. Since people experience an emotion, for which they have words, different emotions are in existence and experienced across cultures. For example, there is no precise translation in Polish for the emotion of disgust and the emotion *amae,* representing affective dependency upon another's love, only exists among the Japanese. It is analogous to infants' feelings toward their mothers; however, this emotion also exists in adults (Morsbach and Tyler 1986).

(ii) Emotions are present during interpersonal and social relationships, but are rather interactions between people and their environments. Interpersonal factors are thus what triggered emotion and either make humans participate or withdraw from specific interpersonal contacts.

(iii) Societal norms, practices, values and beliefs do regulate emotions and contribute largely to which events lead to what emotion and how that emotion is expressed.

2.2.3 Transitory Social Roles

Emotions have been described here as a socially constituted syndrome involving human's evaluation of a situation such that it is construed as a passion rather than an action (Averill 1980). The syndrome as well as transitory social roles is produced by beliefs and societal norms, which governs human emotions. Syndromes are aggregation of all emotional responses for a particular emotion, but none of them is essential for that emotion syndrome (Averill 1980). All elements, including the eliciting stimuli and other non-social elements contributing to an emotional response are included in the syndrome.

2.2.4 Emotion Process Theories

This approach is centred on the emotion process proper. The process includes stimuli elicitation and perception, which may include thoughts or recalling, which is followed by activities between the perception and generation of a body response, while the last stages are the facial and physiological responses, such as increased heartbeat, skin temperature, skin conductivity, respiration etc. The stimuli elicitation and perception, which is considered in an emotion process as the early portion, is very germane, because the nature of the emotion experienced and expressed by humans, is determined at this stage.

2.3 Emotions Recognition Research

Various researchers in diverse scientific fields, including affective computing, psychiatry, speech analysis, computer vision, biology, linguistics, sociology, neuroscience, anthropology, psychology and information and communication technology have crisply acknowledged more than 300 emotions. Nevertheless, not all these emotions are experienced in human day-to-day life and researchers have tried to narrow the number of emotions down to six basic emotions, introduced by Ekman (1982) and argued, in support of Palette theory, that “any other emotion is the composition of the six basic emotions.

However, the emotion types that have been recognised with the various elicitation stimuli include: amusement, anger, annoyance, anxiety, boredom, calm, confusion, contempt, contentment, disbelief, disgust, distress, elation, embarrassment, emphatic, engagement, fatigue, fear, frustration, grief, happiness, hate, helpless, hot/cold anger, interest, ironic, joy, laughter, motherese, neutral, relaxed, sadness, shame, stress, surprise, panic, platonic love, pleasure, pride, puzzlement, rebelliousness, reprimanding, rest, reverence, romantic love, touchy, uncertainty. Classes of arousal – low, medium and high; and valence – positive, negative, pleased and displeased (Chanel et al. 2009; Zhihong et al. 2009; Jerritta et al. 2011; Soleymani et al. 2012) have also been identified.

Studies using physiological signals to recognize emotional states have been on the rise in the past one decade. Some of these research studies are described below and the improvement of accuracy using different combinations of signals and methods can be seen.

2.3.1 Physiological Signals

Six emotions, namely amusement, contentment, disgust, fear, neutral and sadness were classified with an accuracy of 90% and 92%, using the Support Vector machine (SVM) and Linear Discriminant Analysis (LDS) classifiers respectively, along the subject dependent approach as reported by Maaoui and Pruski (2010). Time domain statistical features such as mean, standard deviation of raw signals, absolute values of first and second differences of raw signals were extracted from Blood Volume Pulse (BVP), Electromyography (EMG), skin conductance, respiration rate and skin temperature physiological signals, acquired from subjects induced by the International Affective Picture System (IAPS) stimuli.

The study reported by the Soleymani et al. (2012) utilised physiological signals, including respiration amplitude, skin temperature, Galvanic Skin Response (GSR) and Electrocardiograph (ECG), harvested from experiment, subjects induced with video clips to classify emotion into the arousal (calm, medium and excited) and valence (unpleasant, neutral and pleasant) states. Time domain statistical features, including average, standard deviation of raw signals, as well as the absolute values of the first and second derivative, were extracted and trained with an SVM classifier. A classification accuracy of 46.2% was obtained in the arousal state while 45.5% was recorded for the valence state both of them using a subject independent approach. The study was also extended to using Electroencephalography (EEG) signals from where Power Spectral Density (PSD) and spectral power asymmetry features were extracted and trained using an SVM classifier. Using the subject independent approach, a classification accuracy of 52.4% and 57.0% was obtained from the three states of arousal and valence respectively.

This study did not meet the required level of accuracy to be implemented in a real-time environment. However, Noppadon, Setha and Pasin (2013b) extracted PSD features from EEG signals acquired from pictures induced experimental subjects and obtained a classification accuracy result of 85.4% along two valence states with an SVM classifier and subject independent approach. This result was much more promising.

In the research study conducted by Mimma et al. (2015), Heart Rate Variability (HRV) derived from ECG physiological signals acquired from IADS stimulated healthy volunteers were utilised in recognizing emotional state along four and two classes of arousal and valence respectively. The Autonomic Nervous System (ANS) dynamics estimated through the standard and nonlinear analysis as well as the Lagged Poincare Plots of the HRV were used as features. A leave-one-subject-out validation approach was employed and a quadratic discriminant classifier was applied to the extracted features. The arousal dimension gave a recognition accuracy of 84.2%, while the valence approach posted 84.7%.

Ateke, Ataollah and Atefeh (2016), who utilised images as stimuli to acquire Heart Rate Variability (HRV) physiological signals from college students, obtained the next level of accuracy. They extracted 17 standard and non-linear features from the intrinsic mode functions decomposition of the HRV signals. A probabilistic neural network algorithm was applied to the extracted features to classify emotions into four classes of fearfulness, happiness, sadness and peacefulness. A classification accuracy of 99.09% was obtained.

Jun-Wen et al. (2016) separated ten groups of emotional pictures separated into five classes based on the valence and arousal dimensions. Facial EMG of 113 experimental subjects consisting of young and senior adults stimulated with these pictures was acquired for each of the mapped classes and 16 sets of features relating to the frequency, amplitude, variability and predictability of the EMG signals were extracted. An SVM classifier was applied to these features and a classification accuracy ranging from 75.6 to 100% was obtained for the five affective classes and the baseline for all individuals.

2.3.2 Facial expressions and speech

In a similar vein, various studies have been conducted using facial expression and speeches for emotion recognition. The Cohn-Kanade corpus (Kanade, Cohn and Tian 2000) was utilised as elicitation material by Uroš and Božidar (2015) and extracted 1232 facial images of 106 experiments, subjects to classify human emotions into six classes consisting of anger, disgust, fear, happiness, sadness and surprise. The histogram of oriented gradient (HOG) difference features was extracted from these emotional facial expression images and obtained a recognition accuracy of 95.6%, with the subject independent approach based on the SVM classifier.

In another study, the seven emotions of angry, disgust, fear, happy, sad, surprise, neutral were classified by Yunan, Yali and Shengjin (2015) using the Japanese Female Facial Expression Database (JAFFE) corpus as stimuli and the fusion of HOG and uniform local binary patterns (uLBP) as features. Coarse-to-fine classifiers were applied to the extracted features and a classification accuracy of 95.3% was achieved.

The Linear Prediction Coefficients (LPC), Mel Frequency Cepstral Coefficient (MFCC), Linear Prediction Cepstral Coefficient (LPCC) and Perceptual Linear Prediction (PLP) were extracted from speech signals by Palo, Mohanty and Chandra (2015). The multilayer perceptron artificial neural network classifier algorithm was applied to the features and recognition accuracies of 80.0%, 48.6%, 54.5% and 70.0% respectively were obtained for four emotion classes namely angry, bore, sad and surprise.

2.4 Applications of Emotions

The research niche of human emotion is fundamentally a multidisciplinary domain comprising assorted fields such as linguistics, speech analysis, medicine, human computer interaction, business management, psychology, behavioural science, computer vision, psychiatry, marketing, safe transportation, advertising and security, as well as machine intelligence.

An automatic emotion recognition tool can be successfully deployed in various application domains, including social and emotion development research, multimedia sector, depression detector and management, customer services, tutoring and learning, telemedicine and psychopathology, studies on affective expressions such as deception, safe driving and tracking of people's well-being and safety, mother-infant interaction and call centre s (Fried 1976; Cohn and Tronick 1988; Ekman, Matsumoto and Friesen 1997; Larson and Rodriguez 1999; Roisman, Tsai and Chiang 2004; Ekman and Rosenberg 2005; Jonghwa and Andre 2008).

2.5 Emotion Representations

As a result of the variations involved in how an individual experiences and expresses emotions, various approaches have been utilised to represent human emotions. These emotion representation models include the discrete and dimensional or continuous models.

2.5.1 Discrete Emotion Model

The discrete emotion model was inspired by the notion of considering human emotions as necessary for survival, as postulated by Charles Darwin. The naming of the emotions originated from the everyday usage of words by human beings in order to convey his/her emotional feelings. It is arguably the earliest model used by a psychologist (Ekman 1982) while it also takes into consideration the existence of some universal and basic emotions. These basic emotions include anger, sadness, fear, happiness, disgust and surprise (Ekman 1982). The universal characteristic of these basic emotions is supported by using facial expressions (Ekman 1982). It is believed that despite the racial or cultural differences among people, many people can still perceive emotions that correspond to a specific human facial expression. Other emotions mixes have also been argued to be derived from a blend of the core discrete emotions; sadness and surprise, for instance, can result in a different emotion namely disappointment, while happiness and contempt can result in smugness. Some of the characteristics that clearly distinguish the basic

emotions include distinctive universal signals, universal antecedent events, emotion specific physiology and automatic appraisal mechanisms (Ekman 1999).

The distinctive universal signals combine all the emotion evolution phases, including physiological changes and memories that occurred in a person, to bring an emotional expression and what is expected to follow next, including consequences and coping. Also, the appraisal mechanism is concerned with determining or measuring the stimuli that pertains to a specific emotion and this mechanism could be automatic or extended. On the other hand, the physiological characteristics provide information regarding the physiological changes in a human that prepares himself/herself to respond in different ways to the diverse emotions being experienced (Ekman 1999). It has been observed in the literature that unique patterns of Autonomic Nervous System (ANS) activity connect to the emotions of Disgust, Fear and Anger (Ekman, Davidson and Friesen 1990). The characteristics, the universal antecedent event are related to the presence of common elements in the perspectives in which emotions occur, while some other characteristics of basic emotions, such as quick onset of emotions, brief duration and their presence in other primates, have also been added by another author (Reeves 1993). It should also be noted that these characteristics serve as a guide and should not be considered as indispensable and essential conditions that must be present in all basic emotions.

The discrete model, despite its easy understandable, popularity and wide usage, is however challenged by the fact that there exists a variation in the naming of an experienced emotion across different cultures, because emotion words do not have precise translations in diverse languages and cultures. To support this fact Russell (1991) stated that the Disgust emotion, for instance, does not have a corresponding word in Polish language that can convey the meaning of the emotion experienced.

2.5.2 Dimensional Emotion Model

The dimensional model of emotion representation uses a continuous space to represent emotion, so as to cater for the inherent cross-lingual interpretation challenge of the

discrete model. It originated from the cognitive theories and utilises n-dimensional space to represent an emotional state (Hanjalic and Xu 2005). The most popular are the 2- or 3-dimensional spaces, in which the emotions experienced are mapped to the bipolar valence and arousal scales. The valence scale measures the pleasantness, pleasureness or otherwise of an emotion, while the activation level and intensity experienced in an emotion is captured by the arousal scale (Hanjalic and Xu 2005). The arousal level could be calm/low, average/medium and excited/high, whereas the negative (unpleasant), neutral and positive (pleasant) valence levels can be attributed to an experienced emotion (Uhrig et al. 2016). The sadness emotion, for example, has a low arousal but negative valence; fear has a negative valence but high arousal; surprise has a low arousal and positive valence, while happiness has positive valence but high arousal.

The third scale in the 3-dimensional space is called dominance, which measures the level of control an individual has over the felt emotion and ranges from submissive (without control) to dominant (empowered). The 3D space popularly named Pleasure-Arousal-Dominance (PAD) was described by Russell and Mehrabian (1977) and has also been recently validated by Iris et al. (2014) to replace the 2D space. A fourth dimension called predictability was also proposed by Fontaine et al. (2007) in addition to the PAD dimensions. The level of certainty or likelihood of occurrence of a sequence of event being watched by a subject is measured by the predictability dimension. It is essential to mention that though the dimensional model has the advantage of people being able to situate emotional content in comparison to a reference point, this does not make the model superior to the discrete model. Both emotion models are popular in emotion recognition studies and are transformable to each other, while their contributions and applicability differs with variabilities introduced by contexts, subjects, time and stimuli among others (Soleymani *et* al. 2014).

2.6 Emotional Stimuli Datasets

The elicitation of emotional responses from subjects in experiments is a very challenging task and requires the selection of the most effective stimuli to achieve a valid response.

In the course of eliciting an emotion in a subject under experimental conditions, many external stimuli have been used by different researchers, depending on the contexts and the subjects. These external stimuli include pictures, static images and vignettes, robot actions, voice, games, audiovisuals, films/movies/video clips, audio music, music videos, odour, self-elicitation and the recall paradigm (Lan and Ji-hua 2006; Noppadon, Setha and Pasin 2013a).

OASIS -The Open Affective Standardized Image Set (Benedek, Shayn and Mahzarin 2017), Geneva Affective PicturE Database (GAPED) (Dan-Glauser and Scherer 2011), International Affective Picture System (IAPS) (Lang, Bradley and Cuthbert 2008), International Affective Digitized Sounds (IADS) and the Affective Norms for English Words (ANEW) (Bradley and Lang, 1999a; Bradley and Lang 1999b; Bradley and Lang 2007) are among the popular stimuli datasets for emotion elicitation, while in other studies, subjects have been asked to personally obtain various stimulus materials capable of inducing a target emotional state in themselves by survey or otherwise (Panagiotis and Leontios 2010; Xu and Plataniotis 2012).

The OASIS stimulus data set (Benedek, Shayn and Mahzarin 2017) contains 900 images carefully organized under four themes: scenes, animals, humans and objects, along with normative scoring on the arousal and valence affective dimensions. It is an online open-access dataset, collected in the year 2015 and therefore contains recent images as well as current valence and arousal ratings than other stimulus datasets. The data set is also not under any copyright restrictions, allows free download and usage while the huge number of images under the four themes allows the users to interactively reconnoiter the images by its categories. Image stimuli reflect both the social and physical worlds and could use its pixel characteristics to represent scenes, animals, humans and objects that are capable of eliciting emotional reactions, including anger, contentment, disgust, love, amusement, fear or happiness. Anthony and Seth (2018) and, Alarcao˜ and Fonsec (2017) have utilised the OASIS data set for research studies.

The GAPED (Dan-Glauser and Scherer 2011) was developed in order to escalate the obtainability of visual emotion stimuli to the research world and resolves the IAPS' data set limitation of inadequate number of pictures under specific themes. The data set

contains 730 pictures that contains both negative and positive contents capable of eliciting different emotions in subjects participating in the experiment. Some of the negative content pictures include snakes and scenes relating to violation of moral and legal norms, such that the negative emotions can readily be elicited. The positive contents on the other hand contains human pictures and animal babies, while the neutral pictures contain non-living objects. The ratings of arousal and valence were done for all the pictures and the results obtained are presented. The GAPED data set has been utilised in (Lakens et al. 2013; Zhao et al. 2014; Stöckli et al. 2017).

The IAPS (Lang, Bradley and Cuthbert 2008) since inception has contributed immensely to the advancement of research and it is perhaps one of the most frequently used emotion stimulus datasets with several thousands of research papers published using the IAPS images. The data set contains 1,195 emotionally evocative colored images, normatively rated on the 3D – arousal, valence and dominance scale. Despite its success, the data set is criticized for being developed during the pre-internet research period and does not contain recent images and normative affective ratings, as well as a limited number of images under specific themes of interest. The data set has been used in studies reported in (Yisi and Olga 2013; Brouwer et al. 2013; Lachezar et al. 2015; Betella and Verschure 2016; Martin et al. 2018).

The IADS (Bradley and Lang 2007) contains 167 naturally occurring sounds, mapped to a number of contexts that elicit a wide range of emotional responses in subjects participating in experiments. The data set contains a non-verbal emotional set of sounds, which is analogous to the emotional facial expression. It is rated on the affective dimensions of arousal, valence and dominance. These sounds can be mapped to musical instruments, human sounds, means of transport, objects, animals and other scenarios. Research studies have used this data set for emotion recognition studies classifying emotions along the discrete and dimensional scales. Some of the studies that have utilised the IADS data set include (Brouwer et al. 2013; Soares et al. 2013; Yisi and Olga 2013)

The Affective Norms for English Words (ANEW) is another emotion eliciting data set that offers a group of normative emotional ratings for a huge number of words in the

English language (Bradley and Lang 1999b). The ratings of these verbal materials are done using the arousal, valence and the dominance scale as the data set complements the IAPS and IADS emotional rated stimulus datasets. Montefinese et al. (2014) and Soares et al. (2012) are some studies that have utilised the ANEW data set of experimentations.

As a result of some glaring limitations associated with the pictures or static images, emotional stimulus materials, many research studies have considered using audiovisuals or films as emotional stimuli (Schaefer et al. 2010) for eliciting discrete emotions such as amusement, happiness, anger, disgust, sadness and fear, including both negative and positive emotions. As compared to the use of other methods of emotion elicitation stimulus, the film clips offer numerous benefits, which include the provision of strong emotional context, dynamic and biologically appropriate stimuli within a relatively short duration as well as the ease of standardization. The integration of auditory, visual and sometimes pictorial information on the films offers the added advantage of meaningful and seamless information, communication across the sensory modalities (Schaefer et al. 2010). In addition, as opposed to the experimental and laboratory-like models of emotion elicitation stimuli, which are considered rather manipulative, film clips/videos are relatively observed by the viewers as a pleasurable and conversant activity undertaken by people in their day-to-day living and are less manipulative by the viewers (Gross and Levenson 1995). All these benefits, among others, motivated the choice of using film clips/videos as the emotion elicitation materials in this research study.

2.6.1 Film Emotion Stimuli Datasets

The FilmStim film clips database was developed by Schaefer et al. (2010) as an emotion stimuli data set for affective related experiments. It is made up of 70 film clips of between 1 – 7 minute duration, with 10 films each catalogued under 7 human emotions, which are fear, anger, tenderness, sadness, neutral, disgust and amusement. Each of the film clips was rated by 364 subjects participating in the experiment, along 24 classification norms, such as positive and negative affect, subjective arousal, the positive and negative affect scores obtained from the Differential Emotions Scale, 15 mixed feelings scores and six

emotion discreteness scores. Despite the number of videos utilised, the data set is criticized for global labelling of the videos, which is considered inadequate to construct ground truth data for induced emotion labels (Baveye et al. 2015) especially as emotions are known to last only a few seconds from onset to offset (Rottenberg, Ray and Gross 2007).

The Emotional Movie Database (EMDB) is another emotion elicitation stimuli data set and consists of 52 non-auditory film clips. The data set was introduced by Carvalho et al*.* (2012) with each film clip lasting 40 seconds and extracted from commercially produced films. The arousal, valence and dominance ratings of the clips were done on a 9-point scale of the experiment subjects. Although the non-auditory nature of the clips offer the advantage of future experimental manipulations of the clips, the data set has a drawback of introducing a certain degree of artificiality as well as the inability of multimodal processing when using the clips, especially in speech based emotion studies, since only the visuals are available and not the audio (Carvalho et al. 2012).

LIRIS-ACCEDE was introduced by Baveye et al. (2015) and contains 9,800 segmented short video clips, lasting 8 to 12 seconds, which translate to about 26:57:8 hours for all the 9,800 clips, with each segment considered large enough to obtain consistent excerpts for specific emotions to be readily elicited in the experimental subjects. The segmented clips were annotated using the valence arousal 2D space; it is freely available and shared under the Creative Commons licenses. The 9800 excerpts were extracted from 160 films obtained from the video platform VODO and the database is reputed to be the largest video database currently available. Animation, comedy, action, horror, thriller, romance, documentary, adventure and drama are the 9 genres under which the 160 movies are classified. The excerpts reflect a wide range of scenes such as violence, murders, landscapes, sexuality and some other positive scenes. Sabyasachee, Rahul and Shrikanth (2017) and Leimin et al. (2017) have used the LIRIS-ACCEDE data set of experimentations.

Some of the multimodal and physiological signals based emotion datasets that have been developed using videos/films for elicitation of experimental subjects are hereby discussed. The MAHNOB-HCI data set (Soleymani et al. 2012) utilised 20 short emotional

videos, shown to 30 experimental subjects as stimuli to elicit emotions in them, while harvesting peripheral physiological data, Electroencephalograph (EEG) readings, audio recordings, eye gaze and facial videos from the subjects. The subjects rated the videos using emotional keywords, valence, arousal, dominance and predictability. This emotional stimulus data set was used for the study described in this dissertation. The list of the 20 videos utilised as well as their sources are provided in (Soleymani et al. 2012).

The Database for Emotion Analysis using Physiological Signals (DEAP) is a multimodal data set developed by Koelstra et al. (2012) for the investigation of human affective states. An aggregate of 40, one-minute duration each, of music videos was used as emotion elicitation stimuli for 32 experimental subjects whose peripheral physiological signals as well as EEG readings were harvested as they experienced the affect (Koelstra et al. 2012). The videos were rated by the experimental subjects along the valence, arousal, dominance, familiarity and like/dislike spaces. Classification results were obtained by utilizing the peripheral physiological and EEG modalities along the valence, arousal and like/dislike ratings while a decision fusion of the classification results was also performed. Some recent research studies that have utilised the DEAP data set include Xiang et al. (2018) and Thammasan (2017), Yin et al. (2017), Wang and Shang (2013), Li et al. (2015), Jirayucharoensak, Pan-Ngum and Israsena 2014.

These emotional stimuli datasets have varied characteristics such as the number of subjects from whom data were collected, emotion channels and stimuli (audio, gesture, visual, audio-visual, and physiological), nature of expressions (acted, simulated or spontaneously) and whether these emotions were induced/elicited or naturally expressed.

Out of the emotional databases stated above only the DEAP and MAHNOB-HCI contain physiological emotion datasets, are publicly available for experimentation and have been used to recognize emotions, while other research studies have developed several in-house corpuses that are not publicly available.

2.7 Identifying Emotions

Different methods may be used to identify the emotions felt by subjects from whom an emotion is elicited while participating in an affective experiment.

2.7.1 Self-reporting

In emotion recognition research studies, it is essential that an individual's experienced emotion is the same one being analysed and reported, in order to give proper credence to the results obtained and arguments posited. Knowing the emotion experienced is mainly achieved by the subject participating in the experiment reporting his/her emotion, while the researcher compares the emotion with the ground truth.

Both the dimensions and discrete models of emotion representations have been employed in self-reporting emotion. The subject participating in the experiment either liberally uses emotional keywords to state his/her emotion experienced, or is asked specific questions, or to choose from a pre-determined list of emotions, keywords, range of values or symbols (Soleymani et al. 2014). Several methods have been adopted for emotional self-reporting, which involve the collection of subjective affective ratings from experimental subjects. These include the use of the Affective Slider (AS), the Geneva Emotion Wheel (GEW), Positive and Negative Affect Schedule (PANAS) and Self-Assessment Manikin (SAM), among others. These methods are briefly discussed below.

2.7.1.1 Affective Slider

The Affective Slider (AS) is a tool that was introduced by Betella and Verschure (2016) for emotion self-reporting. The AS is digital in nature and uses two slider controls to achieve a quick assessment of pleasure and arousal in human emotions. The results achieved by the authors suggested that the AS matches the SAM on the pleasure and arousal scales of emotional self-reporting. It can also be readily reproduced in the modern day's digital devices, such as tablets and smartphones, while it also does not require written instructions.

2.7.1.2 Geneva Emotion Wheel

The Geneva Emotion Wheel (GEW) method is a semantical self-reporting tool that was developed by Scherer (2005), using 20 emotions distributed in a circular manner to combine both the discrete and dimensional approaches of emotion representations. Each of the 20 emotions arranged around the wheel, has five circles of increasing sizes from the centre outwards. The circle size indicates the activation level of the emotion experienced. Subjects participating in the experiment are required to choose two emotions considered to be the closest to their experienced emotion and also report the activation level by marking the corresponding circle size. The emotions have been carefully arranged in the circle in such a manner that the high-control emotions are placed at the top while the low-control emotions are at the bottom with the horizontal axis representing the pleasure/displeasure dimension.

2.7.1.3 Positive and Negative Affect Schedule

Watson, Clark and Tellegen (1988) developed the Positive and Negative Affect Schedule (PANAS) emotional self-reporting method. The method is made up of two scales that measure the positive and negative affect, while it shows the relationship between the two. 10 affect descriptors each were used for both the negative and positive affect as experimental subjects were asked to what extent they feel at the present moment or over a week. They responded to this 20-item affect descriptors test by scoring each on a 5-point scale with values of very slightly to extremely i.e. 1 to 5 points. This study was later expanded by Watson and Clark (1994) by reporting up to 11 discrete emotion groups, using about 60 emotional keywords. The work reported by Crawford and Henry (2004) validated and confirmed that PANAS is a dependable assessment of the constructs being measured, while rejecting the hypothesis of the complete independence between the positive affect and the negative affect. PANAS method was considered unsuitable for experiments of short time duration and multiple stimuli, due to the long duration required to answer the questionnaire.

2.7.1.4 Self-Assessment Manikin

The Self-Assessment Manikin (SAM) is made up of manikin-like shapes, which depict expressed emotions and is popularly used by psychologists as an emotional self-reported tool. The Pleasure-Arousal-Dominance 3D scale is normally used to represent the expressed emotions (Soleymani et al. 2014), while other studies have also used the picture of the shape of a human thumb to express likeness or dislikeness to emotions (Jirayucharoensak, Pan-Ngum and Israsena 2014). The Manikins are pictorial and designed in such a simple way as to make easy understanding of experimental subjects to facilitate their choice of manikins best expressing their emotion, as well as values that can be attached to arousal and valence levels. The SAM method is constrained by its inability to support the expression of co-occurring emotions, though the tool is language-independent and does not require experimental subjects to verbally state or write down expressed emotions (Soleymani et al. 2014). For the purpose of emotional self-reported by subjects in the research study reported in this dissertation, in addition to other instruments utilised, the SAM method was adopted by Soleymani et al. (2012) in constructing the MAHNOB-HCI dataset. This is because of its popularity, ease of understanding by experimental subjects and language independence. It is also preferred over the AS because of the 3D PAD scales rather than the 2D pleasure-arousal scales of AS.

2.7.2 Emotion Measurement

In the literature, emotions have been measured, analysed and recognised using facial expressions, gestures, visuals, voices and physiological signals. Facial expression emotion measurement applies different techniques such as Facial Action Codings (FACS), Facial Detection Parameter (FDP) and Facial Animation Parameter (FAP) on the emotional facial images. The parameters including positions, outlines and movements, thus determine readings such as nose, mouth and eye positions, as well as eyelid and head movements, pupil diameter, eye blinking, gaze distance and coordinates amongst others (Kollias and Karpouzis 2005).

Emotional speech signals bear lexical, gestural and textual information. Various parameters such as pitch/intonation, intensity, speech quality, duration - speaking rate and pauses have therefore been measured in the speech based emotion analysis and recognition (Ramakrishnan 2012). Desired features are extracted from these readings for consequent analysis and recognition. On the other hand, gesture based emotion recognition systems, such as the hand tracking systems have measured users' movements, including speed, direction, acceleration and variation, from which desired feature vectors were extracted/selected for emotion recognition.

However, the peripheral and autonomic central nervous system's signals that have been collected, according to the literature, were centred on the human body systems, such as the cardiovascular, electrodermal activity, respiratory, muscular and the brain activity/nervous systems. The physiological signals include blood pressure (BP), Galvanic Skin Response (GSR), skin glucometer, Blood Volume Pulse (BVP), skin temperature, Electrocardiogram (ECG), Heart Rate (HR), Electrodermal Activity (EDA), Electroencephalogram (EEG), Electromyogram (EMG), Electro-oculogram (EOG), respiration, spirometer and Heart Rate Variability (HRV) (Chun-yan, Hai-xin and Wang 2013). These variable signals can be pooled to extract a group of discriminating features to train a pattern recognizer in a physiological based emotion recognition system.

In the next sections, Human Activity Recognition (HAR) and location detection are presented as the duo has capabilities that can be combined with physiological signals based emotion recognition procedure with a view to tracking an individual in a distress phase situation.

2.8 Human Activity Recognition

This section will discuss Human Activity Recognition (HAR), since the study reported in this thesis focuses on emotion recognition with a view to utilizing it with activity recognition as well as location information for an individual tracking.

The essential features of very high computational power, affordable cost, small, lightweight, portable sizes offered by sensors and mobile devices, allow people to

manipulate and communicate with the devices as an integral part of their lifestyle and daily living. This has given birth to ubiquitous sensing, which is now a vibrant research area concerned with the extraction of knowledge from data acquired through pervasive sensors (Perez, Labrador and Barbeau 2010).

The need to recognize and classify human ambulatory activities, such as running, walking, sitting, cycling or riding a bus, has become quite germane in offering evaluative responses to medical doctors about a particular patient's behaviour. Accurate information (remotely harvested from sensors) on the soldier's activities in addition to their locations and health conditions through acquired vital signs or physiological signals, including heart rate and temperature, are significant for their remote monitoring, safety and performance evaluation.

Some very earlier research on Human Activity Recognition (HAR) began in the late 1990's (Schlmilch et al. 1999; Foerster, Smeja and Fahrenberg 1999). Various inherent challenges during this early stage have given birth to the emergence of new technologies to address and improve the accuracy under more feasible conditions using algorithms. These challenges include among others, the construction of miniaturized, affordably cheap and small sized data acquisition systems; the collection of data under practicable conditions; identification, pruning and selection of attributes to be evaluated; implementation in mobile devices satisfying processing and energy requirements; and reusability that does not require re-training the system (Kim, Helal and Cook 2010).

2.8.1 External Sensors in Human Activity Recognition

Human Activity Recognition can largely be investigated using two distinct approaches, namely through external sensors and wearable sensors. Sensors are the primary sources for raw data collection in activity recognition. External sensors are made up of devices that are placed in a location of interest that are pre-defined, such that inference and decision making are purely based on the deliberate interaction of the subject being monitored with the sensors. Some external sensors include the laser range finders,

cameras, wireless networking infrastructure, infrared and ultrasound sensors (Hightower and Borriello 2001).

The Smart Home (SH) (Sarkar et al. 2010; Kasteren, Englebienne and Krse 2010; Tolstikov et al. 2011; Yang, Lee and Choi 2011) provides another very good example of what external sensors are. SH technologies capable of automatic detection of need for assistance and sending/delivery of alerts (Short Message Service (SMS) and emails) through mobile devices could, for instance, significantly reduce the costs of elderly care or patients diagnosed chronic diseases such as memory loss disease (dementia) among others. The ability to recognize human activities, such as washing clothes, cooking, opening the door, making tea and eating, by the system is fairly complicated, because data acquisition is achieved via a number of external sensors placed on target objects such as the stove, washing machine, door, tap and the individual being monitored is expected to interact physically with these objects for the required data to be acquired. Among the limitations associated with external sensors and restricting its wider adoption is the inability to read or acquire data the moment the user is not within the reach of the sensors or deliberately does not carry out activities that interact with the objects on which the sensors are placed. The maintenance of the sensors as well as their installation also comes with a high cost to the system (Oscar and Labrador 2013).

A camera provides another typical example of the external sensors for an HAR. Activity recognition is achieved via the device from the video sequences being offered to track a user's daily life activities. The video sequences are ideal in military and security systems for intrusion detection, surveillance and interactive applications. The use of video sequences in HAR has the glaring limitations of privacy, complexity and pervasiveness (Oscar and Labrador 2013). It is not every individual that would readily allow being tracked or monitored and recorded by cameras. In addition, video processing techniques are also quite expensive, thereby challenging the scalability of a real time HAR. In addition, video recording devices pose the difficulty of attaching them to a target individual for obtaining images of the whole-body frame during the performance of activities of daily living, coupled with the challenge of the individual being monitored to continuously be in a predefined perimeter covered by the location and proficiencies of the camera device.

These weaknesses in the use of external sensors brought the emergence of wearable sensors with a view to overcoming the observed limitations.

2.8.2 Wearable Sensors

Most research works in smart wearable systems have been engrossed with smart mobile devices and other intelligent environments which constitute wearable computing, thus moving the interface for computational environments from the (smart) home to individual users' bodies, though the wearable, implantable and swallow-able means (Ciuti, Mencassi and Dario 2011), clothes - smart clothing (Smith 2007) and also as portable accessories or body jewellery like wearable sensors.

A wearable sensor device and system is essential in security tracking and monitoring, healthcare, military and sporting. It has features capable of tolerating the deployment of miniaturized portable and wearable sensors around the human body or clothing to instinctively collect location, activity, mobility and physiological data of individuals and transmit them over a secure network to a device or application for processing. Leveraging on these features enables the wearable sensor to be suitable for tracking of individuals in a distress phase situation, such as vehicular or fire accidents and abduction phase scenarios. Because of impressive progress recorded in the field of computing and mobile sensing, wearable computing has brought forth very innovative approaches using various activity recognition algorithms to recognize, classify, label human actions and environmental conditions automatically (Roggen et al. 2011).

Some physiological sensors monitoring vital signs (heart beat rate, SpO_2, body temperature), environmental sensors (that measure humidity, air temperature, sound, light) and a location tracker sensor can all be coupled into a wearable sensor network (WSN) system, but the obtrusiveness factor and weight of the device must be taken into consideration for ease of use, carriage and mobility of the individuals being monitored. The progress made with wearable devices, particularly watches, has brought this type of monitoring device within reach of a very large section of the population.

The attributes of interest to be measured in a wearable sensor system are defined by the user's ambulation or activity movements using accelerometers and location information via GPS devices; environmental variables like air temperature, wind speed, precipitation and humidity; and physiological signals including heart beat rate or electrocardiogram and body temperature all using various sensors for measurement.

Broadly categorized, the types of activities recognised by an HAR system include; walking, running, sitting, standing, climbing/descending stairs, riding an elevator, riding an escalator, pulling, pushing and struggling). Riding a bus, cycling and driving are grouped under transportation, while phone usage activity includes text messaging and making/receiving a call. Furthermore, activities of daily living are made up of brushing teeth, drinking, reading, eating, TV viewing, scrubbing and vacuuming. Sporting exercise and fitness include activities such as rowing, lifting weights, spinning, fast walking, doing press-ups, while military operational activities are classified as crawling, kneeling and opening a door among others. Upper body movement activities include moving the head, swallowing, sighing, speaking and chewing (Oscar and Labrador 2013).

2.8.3 Architecture of a Typical Human Activity Recognition System

In an attempt to recognize human activities, specific wearable sensors are affixed to the individual's body (see Figure 2.1) in order to acquire desired attributes, including location (Iglesias et al. 2011), temperature (Choujaa and Dulay 2008) and ambulatory movement (Shotton et al. 2011). These wearable sensors are calibrated, initialized, synchronized, configured and programmed to communicate with an Integrated Device (ID) such as a laptop (Parkka et al. 2006; Oh, Park and Cho 2010), mobile phone (Jatoba et al. 2008; Brezmes, Gorricho and Cotrina 2009) or a customized embedded system (Tapia et al. 2007; Kao, Lin and Wang 2009).

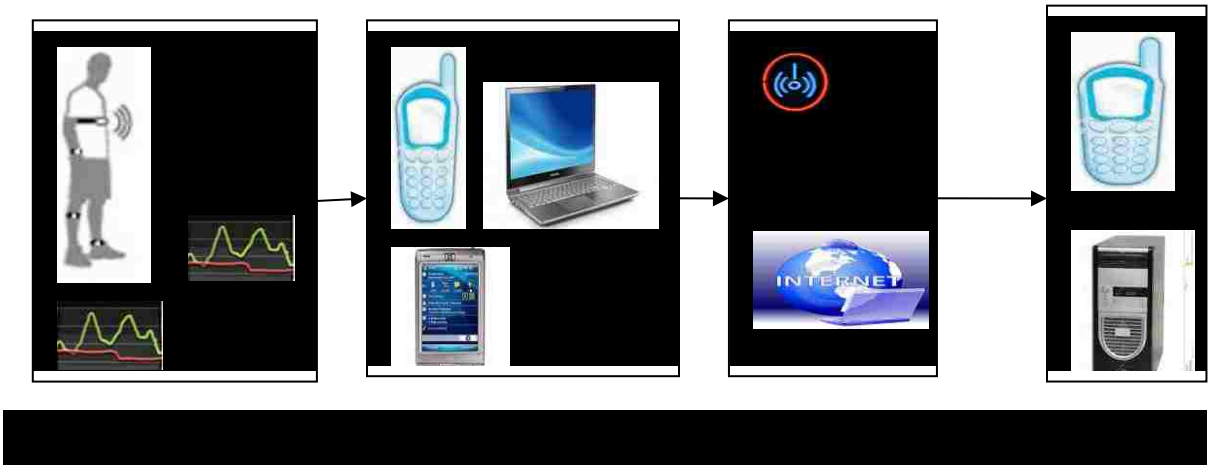


Figure 2. 1: Generic Architecture of a HAR System (Oscar and Labrador 2013).

The ID is primarily tasked with strategically pre-processing the data acquired from the wearable sensors and could in some systems transmit the acquired data to a dedicated application server to enable real-time monitoring, analysis and visualization (Parkka et al. 2006; Maurer et al. 2006). The data transfer communication protocol used to transmit the data might be the User Datagram Protocol (UDP) or Internet Protocol using the TCP/IP.

Generally, each HAR system differs from the other. Sensor data were collected offline in Bao and Intille (2004), Hanai, Nishimura and Kuroda (2009) and Lara et al. (2011), therefore, real time communication or server processing was not required. He and Jin (2008), He and Jin (2009), Berchtold et al. (2010a) and Riboni and Bettini (2011) incorporated sensors within the integration device, while some other studies directly executed the inference process on the ID instead of the server (Brezmes, Gorricho and Cotrina 2009; Berchtold et al. 2010b).

2.8.4 Related Works in Human Activity Recognition

Earlier systems used different approaches and technology to identify activities for various purposes. Environmental attributes such as air temperature and humidity can also be

measured to give contextual information with respect to the user's immediate environments or surroundings and enhance the knowledge driven capability of the system. For instance, a low audio volume or low light intensity might indicate that the individual being monitored is probably taking a nap.

A health care system was developed by Jovanov et al. (2005) based on wearable biomedical sensors with the architecture made up of distinct layers of biomedical sensors, personal server and the health care server respectively. The system offered low-cost, small weight, miniatures and an intelligent sensor platform for integration into a body area network for health monitoring. Real time sensors' data are processed by the system while providing feedback to the user as well as generating information on the user's state, level of activity and environmental conditions.

Maurer et al. (2006), using various sensors, worn on different body parts, developed a eWatch system to monitor users' daily activity and classify these activities. Using multiple sampling rates and time domains, features were collected and analysed. Computational complexity and accuracy of recognised activity were compared and a trade-off established between the classification accuracy obtained by wearing the electronic devices on various body positions and the computational complexity.

Jung et al. (2008) proposed a ubiquitous healthcare system deploying vital signs and environmental sensor devices. These devices received data that the system intelligently used to monitor health conditions of patients and administer treatment in real-time, saving the medical costs of rendering such services in physical hospitals and treatment rooms.

Atallah et al. (2009) investigated activity recognition, using a lightweight ear-worn device and other wireless ambient sensors for recognizing common Activities of Daily Living (ADL) using two-stage Bayesian classifiers with data acquired from both types of sensors. Choujaa and Dulay (2008) and; Kao, Lin and Wang (2009) who designed some early HAR systems used microphones, light sensors, thermometer and humidity sensors to measure environmental attributes. Most of these sensors that monitor environmental attributes were largely coupled with accelerometers and other sensors (Yin, Yang and

Pan 2008) to measure other attributes of interest and integrating the data acquired with those of the environmental sensors for efficient and accurate decision making in the system.

Pung et al. (2009), introduced context-aware middleware for pervasive elderly home care, using a peer-to-peer based context query processing, context reasoning for activity recognition and context-aware service management for patient monitoring, emergency response service based on location, abnormal activity detection, social networking and ubiquitous medical data access.

Narayanan et al. (2010) implemented a wearable tri-axial accelerometer system using data acquired from one waist-held tri-axial accelerometer for classifying various human movements in real time. The system performed signal processing using a wearable unit platform and intelligently differentiated periods of rest and activity. Expended metabolic energy was calculated against recognised activity such as fall and orientation posture of users.

A mobile application based on context-awareness was used by Iglesias et al. (2011) to proactively determine and evaluate the activity being performed by a subject during a daily living activity. This application applied fusion method to estimate movement and location by combining acceleration and radio data from in-device and external sensors. The aim was to offer substantial context-aware notifications by making the user aware of his level of activity.

Olugbara, Ojo and Adigun (2011) developed a ubiquitous healthcare service system using a grid-enabled framework, aiming at an improved healthcare service provisioning at a minimum cost. The research focused on distributed healthcare resource provision with more inclinations for resource sharing and re-use through grid computing and mobile technology.

Ogunduyile, Olugbara and Lall (2013) implemented a prototype wearable ubiquitous healthcare system to monitor physiological vital signs of an elderly person. The system consisted of an integrated biomedical sensor, including an accelerometer and SpO_2, which collected physiological signals of the elderly and then sent the data to the

Intelligent Base Node (IBN) via wireless transmission, from where the data was uploaded to the health server for managing the health status of the elderly.

An automated fall detection prototype system was developed by Ozdemir and Barshan (2014) using wearable motion sensor units fitted on six different locations to the body of the user. Each unit consisted of 3 tri-axial accelerometers. Voluntary falls and various Activities of Daily Living (ADL) were measured. Data were analysed with activities classified using various machine learning methods, such as Artificial Neural Networks (ANNs), Least Square Method (LSM), Bayesian Decision Making (BDM), k-Nearest Neighbour (k-NN), Support Vector Machine (SVM) and Dynamic Time Warping (DTW). This work introduced the glaring obtrusiveness of HAR systems, making it unsuitable for the research area of application proposed in this study, even though similar machine learning techniques could be adopted in the proposed affective system.

Similarly, the research work carried out in this current study on emotion recognition for utilization with human activity recognition significantly differs from this early work in terms of the algorithms tested for data preprocessing, feature extraction, selection and classification for emotion recognition, which are obviously not adequately covered in the prototype.

Physiological signals in an HAR system consist of vital signs data including, galvanic skin response, heart rate, SpO_2, body temperature and pulse. In measuring the physiological signals, other additional sensors are required. To measure the heart rate of an individual being tracked, a heart rate monitor and accelerometers are required. This system increases the costs and thereby introducing obtrusiveness (Choujaa and Dulay 2008).

Major research issues in the design and implementation of the specific HAR system are classified as obtrusiveness, data collection processes, activity recognition performance, methods of data processing, energy consumption by the system, flexibility (reusability) and selection of attributes. Attributes of interest for selection in HAR are mainly environmental attributes; ambulation/acceleration, location and physiological

signals. These are collected and measured using sensors worn by the individual to be monitored in an emotion recognition and HAR system.

2.8.5 Design Considerations

Obtrusiveness is a critical factor for consideration in the design of HAR systems. Wearing many sensors at the same time should not be required in HAR systems or interacting too often with the application, such as in the use of external sensors thereby contributing to their limitations. Though, if the available sources of data from sensors are many, the processed information would be richer, but HAR systems that demand a user being monitored to wear four or more accelerometers (Oh, Park and Cho 2010; Lara et al. 2011; Vergara-Laurens and Labrador 2011) may make configurations of sensors to be complex, invasive, and expensive, which would render them unsuitable for activity recognition, especially in an abduction phase scenario, where professional abductors aim at first dispossessing potential victims of valuables, such as a mobile phone, and conspicuous personal effects.

A data collection protocol is also of huge importance in the design of HAR systems. It involves the step-by-step procedures followed by an individual while collecting data. Features such as the number of individuals for whom testing or training data would be collected, their physical characteristics and demographics such as age, gender, height, weight and other varied health conditions are all significant in the data collection methods. HAR systems that are person dependent suffer reusability limitations. Though data collected from an individual should be huge enough to cater for intra person variability in the data signals collection and activity performance, a model should be trained that allows person independent activity performance and recognition. Such system model will allow reusability and reduce cost of having to train the system for different users over and again.

The accuracy and recognition performance in a specific HAR system rests on several interrelated factors like the activity data set, the quality and size of the training data set, feature selection and extraction algorithms deployed and the learning algorithms to be used.

The issue of energy consumption should also be considered vital in HAR systems. Wireless networks such as Wi-Fi and Bluetooth that are short ranged are preferred to long-range networks like WiMax or cellular networks as short-range wireless network need low power to function. However, consideration should be carefully made in respect of suitable transmission networks or models for a specific research. Some energy savings mechanisms being currently adopted in HAR systems are data aggregation and data compression without necessarily involving more computations that could hinder the application's performance. To prevent raw signals being continuously streamed to the server, feature selection, extraction and classification are better carried out at the integration device stage (Maurer et al. 2006; Berchtold et al. 2010b) though consuming more energy and making the system expensive. In addition, not all sensors might necessarily be in use simultaneously. In saving energy, some of the sensors not in use might be turned off or their sampling rate transmission reduced. For a user's activity of "standing still" as an example, turning off the GPS sensor could significantly save energy (Chen et al. 2008).

Another important issue in HAR is the signals or data processing phase, this requires the signals to be processed for discriminatory features to be extracted for machine learning and for an activity to be consequently recognised. The HAR design should determine whether the activity recognition processing would take place on the ID or on the main server. Resource constraints in the form of processing capacity, storage and energy, hinder data processing on the ID. Feature selection and extraction as well as machine learning algorithms need careful blending in order to ensure a rational and realistic response time while also preserving the battery life of the device.

2.9 Location Detection

The location of a person/object is very essential in providing an assistive or emergency response when required. If the location information is not updated, then virtually no assistance can be rendered.

2.9.1 Overview of Existing Devices

Over time, ICT has deployed solutions to address the problems of knowing the exact location of an individual. Closed Circuit Television (CCTV), cameras and satellites as monitoring devices also proved to have limitations in terms of physical attacks on the devices, a predefined perimeter coverage range of the devices and the fact that most of these technologies are data driven only while not knowledge driven, as they do not provide the actual environmental state and physiological health conditions of the person being trapped or held hostage. Though, knowing the exact location of an abducted or trapped victim is essential in providing rescue services, data relating to the physiological state of the victim, such as body temperature, heart beat rate, respiration, oxygen in the blood level, pulse and whether the victim is sitting, standing, crawling, walking, running, shouting or has fallen down are vital to the security agencies or rescue team for prompt decision making and logistic support. In medical health services, the ability to remotely identify the exact location of a patient and his/her environmental situation would also prove useful in the deployment of healthcare services and facilities at the optimal time, thus reducing fatalities and medical services cost.

The Global Positioning System (GPS) plays a vital role in any location detection system by enabling various forms of location-oriented services. Most cellular phones today have GPS functionalities, thus enabling the sensor to be very appropriate for context aware applications, such as recognizing the transportation mode of the user (Chen et al. 2008). The exact location of a person using a GPS device could also be important in making decisions about his/her activity, using the ontological reasoning method (Berchtold et al. 2010a). For example, a person in a hall might probably be standing or sitting but not brushing his teeth.

A GPS tracker harvests, analyses and saves location data acquired from weak GPS satellite signals, process this location information and stores the processed information for subsequent review or transmits it in real-time. For better performance and accuracy, the GPS tracking devices should directly access the open sky to enable efficient capturing of satellite signals and transmitting the signals to the cellular network.

The most common GPS tracking device is a smart phone, others include Garmin GTU10 bearing in mind that for almost every purpose, there exists a GPS tracker, as they enhance security and assist in tracking the location and possession of an individual. For a person tracking device to be efficient, it should be able to accurately track for long sequences, track without dependence on activity being performed, self-start, and be computationally efficient, while offering great robustness in drift changes and continuously track when brief but not permanent occlusions occur (Ramanan and Forsyth 2003).

The tracking services offered by the Global System for Mobile Communication (GSM) service providers in the form of using GPS/GPRS (General Packet Radio Services) to know the exact location of individuals have some limitations. These include; having the individual with a GPS enabled mobile phone or watch to always turn it on, with the attendant draining of the device's battery power (Chen et al. 2008), and the legal implications of not seeking the consent of the person being tracked in terms of his/her privacy rights being breached. GPS devices or sensors also do not work very well indoors and therefore require deployment of other sensors like accelerometers with them (Berchtold et al. 2010b) to aid activity recognition processes. The tracking technology being employed here, associates a mobile device to the nearby base station owned by the service provider. If the mobile device goes out of the coverage area of the service provider, then the tracking fails. Privacy issue concerns are associated with location data, because not all users are willing to be monitored. The concept of encryption, anonymisation and obfuscation are some techniques being deployed to significantly enhance location data privacy (Huang, Kanhere and Hu 2010; Christin et al. 2011).

The advantages of the GPS tracking are that it can be readily accessed by users for varieties of applications, while very useful information on locations and places can easily be acquired through Google places web services (www.code.googlecom./ apis/maps).

2.9.2 Benefits of the Proposed System

The idea of deploying a combination of a data driven and a knowledge driven approach, as offered by wearable sensor devices and supported by machine learning models, such as artificial neural networks and the Markovian Logic Network concepts for modelling emotion and activity recognitions, could prove to be a very good solution in the case of abduction events, counter terrorism, vehicular accidents, fire trappings and missing person scenarios. Machine learning algorithms for pattern recognition and classification are novel approaches to emotion and activity recognition. They have the capability to combine first order logic (common sense), which provides relational information of objects, together with the probability of uncertainty in knowledge, thereby improving the recognition capability of the system.

Some models offer combinations of logic (common sense) and uncertainty in knowledge include the Bayesian Network Model, Naive Bayesian Model, k-Nearest Neighbours, Decision Trees, Markovian Logic Network, Neural Networks, Support Vector Machine and Hidden Markov Models (HMMs). Each of them has its limitations and strength in solving a specific problem relating to training a model for emotion/activity recognition as the strength of the models would be examined in order to discover the best approach. Emotion recognised in the system can however be utilised on its own or combined with human activity recognised as well as the location information of an individual to determine if an emergency response is indeed required to be prompted by the proposed system. In summary, the benefits of the proposed system includes:

(1) The proposed system in this research work combines data driven and knowledge driven approaches while also applying machine learning techniques for modelling human emotion recognition.

(2) Digital image processing techniques will also be utilised for data processing and feature extraction towards obtaining discriminative features and yielding impressive results.

(3) A traditional and shallow learning approach consisting of the artificial neural networks model will also be utilised in the proposed system and has capabilities of better performance than some of the state of the art results based on deep learning techniques, which are also associated with high complexity and computational costs.

(4) The proposed system can also prove beneficial to security agencies, emergency services and rescue teams as well as potential victims and their families in proactively proffering solutions in an emergency event of emotionally induced distress situation.

CHAPTER THREE

Theory of Emotion Recognition and Emergency Response

This chapter introduces an architecture for an emergency response system for tracking persons in distress phase situations, including the physiological signals measured, the sensors utilised, methods and techniques used.

3.1 Emotions in emergency system

An emergency may be defined as a sudden and possibly harmful or dangerous situation that requires instantaneous attention and action. These situations differ and ranges from negligible, which could be personally handled by the subject without the intervention of emergency or rescue services, to neutral/medium emergencies involving the presence of principal emergency personnel, to monumental/major emergencies (Roche, Harney and McDowell 2004).

An event can therefore be referred to as a major emergency if with or without little prompting, causes or threatens death or injury. Such events are also capable of causing substantial disruption of essential services, damage to infrastructures and the environment by outstretching the resource capacity of the principal emergency services. Specific additional procedures and resources therefore need to be activated to enhance a well-coordinated and effective rescue operation. It has been noted that accidents and injuries are critical health threats, which could result in permanent disability and or extended periods of hospitalization and convalescence, which further aggravates health problems (Krug, Sharma and Lozano 2000). The tracking of persons, using emotional information, activities and location properties, with a view to averting death or related injuries can thus be termed a major emergency system.

Various advanced technologies, including mobile and wearable sensor technologies for emergency responses, are now readily available, but rescue efforts may nonetheless be constrained by issues relating to interaction among rescuers and rescue agencies, including firefighter, police, ambulances and paramedics. Humans, combined with

technology co-agency, are embedded in an emergency response system, as both human and technical artefacts' interactions are needed to accomplish system goals (Staffan 2011). This viewpoint requires that the emergency system performance necessarily depend on the understanding of core issues, including the design and utilization of artefacts for coping with the expected complexity within the system, management of time and improvement of control. The ability of an emergency response system to efficiently function and coordinate is inherently determined by its architecture. This further depends on the efficiency and effectiveness of tools and structures designated to information handling, communication within the system, as well as an effective information system. This is due to calamitous consequences often being associated with real incidents, coupled with limited likelihoods of mid-event learning. It is therefore essential for emergency response systems to have structures for training and learning embedded within the system.

This study offers a solution towards tracking a person, capturing emotions-based physiological signals capable of being linked with the human activities being performed, with a view to alerting an emergency service, if the need arises, in order to avert a possible casualty or disaster. It is therefore imperative to consider the steps involved in the emergency system management, especially as it relates to the system architecture. Salasznyk and Lee (2006) divided the response stage into detection, preparation, response travel and clearance. Responses occur just before, during and directly after an emergency; the focus here in this study is on the "just before emergency" timing, so as to avert a casualty. The response stage, however, broadly includes activities such as search and rescue, emergency medical assistance and treatments, the extinguishing of fires, evacuation as well as the management and coordination of other support services. Chen et al. (2011) voiced the opinion that emergency management systems should involve four different stages, including preparation, mitigation, response and recovery. The response stage is specifically concerned with various actions being undertaken during rescue/evacuation operations, where concerted efforts are launched towards saving lives and reducing facilities/infrastructure damage.

On the other hand, Borges, Engelbrecht and Vivacqua (2011) characterized the recovery stage as involving activities being undertaken to return to normality; both for people and infrastructure. This includes immediate and long-term recovery. Activities such as cleaning of debris, restoration of normalcy, which includes essential services and supplies, and investigating the remote cause and effects, are classified under the immediate recovery. Conversely, activities such as support of casualties, reconstruction of collapsed infrastructure, buildings, services and further restoration of normalcy are grouped under the long-term recovery. In addition, actions taken to either eliminate or shrink the chances of occurrence of an accident or its effects on the people, infrastructure and the environment, are included in the mitigation stage. The preparation stage involves getting ready for expected future incidents and it entails systems development, resource procurement and management to underpin preparedness, developing scenarios, engaging in real-world learning and training, testing of systems, plans and procedures, maintenance, audit and assessments of preparedness. The harvesting of emotion physiological signals and data processing procedures in the emotional tracking system falls under the preparation stage while the emotion recognition results, which could trigger an emergency response, could be classified under the response stage.

3.2 Proposed Emergency Response System

The proposed architecture utilizing Human Emotion Recognition (HER), Human Activity Recognition (HAR) and location data is shown in Figure 3.1. In the first module of the architecture as shown in Figure 3.1, the first step consists of elicitation of the emotion to be recognised while concurrently acquiring emotion physiological signals and activity data from the subjects using the appropriate sensors. The cooking hacks' physiological sensors can offer a practical example of how the physiological signals may be captured. Some of the physiological signals that could be collected include pulse, oxygen in blood, GSR, skin temperature, airflow/respiration and ECG. These physiological signals have been shown in various studies to be related to human emotions (Healey and Picard 2005; Collet et al. 2009; Schneegaas et al. 2013; Karthik, Varghese and Chaitanya 2013;

Solovey et al. 2014; Salazar-Lopez et al. 2015; De Witte et al. 2016; Dawson, Schell and Filion 2017).

However, in order to collect the emotion physiological data, emotions need to be first elicited in the subjects. The elicitation of emotional responses from experimental subjects is a challenging task and requires the selection of the most effective stimuli to achieve a valid response. In the course of eliciting an emotion in a subject under experimental conditions, many different external stimuli have been used by different research works, depending on the contexts and the subjects. These external stimuli include pictures, static images and vignettes, robot actions, voice, games, audiovisuals, films/movies/video clips, audio music, music videos, odour, self-elicitation and recall paradigm (Noppadon, Setha and Pasin 2013a).

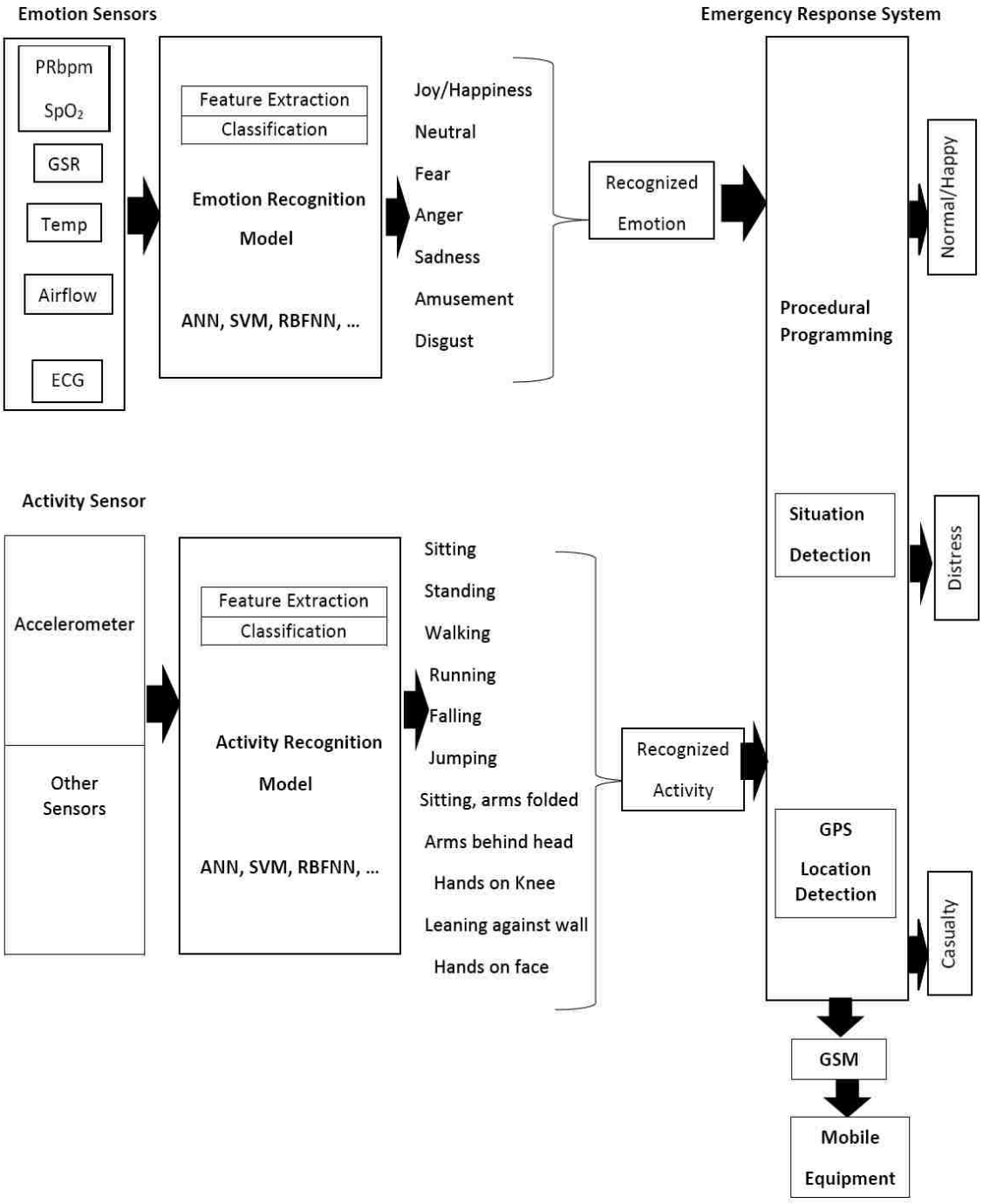


Figure 3.1: Proposed architecture of HER, HAR and location data for remote monitoring (author's own craft).

Films and video clips are popularly used because it offers numerous benefits, which include the provision of a strong emotional context, dynamic and biologically appropriate stimuli within a relatively short duration, as well as the ease of standardization. The integration of auditory, visual and sometimes pictorial information in films also allows the films/audiovisual stimulus to have the added advantage of meaningful and seamless information communication across the sensory modalities (Schaefer et al. 2010). The required physiological signals are collected as the subjects are watching these emotional video clips as it has been noted that carefully selected video clips are capable of eliciting the associated emotions (Soleymani et al. 2012).

In addition to the emotion physiological signals' collection in the first module of the architecture, human activity data could also be collected using the accelerometer sensor. Such human activities include sitting, standing, lying, walking, jogging, running and riding a bus. These activities' data may be collected as the subject is watching the video clips. Activities such as walking, jogging, running and cycling can be performed using a treadmill device and data are collected while the subject is watching the emotional video clips being projected.

The second module in the architecture consists of processing the collected data. The steps in this module include data preprocessing, feature extraction and feature classification in order to be able to detect and recognize human emotions and human activities. Data pre-processing is necessary so as to remove raw signals' artefacts that could affect the quality of features extracted as well as the computation time. Normalization and filtering may also be required in the preprocessing stage. This step is followed by feature extraction and classification into appropriate emotion or activity classes.

However, it is essential to indicate that with respect to the emergency management system, the proposed architecture, shown in Figure 3.1, borrows from the Joint Cognitive System (JCS) (Staffan 2011) viewpoint in the design and implementation of response systems. It encompasses complex socio-technical systems wherein humans and Information Communication Technology (ICT) artefacts interact as a system. The JCS specifically recognises the responsibility of people to manage the subtleties of real-world

situations and that human cognition relies hugely on collaboration, cognitive artefacts and information among systems, people and/or rescue agencies. Response systems such as the one being proposed in this dissertation must thus be designed to enable smooth coordination and communication between the emergency system and the rescue agencies, for instance.

The architecture of response systems is intricately marked by diversity and knotted dependencies between the various components, involving varied configurations of technologies such as wearable sensors, mobile and affective computing. Response systems must be able to anticipate and trigger alerts in order to respond to a situation of concern. Since cognition is a shared process, a number of concepts have been identified by Louise (2007) to be critical in a response system. These are termed the three Cs of emergency response and include control, coordination and communication (from Cognition) (Louise 2007). These concepts are discussed below.

3.2.1 Control

The concept of control in the proposed response system is germane in order to realize the goal of averting a casualty in a distress phase situation. This is addressed by Hollnagel and Woods (2005) in the Contextual Control Model (COCOM), which requires the operators to operate the process, create necessary feedbacks and maintain a dynamic interpretation of the current situation being handled. The control process covers the information integration, risk assessment, planning and monitoring processes (Louise 2007). This obviously requires the response system to plan and take certain actions to achieve this goal (Bergström et al. 2010). Such actions as related to our study, include the timely merging of human emotion, human activity and location information with a view to arriving at a decision of prompting an emergency team response if needed. In the proposed emergency response system in Figure 3.1, various emotions and activities recognised will be considered and interpreted, whether it is a normal, distress phase or a casualty situation in order for the required alerts to be triggered or not.

The concept of situation awareness, which is described as the “knowledge of the state and the physical location of people within an area of interest”, aids understanding/recognition of the situation (Lass et al. 2008). The understanding of the processes in a response system, its history and likely future states – emotions and activities performed, is key to the control process to keep track of communications in the system (Lundberg and Asplund 2011). The states referred to in this study’s context are the human emotional states and the human activities being performed during those emotional states. Situation awareness is thus like a vigilance process that hugely relies on communication as well as information management within the response system. In addition, it is strongly linked to the ability of the emergency response system to predict future outcomes from the combination of emotional states and human activities. White et al. (2011) proposed that situation awareness can further be subdivided into the perception of the problems to be solved, predicting future events/outcomes (which could be achieved using machine learning methods) as well as understanding the implications, bearing in mind that timing is critical in an emergency response system while the stakes are often high and the conditions/circumstances do change.

3.2.2 Coordination and Communication

A well-designed response system should also support shared cognition. This requires tools and structures in the system to be thoroughly coordinated to enable the seamless functioning of technology and people. The architecture of the response system will largely determine its ability to handle coordination in the system while communication and information handling, including an active information system, requires efficient tools and structures in the system. Therefore, emergency response systems need to be carefully designed to handle coordination and communication. For coordination to be successfully achieved, an effective communication within the system is essential (Louise 2007).

Sensor networks, mobile technology, Geographic Information Systems (GIS), peer-to-peer communication platforms, procedural programming and wireless technology are some of the ICT technologies that have evolved over the past years, aimed at

supporting coordination and communication in an emergency response system (Chen et al. 2011). Most of these technologies are involved in our proposed architecture shown in Figure 3.1. The harvesting of data, processing and communication of information is anchored by these technologies.

However, the generic architecture for the design and implementation of the procedures for the 3 experimental models consisting of the 45 experiments that is conducted in this study is shown in Figure 3.2. The first step in the architecture is the DEAP data set that is used for the experimentations. Acquisition of the physiological signals as the experimental data from the DEAP data set constitute the next step which was followed by the data preprocessing procedure wherein the inverse Fisher transform algorithm and mapping of the transformed data to image space was conducted. Feature extraction using the signal and digital image processing algorithms, including the HOG, LBP and Histogram of Images was subsequently applied. The extracted features were passed to a RBFNN pattern classifier for training or testing and the training or testing result of human emotion recognition is obtained.

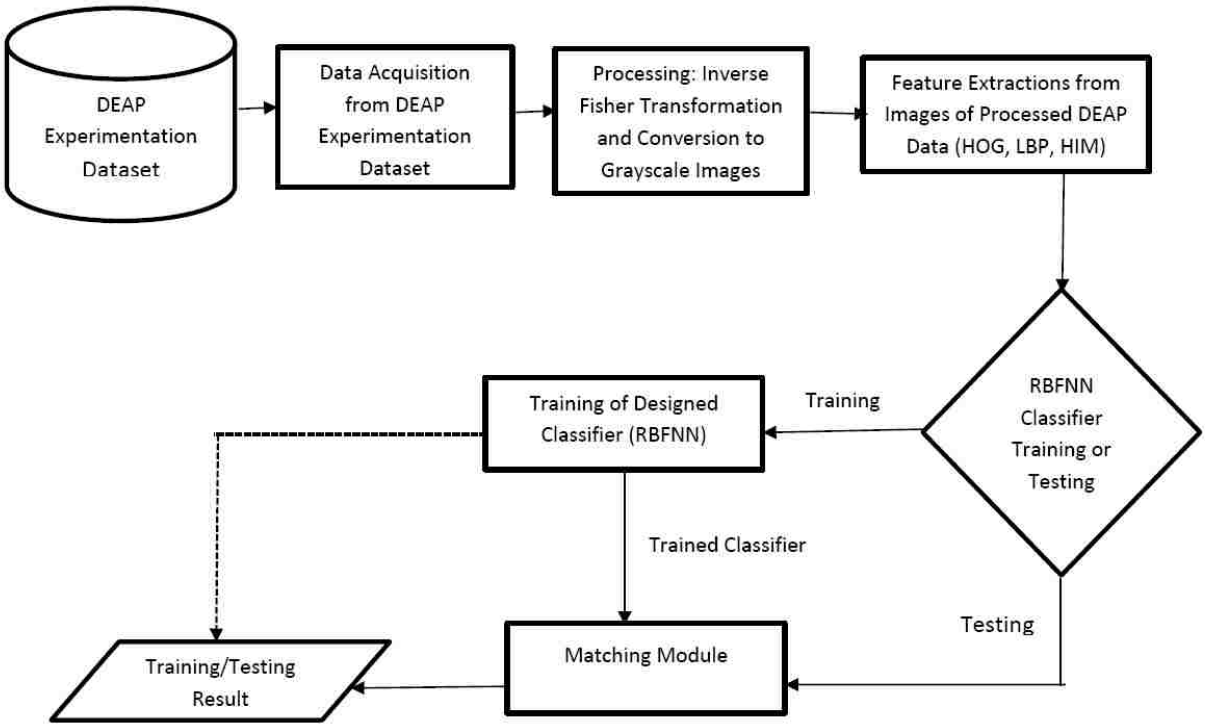


Figure 3.2: The generic architecture of the human emotion recognition model (author's own craft).

3.2.3 Human emotions and human activities

Philosophers, neurologists and psychologists have investigated relationships between human emotions and actions taken, as a result of the experienced emotions. It was discovered by De Sousa (1987) that human emotions hugely influence the action generation processes, including how these actions are executed and controlled. The speed of execution and efficiency were identified (De Sousa 1987) as benefits of the emotions based decision-making, since human emotions significantly contribute to and determine all the available actions for evaluation. Human emotions are not necessary actions, but cannot be isolated from understanding the nature of actions taken. As opined by Damasio (1999), human emotions lie between deliberate action and involuntary physiological processes of the body. The theory of emotion coherence was thus utilised by Thagard (2001) to develop an affective model that investigates emotion based actions, and human actions based on deliberate decision-making processes.

3.3 Signals Measurement

The cooking hacks' e-Health Sensor Platform V2.0 for Arduino and Raspberry Pi (Biometric/Medical Applications), as shown in Figure 3.3, is manufactured by Cooking Hacks (2017). It contains sensors to measure the electrocardiogram, body temperature, Galvanic Skin Response (GSR – sweating), airflow/breathing as well as the pulse and oxygen in the blood (Sp02).

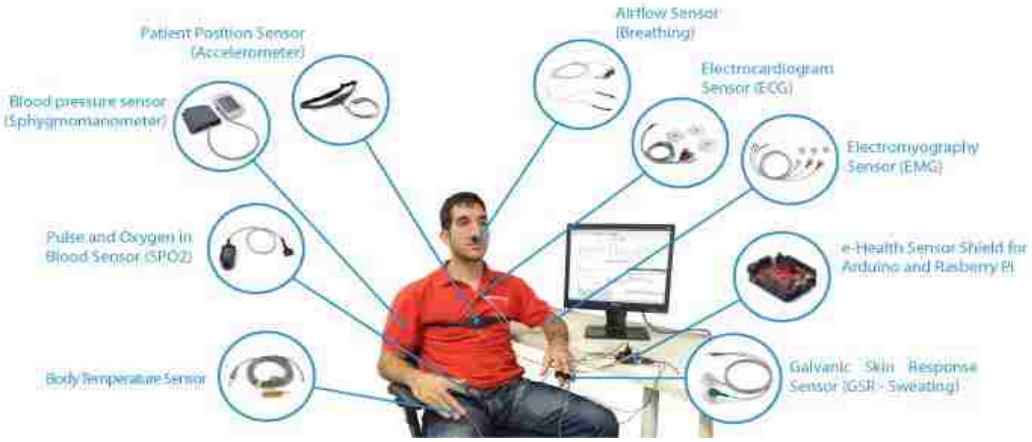


Figure 3.3: e-Health Sensor Platform V2.0 (Cooking Hacks 2017).

The e-health platform, when mounted on the Arduino Uno processor, is as shown in Figure 3.4.

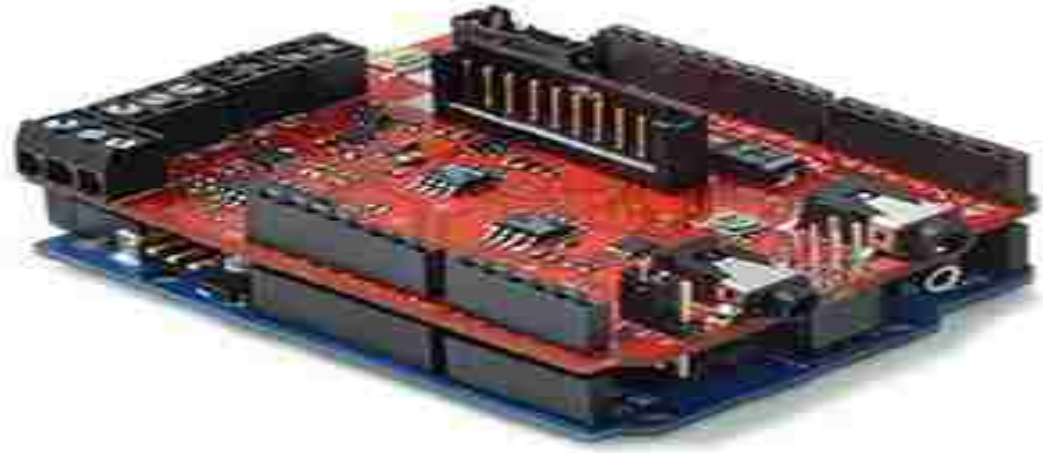


Figure 3.4: e-Health Sensor Shield over Arduino (Cooking Hacks 2017).

The signals relating to the recognition of emotions include heart rate, the body temperature, airflow/breathing and Galvanic Skin Response (GSR – sweating) among others.

In a study to identify a stress situation, Healey and Picard (2005) continuously recorded Electrocardiogram (ECG) signals, Electromyography (EMG), Heart Rate (HR), respiration and skin conductance data from drivers while performing a real-life driving task. The data were analysed using 5 minute data intervals collected at rest, city driving, and highway driving to recognize three different levels of a driver's stress – low (rest), medium (highway), and high (city). A classification accuracy of 97.4% was recorded for multiple drivers and different driving days, using a linear discriminant classifier and a fisher projection matrix. A metric of observable stressors was also created for the driving duration, using features of the continuous physiological data calculated at 1s intervals to measure the correlations between the data. The authors observed that skin conductance and heart rates were the most correlated physiological data for drivers' stress detection.

In related research by Rigas (2011), a methodology for a driver's stress and fatigue detection was presented along with driving performance prediction. Statistical features were extracted from the driver's physiological signals, consisting of ECG, Electrodermal Activity (EDA) and respiration. Along with video recordings of drivers' faces and environmental situations, a classification accuracy of 88% and 86% was obtained with three fatigue levels consisting of normal, low fatigue, and high fatigue; and an alternate classification of two fatigue levels namely normal and stress respectively.

Lanatà et al. (2015) reported an Autonomic Nervous System (ANS) fluctuation amidst driving style modifications in response to stimulated incremental stress load during simulated driving. The driving sessions involved steady motorway driving and two other sessions consisting of added arithmetic questions and vehicle mechanical stimuli that induced an incremental stress load. The respiration activity, Heart Rate Variability (HRV) and EDA response physiological data of 15 participants were measured. The physiological data in addition to the vehicle's velocity, steering wheel angle and time responses recorded significant changes in stress levels during the three driving sessions. A recognition accuracy of over 90% was obtained using a pattern classification algorithm.

Other signals such as ambulatory activities and location information may also be measured using the accelerometer sensor and GPS/GPRS devices. The signals measured and the relevance of these signals in human emotions, as well as the body organs, mainly connected to the signals is hereby discussed.

3.3.1 Heart Rate

The ECG sensor measures heart activities. The activity level of an individual can be read from his heart rate by monitoring the contractile activity of the heart. The organ in the human body that is charged with the responsibilities of supplying blood, oxygen and nutrients to all parts of the body is the heart (Tu, Inthavong and Loong 2015). It is surrounded in the chest region of the body by the pericardium, which is a double wall sac containing fluid.

The heart is the centre of the cardiovascular system pumping blood to the body's vital organs as well as providing oxygen and nutrients, while also removing harmful chemicals and other waste products from the organs. It has also been described as the human emotional centre, because of the relevance and relationship of heart rate changes to human emotions. In fact, it wears a romantic "cloak" which links it to love, affection, hate, romance etc. all of which are human emotions.

The neurotransmitters in the human body are chemicals that transmit messages between the cardiovascular and the nervous systems. These chemicals are capable of triggering a response in the specific tissue as they travel between the cells. Norepinephrine is synonymous to the adrenaline hormone in the human body. It is a neurotransmitter capable of constricting the blood vessels, enhancing the contraction force of the heart and also increasing the heart rate. While experiencing fear, for instance, adrenaline is supplied in huge quantities to enhance the pumping of more blood by the heart to the body muscles, thus preparing the human body for the fight or flight response. On the other hand, the heart rate can be slowed down by acetylcholine, which is another neurotransmitter in the body.

3.3.1.1 Link between heart rate and emotional states

Emotional states such as fear, anger, stress, and fatigue have been shown to be closely linked to heart rate (Healey and Picard 2005; Collet et al. 2009; Schneegaas et al. 2013; Solovey et al. 2014). A lower heart rate can be noticed in subjects in a relaxed and happy emotional state while an increase can be associated with unpleasant feelings such as fear and anger. Therefore, monitoring the ECG signal is essential in detecting emotional stress levels.

Research studies (Grossman et al.1996; McCraty and Childre 2004; De Witte et al. 2016) have shown that emotions and heart rates are inextricably entwined. The research by Grossman et al. (1996) investigated fluctuations in cardiac vagal control in patients with coronary heart diseases and reported cardiovascular responses to behavioural stress as responsible for the observed fluctuations in heart rate. Human emotions, including anxiety, fear and anger, have been shown by McCraty and Childre (2004) to cause irregular changes in heart rate while positive valence emotions such as appreciation and compassion result in a regular and orderly heart rate. The research study by Jonghwa and Andre (2008) also highlighted the inter-beat interval, heart rate variability and heart rate as common features of electrocardiogram signals. They stated that the heart rate mirrors emotional activity and utilised it to distinguish negative and positive valence emotions, while the heart rate variability was employed as an indicator of stress and mental workload in adults.

Studies also indicate that regulating the heart rate could influence the emotional state. In a study by Peira, Pourtois and Fredrikson (2013), a learned heart rate regulation method with biofeedback transfers to emotional situations without biofeedback was investigated. Biofeedback was used to train participants to decrease their heart rate and thereafter they computed the inter-participant differences of the acquired skills to predict their expertise in decreasing heart rate reactivity to being stimulated with negative arousing pictures without feedback. It was proposed that the learned ability could be transferred to emotionally challenging conditions without biofeedback, thus enabling regulation of bodily aspects of human emotions when heart rate biofeedback is unavailable.

This means that the heart rate is an important indicator of specific emotions. In addition, Ménard et al. (2015) recognised six basic emotions, using heart rate and skin conductance signals that allowed the identification of emotional states with a good accuracy ratio in a market research context. A partial correlation was also observed by the authors between objective and subjective data.

De Witte et al. (2016) proposed that an autonomic nervous system, such as a vagally mediated heart rate variability and interoceptive sensitivity could have an impact on human emotion regulation. In the study, the connection between self-reported emotional regulation methods and heart rate variability, when at rest, was investigated. A high heart rate variability was found to be related to an increased utilization of external emotional regulation methods.

Dimitriev, Saperova and Dimitriev (2016) among other studies, demonstrated the impairment of heart rate variability in humans as a result of experiencing anxiety emotion. Changes in the state of anxiety were discovered as nonlinear dynamics of heart rate.

These studies have confirmed that emotions and heart rate are strongly related, hence the choice of the heart rate as a physiological signal in this study, for the purpose of human emotion recognition to be used for the tracking of persons in a distress phase situation.

3.3.1.2 Sensors for measuring heart rate

The various methods and devices, include sound/acoustic, electrodes, pressure, and light sensors have been used to measure Heart Rate (HR). The observed fluctuations in the optical characteristics of human blood is used by the light sensor to measure the heart rate while the variations in the sounds produced around the heart were used by the acoustic sensor for the measurement. The pressure sensors on the other hand utilised vibrations produced by the heart to measure the HR.

(i) Acoustic sensors

Phonocardiography (PCG) is a popular and non-invasive method adopted in HR measurement using sound technology by listening to the heart's sounds. The PCG has been used in some cases to diagnose cardiac conditions. A full heart cycle is made up of two major sounds namely S1 and S2. The piezoelectric transducers are utilised in detecting vibrations and sounds produced by the heart, the muscles' activities, blood flux and valves by placing sound sensors on the human chest where the audibility of the sound is the highest. HR is computed from the PCG by detection of the first and second heart sounds (S1 and S2) having segmented the sound to heart cycles and identifying the different cardiac phases. A number of studies (Gamero and Watrous 2003; Brusco and Nazeran 2005; Yamacli, Dokur and Ölmez 2008; Zhang et al. 2010; Chen et al. 2011; Chen et al. 2015) used different approaches to classify S1 and S2 acoustically.

A statistical method utilizing the Hidden Markov Model was used by Gamero and Watrous (2003) to classify S1 and S2 sounds while using a data set of 80 subject's, consisting of 20 seconds recordings each, to achieve a 95% sensitivity result. In the study reported by Brusco and Nazeran (2005), different heart sounds were classified while utilizing the peaks obtained from the normalized Shannon energy. The distances between the peaks obtained were measured and classification was done using a threshold while the S1 and S2 heart sounds were appropriately segmented. Heart cycles were detected at a 79.3% overall accuracy.

Data from 53 patients from whom 326 heart cycle's samples were harvested were used by Yamacli, Dokur and Ölmez (2008) to classify the S1 and S2 heart sounds from energy peaks detected by varying thresholds. A wavelet decomposition of the normalized input signal was performed before a window integration movement of the energy signal. The results obtained showed an 88.9% sensitivity for the S2 and 91.4% for S1 respectively.

An air conductive microphone was utilised by Zhang et al. (2010) through harvesting sound signals from the device and applying a novel algorithm to it in order to estimate the HR of a subject without direct contact with the skin. The HR measurements

were obtained with the subject wearing the device while performing some activities such as coughing, jumping, sitting and reading. Results obtained indicated the suitability of the method in estimating cardiac heart sound in a natural environment.

Chen et al. (2011) in another study, measured HR from PCG by utilizing a template extraction including the filtered input signals being matched with the templates. A root mean square (RMS) error of 2.4bpm (beats per minute) was obtained in the calculation of HR using this method and tested with three human subjects at rest position.

In Chen et al. (2015), a small sensor was placed on the neck to acquire acoustic signals to which a novel algorithm was applied to detect the heart sounds - S1 and S2 from which the HR signals was computed in a window of 60 seconds. The algorithm analysed the filtered input acoustic signal using Continuous Wavelet Transform (CWT) to harvest peak frequencies likely to be the S1 and S2 heart sounds. A classification accuracy of 90.73% and 90.69% respectively was achieved when compared and evaluated with HR values offered by PULSOX 300i pulse oximeter (Konica Minolta 2014) and SOMNOscreen by SomnoMedics (SOMNOmedics GmbH 2014).

(ii) Electrode sensors

The use of electrodes, combined with various conducting materials, for measuring HR has been documented by a number of researchers (Komensky et al. 2012; Nemati, Deen and Mondal 2012; Tseng et al. 2014; Lee et al. 2014; Andreoni, Standoli and Perego 2015; Majumder, Mondal and Jamal Deen 2017).

An ECG monitoring system was proposed by Komensky et al. (2012) without the driven right-leg (DRL) electrode circuit. A DRL electrode usually requires a long, wired connection as it is often positioned at a far distance from the measurement electrodes. The system proposed by Komensky et al. (2012) however, utilised two active capacitive sensors enclosed in a flexible band placed on the chest. The system offered good measurement of ECG signals from resting/stable participants even though the P waves were not distinguishable. This was traceable to the position on the body where the electrode was placed or the fact that the common electrode was not available. However,

during walking activity, the ECG measurements were badly impacted by artefacts from the movements, with the QRS complexes still recognizable.

ECG was also measured using a cotton T-shirt integrated with a stretchable belt wherein three capacitive electrodes were embedded (Nemati, Deen and Mondal 2012). The monitoring system derived is compact, wireless and low-powered with the shirt functioning as the dielectric between the human skin and the capacitive electrodes. To achieve low power consumption in the system, electronic components with low power features were utilised while a two-layer printed circuit board was used as a mount on which the signal processing and communication modules were placed. An idle mode sampling method was utilised for the signal as well as a low-powered, short-range wireless technology for sensor networks and similar applications to further reduce power consumption. Inconveniences may however be experienced by users as a result of the stiffness of the electrodes with the potentials of corrupting the harvested signals by motion artefacts.

Tseng et al. (2014) introduced a three-electrodes, elastic fabric based vest to measure the ECG signal. A data acquisition module is embedded in the vest and firm contact of the electrodes with the skin was established. A Ni/Cu (Nickel/Copper) coated compressed urethane polymer foam was used to fabricate the electrodes, which was then encased in an Au (gold), coated conductive taffeta fabric. The proposed system achieved a high correlation using the simulated signal while the electrode motion noise was greatly reduced as a result of the low electrode-skin impedance derived from the conductive characteristics of the substrate.

A system involving flexible capacitive electrodes was developed and embedded in a chest belt to monitor ECG (Lee et al. 2014). The motion artefacts of the electrodes were greatly reduced in the system at the input of the pre-amplifier by utilizing a very high bias resistor. The HR and QRS complex were determined from the harvested raw ECG data by applying a peak detection and noise cancellation algorithm on the raw ECG data.

The R-R interval, HR and ECG physiological signals were monitored in the study reported by Andreoni, Standoli and Perego (2015) using textile electrodes made from

silver rich conductive yarns. These electrodes were embedded in customized belts and T-shirts. Body sweat was used by the electrodes as an electrolyte medium for the enhancement of conductivity between the skin and the electrodes, as well as the physiological signals' quality. A low powered Bluetooth 4.0 was utilised to send the data harvested by the device, while an SpO_2 pulse oximeter sensor was also included in the device, as well as a trial axial accelerometer to detect fall.

The Ag-AgCl electrodes are the most popularly used devices for measuring the HR. They consist of the hydrogel or wet ECG method, which involves placing the electrodes on some, identified parts of the body, especially the chest and the lower part of the abdomen. The conducting gel in the electrodes serves as a conduction medium between the electrodes and the skin. The likely irritating, toxic and staining effects of conducting gel are making it unsuitable for a long duration ambulatory monitoring system, though they are being popularly used in measuring HR in other systems that require less individual daily activities. Majumder, Mondal and Jamal Deen (2017) comprehensively reviewed some of the studies carried out by researchers to measure or monitor ECG signals using electrodes as the electrocardiogram sensor of the cooking hacks platform also measures HR.

(iii) Light Sensors

Detection of variation in the human blood volume has been utilised to measure HR by using light emitting diodes (LEDs) and photodiode at the near infrared region. This method of using a light source to measure HR, as well as blood oxygen volume, is considered non-invasive, easy to use and with a compact size device, hence its popularity in pulse oximetry research studies.

Matsumura et al. (2014) investigated the measurement of Normalized Pulse Volume (NPV) and HR from reflection mode plethysmograms simultaneously obtained at the red, green and blue spectral regions. This was premised on the capability of utilizing the reflection mode photoplethysmography (PPG) to measure these physiological signals

by using the light sensors available in phone cameras and a light source embedded in a smartphone's flash.

The accuracy of the NPV and HR measurements when influenced by motion artefacts were analysed, while it was discovered that the signal-to-noise ratio (SNR) attained with the red-light PPG was lower than those of the green and blue light PPG. The green colour was suggested as the most suitable in measuring NPV and HR from the reflection mode PPG under motion artefact conditions.

The Fitbit Charge HR was described in Dooley, Golaszewski and Bartholomew (2017) as a triaxial accelerometer based device offering estimates of HR among other parameters such as sleep quality, distance travelled and calories extended. LED lights were utilised to measure human HR by the device through a technology termed "PurePulse". This requires the device to be worn on the wrist about 3 fingers above the wrist bone in order to enhance the accuracy achieved.

The Garmin Forerunner 225 device (Garmin 2016) was similarly experimented by Dooley, Golaszewski and Bartholomew (2017) and described as consisting of an accelerometer that offers sleep time and HR measurements. An inbuilt optical sensor utilizing the Mio HR technology that monitors HR using proprietary algorithms for LED light sampling is used to measure the HR at the wrist. The device also includes a Global Position System (GPS).

It was found out that the differences in measurement in the two devices Fitbit Charge HR and Garmin Forerunner 225 are not significant from the Polar HR monitor either during/start of moderate and vigorous intensities respectively.

(iv) Pressure Sensors

Many research studies have measured the HR using piezoelectric pressure sensors. These sensors detect the arterial pulse wave that is produced during the intermittent relaxation and contraction of the human heart.

A study reported by Wang, Jin and Li (2008) utilised pressure sensors to convert the vibrations caused by heartbeats into electric signals. ECG and Ballistocardiograph signals were harvested from 30 subjects and the heart rate signal was found to be synonymous with a one channel ECG.

An ear-worn reflective photoplethysmography (PPG) sensor was introduced by Patterson, Mcllwraith and Yang (2009) and involved attaching the sensor to the skin surface while targeting an artefact reduction of ambient noise. Ambient light was effectively separated from the preferred photoplethysmography signal by reducing wideband noise through the appropriate modulation of LED and a distinct integrating photocurrent demodulator. The experimental results confirmed the robustness of the sensor sensitivity to location and pressure variations.

In a similar study by Yoo and Lee (2011), the Pulse Rate Variability (PRV) obtained from the photoplethysmography signal was considered for measuring the Heart Rate Variability (HRV) in order to assess human stress. The practicality and accuracy achieved in using HRV obtained from the photoplethysmography, as well as fingertip PPG including the standard lead II ECG signals, were experimentally compared. The results obtained indicated the PRV obtained from PPG was a good alternative to HRV in stress assessment studies.

A piezoelectric pressure sensor attached to the human wrist was used by Yoon and Cho (2014) to estimate the HR through the detection of the pulse wave in the human artery. A polyimide substrate was used to manufacture the pressure sensor as a polyvinylidene fluoride – a trifluoroethylene (P (VDF-TrFE)) piezoelectric layer was used to spin-coat a thermally evaporated silver electrode. A mechanical stress is triggered on the piezoelectric layer as a result of the fluctuation in the radial artery, thereby causing a likely change across the electrode.

Park et al. (2015) developed a wireless HR monitoring device that used fluctuation in the pressure of the ear's canal surface to measure the HR. The in-ear pulse waves (EPW) was detected using a piezoelectric film pressure and converted the detected waves to an electric current. After preprocessing the EPW signal, a microcontroller was

utilised in the implementation of a knowledge-based algorithm to discover in real time, the pulse peak. The method was constrained by the fact that body movements can badly impact the pressure variance, as well as the peak height of the pressure waves, thus introducing error in the measurement of the HR.

In another study, a piezoresistive material was made by Tajitsu (2015), using an electro-spinning based non-woven acrylate-modified polytetrafluoroethylene (PTFE) fabric to measure HR. The sensor was inserted in a wristband while monitoring the HR, as the detected pulse wave signal pattern is synonymous with the electrocardiogram signal. The impact of body movements on the harvested signal was very little while a high accuracy was achieved.

3.3.2 Body Temperature

The human skin, shown in Figure 3.5, covers the entire area of the body, thus making it the largest and heaviest single organ of the body, with also the largest surface area of about 16m^2 (Shimizu 2013). The organ consists of tissue layers and offers protection to the other organs and muscles beneath it. The dermis and epidermis are the two main layers of the skin and are the inner and outer layers respectively necessary for maintaining critical functions. A third layer of subcutaneous tissue/ hypodermis is included in the structure and it is made up of connective tissues and fat. The skin is further classified into two types, hairy and non-hairy. The hairy skin has hair follicles and sebaceous glands while the non-hairy skin on the other hand lacks hair follicles and has a thicker epidermis.

The dermis provides softening, lubrication, water proofing and anti-bacterial functions while the epidermis maintains the skin moisture and guarding the body against pathogens, microbes and harmful external chemicals and influences.

The value of the body temperature depends on the parts of the body the measurement is made including skin surface, mouth, armpit and anus. Body temperature is significant as several illnesses as well as the causes, can be detected from changes in the body temperature. Researches have shown strong relationships between human skin

temperature levels and emotions and this physiological signal plays an effective role in measuring human sensations. Some of these research works are mentioned below.

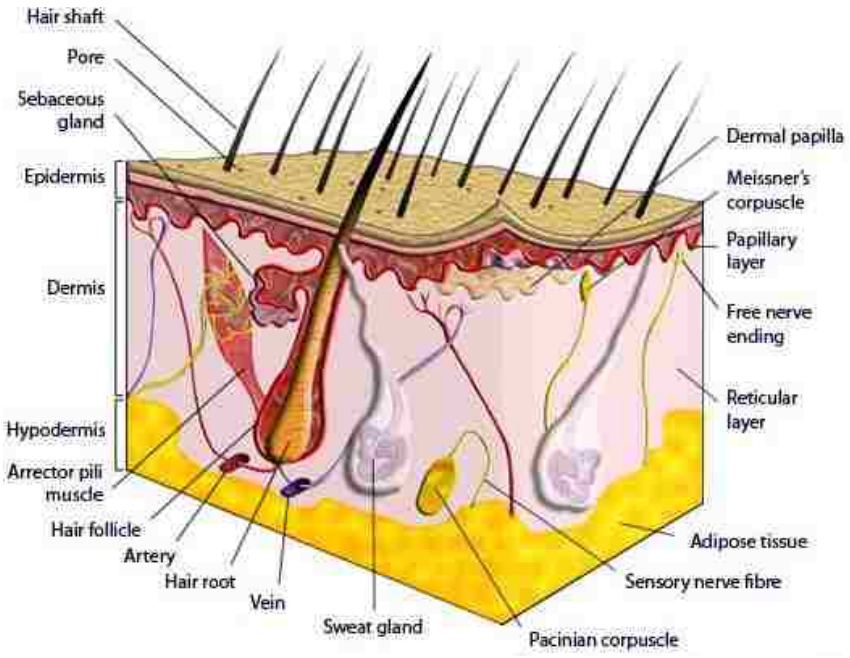


Figure 3.5: The anatomy of human skin (https://www.myvmc.com/anatomy/human-skin/).

A driver's level of awareness was investigated by measuring facial skin temperature of healthy volunteers during a simulated driving activity (Yamakoshi et al. 2008). The peripheral vasoconstriction and sympathetic nervous activity were discovered to be enhanced during a monotonous driving situation, thereby leading to a remarkable decrease in the skin temperature.

Cooper et al. (2014) investigated the influence of physiological responses of body temperature on other persons during social interactions. Thirty-six subjects watched and rated eight 3 minutes duration videos showing actors' left or right hands dipped in either cold or warm water. The temperature of observers' hands was measured using a thermistor as they watched the videos. For control purposes, other videos showed actors placing hands in front of water. The results obtained indicated that the videos displaying hands that were immersed in cold water were rated by subjects as being suggestively cooler when compared with the hands immersed in warm water. In addition, subject's hands were suggestively colder as they observed cold water videos than the hot water videos thus indicating a temperature contagion effect. This further suggests that the

emphatic consideration for primary low-level physiological challenges, including more complex emotions have somatic simulation fundamentals.

Salazar-Lopez et al. (2015) investigated the cognitive neuropsychology of emotions as a somatic marker of subjective experience during emotional tasks by using thermography. The results obtained indicated significant correlations between a mental state and fluctuations in facial temperature. It was observed that the nasal temperature decreased with negative valence stimuli, while the positive valence and arousal emotions caused an increase. The face, cheeks, orofacial area and the forehead are other parts of the body that witnessed these temperature variations. The “in the experiment” subjects’ empathy scores and performance were observed to be positively correlated with the nasal temperature changes, which suggest connections of bodily sensations with various emotional feelings, including love.

3.3.3 Respiration

Breathing is an essential human activity that must be performed continuously in order to safely engage in tasks. Respiration is connected to emotional states such as stress, fear and anger emotions, and could be associated with deeper and faster breathing. This is to trigger a fight or flight response, while rest and relaxation can result in shallow and slow breathing.

Physical kinetic exercise, increased mental workload or emotional excitement can also lead to faster and deeper breathing (Gorman 2004). It has been shown by (Karthik, Varghese and Chaitanya 2013) that breathing rate is correlated to stress levels hence its measurement is significant in stress detection.

The respiration sensor is most often placed on a subject’s diaphragm or thorax and measures the breathing rate as well as the deepness of breath. This placement is premised on the fact that the diaphragm contraction inflates the lungs, thus enabling it to perform a vital role in human breathing.

3.3.4 Galvanic Skin Response

The skin, as shown in Figure 3.5, has electrical properties that fluctuate relatively quickly and are strongly connected to the body physiological processes. It has also been shown that without applying an external current, electrical potential can be measured between two electrodes placed on human skin.

Human skin helps in conveying harmful materials from the body and passing it outside the body while also preventing harmful materials from entering the body. The eccrine and the apocrine are the two sweat glands in the human body. They have different functions, both produce sweat with varied chemical compositions that greatly influence the skin smell, called body odour.

The eccrine sweat glands in particular produce fluid called sweat containing water and electrolytes and are primarily responsible for thermoregulation. During hot conditions, sweat is released onto the skin surface and the body is cooled as the sweat evaporates. The glands are also present on the Palmar and Plantar surfaces and these glands are majorly triggered by emotional stress rather than body temperature (Hölzle 2002) with these surfaces much more reactive to observed psychological sweating than other parts of the body.

The Galvanic Skin Response (GSR) otherwise known as the Electrodermal Activity (EDA) of the skin or the skin conductance activity, signifies the electrical conductivity of the skin, especially the sweat gland, as a result of a subject's emotional state in response to his environmental events. It is a measure of emotion, attention and arousal in subjects and it is specifically associated with the sympathetic nervous system of the human body (Dawson, Schell and Filion 2017). The EDA as a physiological signal measure has been popularly studied and previously applied in anticipatory anxiety studies for determining stress levels, as a lie detector and in evaluating the difficulty of driving tasks (Helander 1978). In a related research work by Rigas et al. (2011), a methodology for a driver's stress and fatigue detection was presented along with driving performance prediction. Statistical features were extracted from the driver's physiological signals consisting of ECG, EDA and respiration. Along with drivers face video recordings and environmental

situation, a classification accuracy of 88% and 86% was obtained with three fatigue levels consisting of normal, low fatigue, and high fatigue; and two fatigue levels namely normal and stress respectively. Fatigue level was also confirmed to be closely associated with driving performance.

The eccrine sweating is regulated by the sympathetic Autonomic Nervous System (ANS), as the galvanic skin conductance physiological signal indicates the arousal of the sympathetic ANS, which is present in several psychological processes. The Galvanic Skin Response (GSR), Skin Conductance Response (SCR) and Electrodermal Activity (EDA) have been interchangeably used in various studies to indicate affective psychological processes, including habituation, cognitive effort, arousal and attention. The EDA indicates electrical occurrences in the skin with the SCR reflecting the capability of the skin to conduct electricity upon application of an external current of constant voltage.

Lanatà et al. (2015) reported an Autonomic Nervous System (ANS) fluctuation amidst driving style modifications in response to stimulated incremental stress load during simulated driving. The driving sessions involved steady motorway driving and two other sessions consisting of added arithmetic questions and vehicle mechanical stimuli that induce incremental stress load. The respiration activity, Heart Rate Variability (HRV) and electrodermal activity response physiological data of 15 participants were measured. The physiological data in addition to the vehicle's velocity, steering wheel angle and time responses recorded significant changes in stress levels during the three driving sessions. A recognition accuracy of over 90% was obtained using a pattern classification algorithm. A higher resistance is recorded by a dryer skin than wet skin as the psycho galvanic reflex of the body is measured by the skin conductance response sensor.

The skin conductance measurement constitutes a component included in polygraph devices and the physiological signal has been used in emotional or physiological arousal research studies. Shock, excitement and stress are emotions that can cause fluctuations in skin conductivity. The unit of measurement of the Cooking Hacks' eHealth GSR sensor is microvolt (μV) though the physiological signal is usually measured in micro Siemens.

Najstrom and Jansson (2007) investigated the predictive value of the skin conductance response (SCR) to pre-attentively perceived threatening pictures on emotional responses following stressful life events. The picture perception task was carried out on 136 police recruits while the skin conductance data were concurrently harvested. Psychological distress was measured after the recruits were exposed to emotionally stressful life events and tasks. The assessment was made by administering the impact of event scale of the subjects after 24 months of taking the SCR data. The results obtained showed that emotional responses to masked threatening pictures in relation to neutral pictures of stressful life events could be strongly predicted by enhanced SCR.

Hein et al. (2011) suggested that people, when empathizing with the pains of another person, display autonomic responses. The authors painfully stimulated some participants or made them observe pains being inflicted on another person, while their SCRs physiological signals were measured concurrently, as well as their affective ratings. The results indicated the prediction of later decisions on costly helping that is, choosing to suffer pains yourself as a result of others experiencing pains, depends on the strength of empathy-related vicarious SCR. A person is also more likely to engage in costly helping if the match between SCR magnitudes while observing pain experienced by others and during self-pain inductions are higher.

In a study by Walla, Brenner and Koller (2011), the hypothesis of the objective assessment of the occurrence of individual likes and dislikes with respect to brand attitude was tested. Individual rated common brands, in relation to subjective preferences. This was followed by the virtual display of the brand names liked and disliked by the individuals, while collecting their skin conductance, heart rate and electromyogram physiological signals, using the participants' eye blinks to the acoustic startle probes as responses. Results obtained suggest that the SCR and HR were lower in the case of liked as opposed to disliked brand names, thus highlighting the existence of emotion related differences in the liked and disliked brands by individuals.

Research on electrical variations experienced in human skin and its physiological effects have dated back a century. The possibility of measuring fluctuations in electrical

potentials between electrodes positioned on the surface of the human skin has been studied over time (Grimnes et al. 2011) as momentary variations in skin resistance could be noticed when elicited by various stimuli. The fluctuations in skin electrical potentials as measured by the galvanic skin response/skin conductance response is very important in emotion studies. The two databases of physiological data, collected during real-life driving tasks and publicly available, confirms the real life applications in which the GSR signals are measured.

These datasets include the stress recognition in automobile drivers' (physionet 'drivedb') produced by the MIT Media Lab (Healey and Picard 2005) and the hciLab driving data set (Schneegaas et al. 2013). The hciLab driving data set consists of contextual data (GPS position, brightness and acceleration), video rating of the driving scenarios as well as physiological data of 10 participants comprising the skin conductance response, skin temperature and the Electrocardiogram (ECG) readings of heart rate and heart rate variability data of the driver. The stress recognition in automobile drivers' (physionet 'drivedb') data set on the other hand consists of the Electrocardiogram (ECG), Electromyogram (EMG), heart rate, respiration and, the Galvanic Skin Response (GSR) of the hand and foot data respectively. These data sets have been utilised with good results in research studies as well as to indicate the relevance of the GSR/SCR for emotion recognition.

However, rather than attempting to collect the various physiological signals to be used for experimentations in this study, since the quality and integrity of data and features are very essential in machine learning studies (Jonghwa and Andre 2008; Solovey et al. 2014), it is opined that it would be ideal to first demonstrate the proposed techniques in this study with a renowned and publicly available standard data set such as the DEAP data set in order to determine the results that would be obtained.

This will give an insight and leeway during real-time deployment on human beings while using the physiological data that would be eventually collected. It is often best practices as also demonstrated in biomedical and behavioural research studies to first undertake preclinical studies involving the use of animals before extending the studies to clinical phase involving human subjects (Hamdam et al. 2013; Redfern et al. 2017).

Hence, the leverage on this, to first conduct experiments offline and evaluate results using existing and standard physiological signals data set of DEAP.

Some of the research studies that have utilised the DEAP physiological signal data set are briefly discussed below. However, it was observed that varied methods and results were recorded in these studies, as the issue of feature extraction which is germane for performance in a pattern recognition system still remain open. This necessitated an attempt in this study to undertake feature engineering using the DEAP physiological data set to compare the proposed method, features and results obtained with those in the literatures.

Wang and Shang (2013) introduced a Deep Belief Networks (DBNs) based system that automatically extracts features from 4 channels raw physiological data consisting of 2 EOG and 2 EMG channels respectively under an unsupervised scheme while building 3 classifiers that predict human emotion along the arousal, valence and liking classes. The classification accuracies obtained were 60.9%, 51.2% and 68.4%, which compares favourably with the results achieved with the Naïve Bayes classifier.

In another study, Li et al (2015) adopted a two hidden-layer Deep Belief Network architecture configured with visible and hidden nodes as 128-10-10 to classify EEG signals of the DEAP data set into a binary label scheme of valence, arousal, dominance and liking respectively using unsupervised training and future learning. Classification experiments were done along individual subjects as well as across all subjects. An SVM classifier was also applied to the power spectral density as well as the DBN features in order to compare the manually extracted features with the DBN features. Recognition accuracies of 58.2% (valence), 64.3% (arousal), 65.1% (dominance) and 66.3% (liking) was achieved with the PSD features across the all subjects while about 58.4% (valence), 64.2% (arousal), 65.8% (dominance) and 66.9% (liking) was recorded for the DBN features. There is no significant difference between the results of the two features as the possibility of learning affective features through a deep learning and manually generated features approaches were explored.

Zhuang et al. (2017) utilised the DEAP data set and applied Empirical Mode Decomposition (EMD) method to extract first difference of time series, the first difference of phase, and the normalized energy features from EEG signals which were decomposed into Intrinsic Mode Functions (IMFs) and classification was done along the arousal and valence classes using the support vector machine classifier. Classification accuracies of 71.99% and 69.10% were respectively, achieved for the arousal and valence classes using 8 EEG channels, which are Fp1, Fp2, F7, F8, T7, T8, P7, and P8, while 72.10% and 74.10% classification accuracies for arousal and valence was however achieved with 32 EEG channels. These performances are better than the results obtained by the authors using other methods/features such as the fractal dimension, sample entropy, discrete wavelength transformations.

EEG signal characteristics such as spatial, frequency domain and temporal were integrated by Li et al. (2017) and mapped to a two-dimensional image from which EEG multidimensional feature images were built to represent varied emotions in EEG signals. A deep learning approach named CLRNN involving hybriding the Convolution Neural Networks (CNN) and Long Short-Term-Memory (LSTM) Recurrent Neural Networks (RNN) was then applied to the EEG MFI obtained from the DEAP dataset. With each subject, an average emotion classification accuracy of 75.21% was achieved, which is reported as better than the current state of the art approaches in emotion recognition domain.

However, the best result was achieved with a 2s time window frame as against the available 60s window size containing all the EEG data per sample. In addition, the results of other classification methods such as k-NN, random forest and support vector machine with the features and the method proposed by the authors are not up to the 75.21% recorded. But aside the computationally expensiveness of deep learning technique, a shallow machine learning approach opted for in this study because it is believed that the classification accuracy obtained by Li et al. (2017) can be improved upon if the discriminatory strength of features is enhanced through novel feature engineering methods attempted in this present study rather than relying on the auto-generation of features as obtained in deep learning approaches.

A multiple-fusion-layer based ensemble classifier of stacked autoencoder (MESAE) for recognizing human emotions was proposed by Yin et al. (2017) involving the identification of the deep structure based on physiological-data-driven approach. Stable feature representations of the physiological signals were obtained as the unwanted noise in the physiological signals' features were filtered by three hidden layers in each stacked autoencoder. The stacked autoencoder ensembles were achieved by using an additional deep model and the physiological features are divided into many subsets based on various feature extraction methodologies with each subset separately encoded by a stacked autoencoder. The derived SAE abstractions were merged based on the physiological modality to create six sets of encodings which subsequently served as input to a three-layer, adjacent-graph-based network for feature fusion whose features were used for human emotion recognition along the binary arousal and valence emotion states.

Average classification accuracies of 77.19% and 76.17% was achieved for the arousal and valence state respectively, using the MESAE scheme with deep learning classifier while the accuracies reached 84.18% (arousal) and 83.04% (valence) with ensemble classification schemes.

Alhagry, Fahmy and El-khoribi (2017) applied Long-Short Term Memory recurrent deep neural network of the EEG physiological signals in the DEAP dataset. An average recognition accuracy of 85.65% was achieved for the arousal class while 85.45% and 87.99% average accuracies were recorded for the valence and liking classes respectively.

In the research study carried out and reported by Menezes et al. (2017), features were extracted from EEG signals for affective state modelling using the Russell's Circumplex model. The support vector machine and random forest classifiers were applied to the EEG features of statistical measures, band power as well as higher order crossing extracted from the DEAP data set to classify human emotion into the valence and arousal classes. The highest classification accuracy was obtained with the bandwaves spectral power density features by the SVM classifier was 69.2% and 88.4% for the bipartite scheme of arousal and valence classes respectively while the tripartite scheme recorded 59.5% and 55.9%. However, the random forest classifier recorded its

best classification accuracy of 74.0% (arousal) and 88.4 %(valence) along the bipartite labeling scheme (high/low) with the statistical band waves features and 63.1% (arousal) and 58.8% (valence) with the same features but for the tripartite labeling scheme (high/medium/low). Diverse lower results were recorded for each and combined band waves (Delta (δ), Theta (θ), Alpha (α) and Beta (β)) by the SVM and random forest classifiers for the bipartite/tripartite labeling scheme and statistical band waves or spectral power density features.

However, a framework to automatically search for the optimal subset of EEG features using Evolutionary Computation (EC) algorithms including the Particle Swam Optimization (PSO), Ant Colony Optimization (ACO), Genetic Algorithm (GA), Simulated Annealing (SA) and Differential Evolution (DE) was introduced by Nakisa et al. (2018). This is aimed at removing inefficiency and redundancy resulting from the high-dimensionality, introduced by combining all possible EEG features. The framework used frequency, time and time-frequency domain features of EEG signals from which some discriminatory features were selected using the EC algorithm and the probabilistic neural network pattern classifier was applied to classify human emotions into four classes. The DE algorithm yielded the best recognition accuracies of 96.97% and 67.47% for the MAHNOB and DEAP datasets respectively, with the probabilistic neural network classifier over 100 iterations.

Though the results obtained are promising, the challenge with this framework and method is its computational complexity as it takes about 80 hours to achieve convergence. This is not ideal for a real-time situation where efficient, prompt and accurate classification is required.

The techniques including data transformation, feature extraction, feature selection and classification that are proposed in the research work conducted and being reported in this dissertation are hereby presented in Chapter 4.

CHAPTER FOUR

Research Methods

This chapter presents the development of a mathematical model for human emotion recognition, including processing of physiological signals for feature extraction and feature classification. The Radial Basis Function Neural Network (RBFNN) pattern recognizer as well as its various configurations that were applied in the experiments that were conducted using the physiological signals of the DEAP (Koelstra et al. 2012) data set is also introduced**.** The essential steps involved in the development of the emotion recognition model are hereby presented.

4.1 Emotion Data set Acquisition

The DEAP (Koelstra et al. 2012) data set was chosen for the experimentations reported in this dissertation, because of its multimodal characteristics, fairly high number of subjects (32) and emotional states (16) considered in building the dataset. The data set as well as the results achieved by Koelstra et al. (2012) provides a platform to measure any improvement that could be achieved with the data set and compare results, using various methodologies developed by researchers who used the same dataset.

For the DEAP dataset, video clips were used as stimuli to elicit human emotions from the subjects and their physiological data were concurrently collected. These stimuli are essential and were carefully chosen to mimic natural inducement of the target emotion in the subject. This ensured the collection of good quality physiological data associated with the target emotion. The data set was developed from 32 voluntary healthy participants (16 females) with ages ranging between 19 and 37 (mean = 26.9) as they watched 40 music video clips capable of eliciting the target/reported felt emotions of anger, contempt, disgust, elation, envy, fear, guilt, hope, interest, joy, pride, relief, sadness, satisfaction, shame and surprise.

These 40 validated emotional music video clips of one-minute duration each was chosen from an earlier pilot study containing 120 one-minute extracts of music videos rated by 14-16 volunteers.

The participants in the main experiment, who were induced with a target emotion at a time, did a self-report of their felt emotions across a discrete approach using an emotion wheel. They did so by selecting the appropriate emotion from the emotion wheel labels of anger, contempt, disgust, elation, envy, fear, guilt, hope, interest, joy, pride, relief, sadness, satisfaction, shame and surprise. The EEG, electrooculogram (EOG), electromyogram (EMG), Galvanic Skin Response (GSR), respiration amplitude, Blood Volume Pulse (BVP) and skin temperature physiological signals were acquired simultaneously while they were experiencing the affectionate. The physiological signals were acquired along with frontal face videos of 22 participants using various sensors and active electrodes with the Biosemi Active II system. However, only two modalities comprising the central nervous system, for instance the EEG, and the peripheral nervous system physiological signals consisting of the EOG, EMG, GSR, respiration amplitude, BVP and skin temperature data as well as a fusion of these two modalities were specifically considered in the study presented here. This is because; it has been shown in the literature that modality fusion is capable of improving the result of an emotion recognition/classification problem (Chanel et al. 2006; Chanel et al. 2009; Soleymani et al. 2012; Koelstra et al. 2012).

During each trial, participants also undertook and reported a self-assessment of their degrees of valence, arousal, dominance and liking on a continuous 9-point scale using a SAM while familiarity was rated on a 5-point scale. For each participant, there were 40 trials, which gives a total of 1280 samples for all the 32 participants.

4.2 Preprocessing of Emotion Physiological Signals

The DEAP raw physiological signals sampled at 512Hz have 48 channels per experimental trial consisting of 32 EEG channels, 12 peripheral physiological channels, 3 unused channels and 1 status channel. These channels were eventually reduced to 40

during preprocessing by merging the same channels such as EOG and EMG channels respectively, while eliminating the unused channels (Koelstra et al. 2012). Each sample of the physiological signals therefore contain 40 channels made up of 32 channels of EEG signals and 8 channels of peripheral physiological signals consisting of the EOG, EMG, GSR, respiration amplitude, BVP and skin temperature data which was down sampled to 128Hz. The preprocessed data was segmented into 60-second trials with a 3-second pre-trial baseline removed while the list of the 40 channels' physiological signals is shown in Table 4.1.

Table 4. 1: Channels' details and categorization of the DEAP physiological data set.

Channel Number	Channel Name	Channel Description/Location	Signal Category
1	Fp1	Left Anterior	EEG
2	AF3	Left Anterior	EEG
3	F3	Left Anterior	EEG
4	F7	Left Anterior	EEG
5	FC5	Left Anterior	EEG
6	FC1	Left Anterior	EEG
7	C3	Left hemisphere	EEG
8	T7	Left hemisphere	EEG
9	CP5	Left Posterior	EEG
10	CP1	Left Posterior	EEG
11	P3	Left Posterior	EEG
12	P7	Left Posterior	EEG
13	PO3	Left Posterior	EEG
14	O1	Left Posterior	EEG
15	Oz	Posterior Hemisphere	EEG
16	Pz	Posterior Hemisphere	EEG

17	Fp2	Right Anterior	EEG
18	AF4	Right Anterior	EEG
19	Fz	Anterior Hemisphere	EEG
20	F4	Right Anterior	EEG
21	F8	Right Anterior	EEG
22	FC6	Right Anterior	EEG
23	FC2	Right anterior	EEG
24	Cz	Centre	EEG
25	C4	Right hemisphere	EEG
26	T8	Right hemisphere	EEG
27	CP6	Right hemisphere	EEG
28	CP2	Right hemisphere	EEG
29	P4	Right posterior	EEG
30	P8	Right posterior	EEG
31	PO4	Right posterior	EEG
32	O2	Right posterior	EEG
33	hEOG	Horizontal EOG ($hEOG_1$-$hEOG_2$)	EOG
34	vEOG	Vertical EOG ($vEOG_1$-$vEOG_2$)	EOG
35	zEMG	Zygomaticus major EMG ($zEMG_1$-$zEMG_2$)	EMG
36	tEMG	Trapezius EMG ($tEMG_1$-$tEMG_2$)	EMG
37	GSR	Galvanic skin response (left middle and ring finger)	GSR
38	Respiration	Respiration amplitude belt	Respiration
39	Plethysmograph	Blood volume pulse	BVP
40	Temperature	Temperature of left pinky	Body temperature

To process the data, firstly, the raw physiological data is normalized. There are different normalization schemes available in the literature amongst which the Min-Max and Z-score are famous. In particular, the Z-score has been used in the study of physiological data (Chubb and Simpson 2012) because it has the capability to dramatically simplify clinical interpretations (Colan 2013). However, both the Min-Max and Z-score normalization have received criticisms as they are both sensitive to outliers and it is not always that their performance is excellent (Singh and Gupta 2007). In addition, the Min-Max normalisation scheme scales data to a fixed range of 0 to 1 thereby giving smaller standard deviations. But the Z-score normalization is often preferred to the Min-Max scheme, especially when applied with Principal Component Analysis (PCA) procedure in order to compare similarities between features as the component that maximizes the variance is often the focus (Raschka 2014). The tanh estimator, therefore, has been suggested as a robust scheme in place of min-max and Z-score because of its robustness, efficiency and elegance (Singh and Gupta 2007). However, the tanh normalization which is adjudged better than the Z-score and Min-Max also has a limitation as the conventional tanh is scaled in the interval [-1,1], which is not desirable in the image processing domain where pixel values are usually between [0,L], where L ϵ (1, 255). The shifted tanh function is introduced in this thesis to correct the limitation of the tanh function.

Therefore, as part of the distinctive contributions to knowledge, the shifted tanh function is introduced. The general tanh function which lies in the interval [-1, 1] is given as:

$$t\big(u(x)\big) = \frac{e^{u(x)} - e^{-u(x)}}{e^{u(x)} + e^{-u(x)}} \qquad \textbf{4.1}$$

where

$$u(x) = \frac{(x - \bar{x})}{\sigma_x} \qquad \textbf{4.2}$$

is the Z-score that ensures a normalization process with a mean ($\bar{x}$) of zero and standard deviation (σ_x) of one. This approach allows the integration of the benefit of Z-score in the tanh function.

The expression in Equation (4.1) can be further simplified by multiplying the right hand side by e^x and at the same time divides it by e^x to give;

$$f\big(u(x)\big) = \frac{e^{2u(x)} - 1}{e^{2u(x)} + 1} \qquad \textbf{4.3}$$

The expression in Equation (4.3) corresponds to the inverse Fisher transform that has the advantage that it is compressive and for large absolute values, the output is compressed to 1 at most while also removing low amplitude variations. The inverse Fisher transform is analogous to edge sharpening in digital image processing. Moreover, it is the exact solution of the standard Fractional Riccati Differential Equation (FRDE) (Salehi and Darvishi 2016) of the form;

$$D^{(\alpha)} f(t) + y^2(t) - 1 = 0;\ \ t > 0,\ \ 0 < \alpha \leq 1 \qquad \textbf{4.4}$$

where

$$\alpha\ = 1, \mathrm{f}\,(0)\ = 0 \qquad \textbf{4.5}$$

The values of f (u(x)) in equation (4.3) still lie in the interval [-1, 1], but it is desired to have a normaliser that computes values in the interval [0, L], where L is the maximum greyscale value such as 255. The addition of 1 to both sides of Equation (4.3) will give:

$$1 + f\big(u(x)\big) = \frac{2}{1 + e^{-2x}} \qquad \textbf{4.6}$$

The expression given by Equation (4.6) will compute values in the interval [0, 2]. To achieve the desired goal of having a normaliser that compute values in the interval [0, L], we multiply both sides of Equation (4.6) by L and at the same time divide by 2 to have:

$$\frac{L(1 + f\big(u(x)\big)}{2} = \frac{L}{1 + e^{-2u(x)}} \qquad \textbf{4.7}$$

By substituting Equation (4.2) into Equation (4.7) and dividing the left hand side of Equation (4.7), a more robust normaliser is obtained as follows:

$$F\big(u(x)\big) = \frac{L}{1 + e^{\frac{-\mathrm{k}(x-\bar{x})}{\sigma_x}}} \qquad \textbf{4.8}$$

Equation (4.8) is the desired normaliser, which is a particular form of the growth function with L being the modifier, $\bar{x}$ is the data mean, σ_x is the data standard deviation and k=2,3,4,5… while the value of F(u(x)) lies in the interval [0, L]. This function corresponds to the inverse Fisher normaliser with k=2.

After the normalization and applying the inverse Fisher transform of the physiological data, the transformed data were thereafter mapped to greyscale image space while forming hyperspectral images. This is to enhance the digital image processing techniques which are intended to be applied to feature extraction. The Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP) and Histogram of the Images (HIM) features were then extracted from the inverse Fisher transformed physiological data. These features are called 'local' features as they were extracted from each channel of each sample.

The step by step procedure for the preprocessing stage and extraction of the features is listed below.

Step 1: Read the raw physiological data from the DEAP dataset

Step 2: Determine physiological data class using emotion representation

Step 3: Preprocess the physiological data by applying inverse Fisher transformation

Step 4: Map the transformed physiological data to hyperspectral images

Step 5: Extract features from the hyperspectral images using different standard algorithms of digital image processing techniques – HOG, LBP and HIM

Step 6: Apply principal component analysis on the extracted features to compute dimensionally reduced Eigen features

Step 7: Select the desired Eigen features using the Kaizer criterion of Eigenvalues greater than one

Step 8: Use pattern recognizer to recognize the selected Eigen features as Happy, Distress and Casualty

This preprocessing and feature extraction procedure in conjunction with the implementation are some of the contributions of this study. The program codes, running or execution and implementation was done using MATLAB 2018a environment.

4.3 Emotion Class Representation

In the DEAP dataset, there are 32 participants who undertook 40 trials each thus making a total of 1280 physiological samples/observations available for analysis. Each sample of the physiological signals was labelled by the participants along the dimensional approach for emotion representations, including valence, arousal, dominance and liking on a 9-point continuous scale ranging from 1-9 using a Self-Assessment Manikin (SAM). These four emotion representations quantitatively describe emotions. The valence scale ranges from happy/joyful/positive/pleasure to unhappy/sad/negative/displeasure; arousal scale ranges from calm/bored/low to excited/stimulated/high; dominance scale from submissive/without control/overpowered to dominant/in control/empowered; and liking ranges from like to dislike (Soleymani et al. 2012; Koelstra et al. 2012).

Valence scale measures the pleasantness or unpleasantness feelings of an emotion and can include happy, joy, peaceful and cheerful emotions for the pleasant/positive emotions and sad, fear, stress and angry emotions for the unpleasant/negative emotions (Soleymani et al. 2012; Koelstra et al. 2012). On the other hand, the arousal scale measures the intensity of emotional feelings and, such measures include both for positive and negative valence emotions. For instance, both pleasure and joy emotions have a positive valence, but low and high arousal respectively, while both sadness and anger have negative valence but low and high arousal respectively. The valence and arousal scales may be represented either way on the vertical/horizontal axis in the valence-

arousal representation plane. This is to enable an easy four quadrants categorisation, including Low Arousal, Positive Valence; High Arousal, Positive Valence; Low Arousal, Negative Valence and High Arousal, Negative Valence as shown in Figure 4.1. But to map the ratings of participants in the DEAP data set with the quadrant, since the quantitative score ranges from 1-9 and no negative figures recorded, we therefore map and represent Negative Valence (NV) as Low Valence and Positive Valence (PV) with High Valence as shown in Figure 4.2.

Dominance denotes the control ability expressed by an individual under a certain emotional state. Dominance and liking emotion models can occur in any emotion type, whether it is a positive(high) or negative(low) valence or low/high arousal emotions (Koelstra et al. 2012). For instance, an individual experiencing joy emotion may have a weak feeling or helpless indicating that he is not in control or has been overpowered by the emotion while another individual experiencing an anger or sadness emotion may still have a strong feeling and be in control. Furthermore, the liking scale does not also always connote a positive valence scale because an individual may like a sadness or anger emotion, which are both negative valence emotions, as well as any materials capable of eliciting these emotions and vice-versa (Koelstra et al. 2012) even when negative valence emotions are described to be characterized with unpleasantness/displeasure.

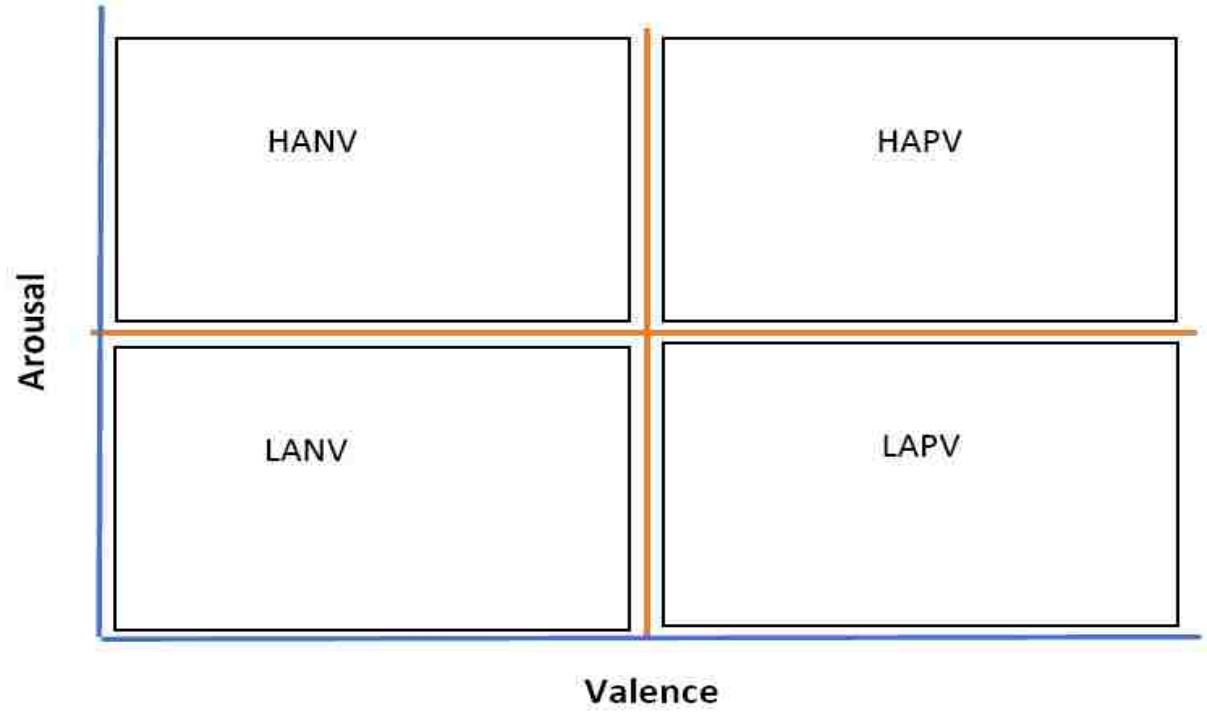


Figure 4.1: The valence-arousal plane in HANV, HAPV, LANV and LAPV dimensions.

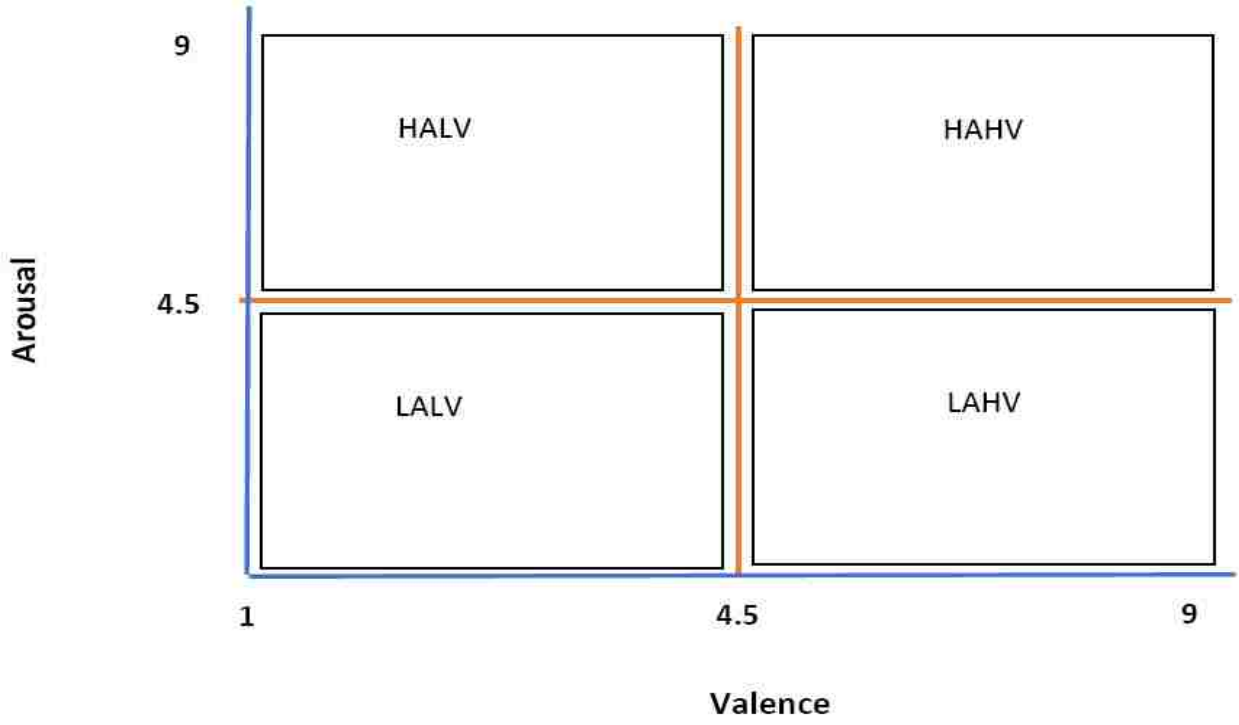


Figure 4.2: Participants' ratings mapped to HALV, HAPV, LALV and LAHV dimensions.

Therefore, a model of human emotion recognition in a distress phase situation problem is proposed along the four emotion representation labels of valence, arousal, dominance and liking as well as a fifth label tagged Distress Phase, which is obtained from the combination of the four emotion representation labels namely valence, arousal, dominance and liking. This entails tackling five classification problems, four of which are binary classifications including low/high valence, low/high arousal, submissive or in control dominance as well as like/dislike for the four labels, all of which recognition results were subsequently obtained. However, the Distress Phase representation has three emotion classes, which are Happy Phase, Distress Phase and Casualty Phase, which were built from the valence, arousal, dominance and liking labels' ratings in the DEAP physiological signals dataset. In addition, the discrete emotions were appropriately mapped to the identified three classes of happy, distress and casualty phases of the emotion model. This was achieved by drawing inspirations from the target/felt emotion wheel namely anger, contempt, disgust, elation, envy, fear, guilt, hope, interest, joy, pride, relief, sadness, satisfaction, shame and surprise reported in the DEAP data set to enable an insight to appropriately categorize these discrete emotions along the valence-arousal-dominance-liking dimensional space while also utilizing the participants' ratings.

Since many people use different words to mean the same emotional feeling, in order to enhance the generalization capability of the Distress Phase emotion model in terms of definitions and meanings of emotions, the emotion wheel in the DEAP data set was extended to the Funto Emotion and Feeling Wheel (Chadha 2016) as shown in Figure 4.3, which contains some other words, meanings, characteristics and features of different emotional feelings. The Funto Emotion and Feeling Wheel (Chadha 2016) was considered over others as a complement to the one utilised in the DEAP data set because it attempt to resolve some inherent challenges in other past emotion wheels that contains more negative emotions than positive ones. In addition, many scientifically identified feelings and emotional words were not on the other emotion wheels and from real world experience, there are many "emotions" that people are identifying but were not included in any, or most, of the other emotion wheels (Chadha 2016). Furthermore, apart from offering a huge number of emotions/affective words to choose from, to compare and

relate to those in the DEAP dataset, the Funto emotion and feeling wheel also have all the six basic human emotions.

Thus, a leverage on Russell (1980) mapping of human emotions and affective words to the valence-arousal dimensional plane as shown in Figure 4.4 was done so as to enable the Distress Phase emotion model, after taking due cognizance of the participants' ratings to appropriately map a detected human emotion into the valence-arousal dimensional plane.

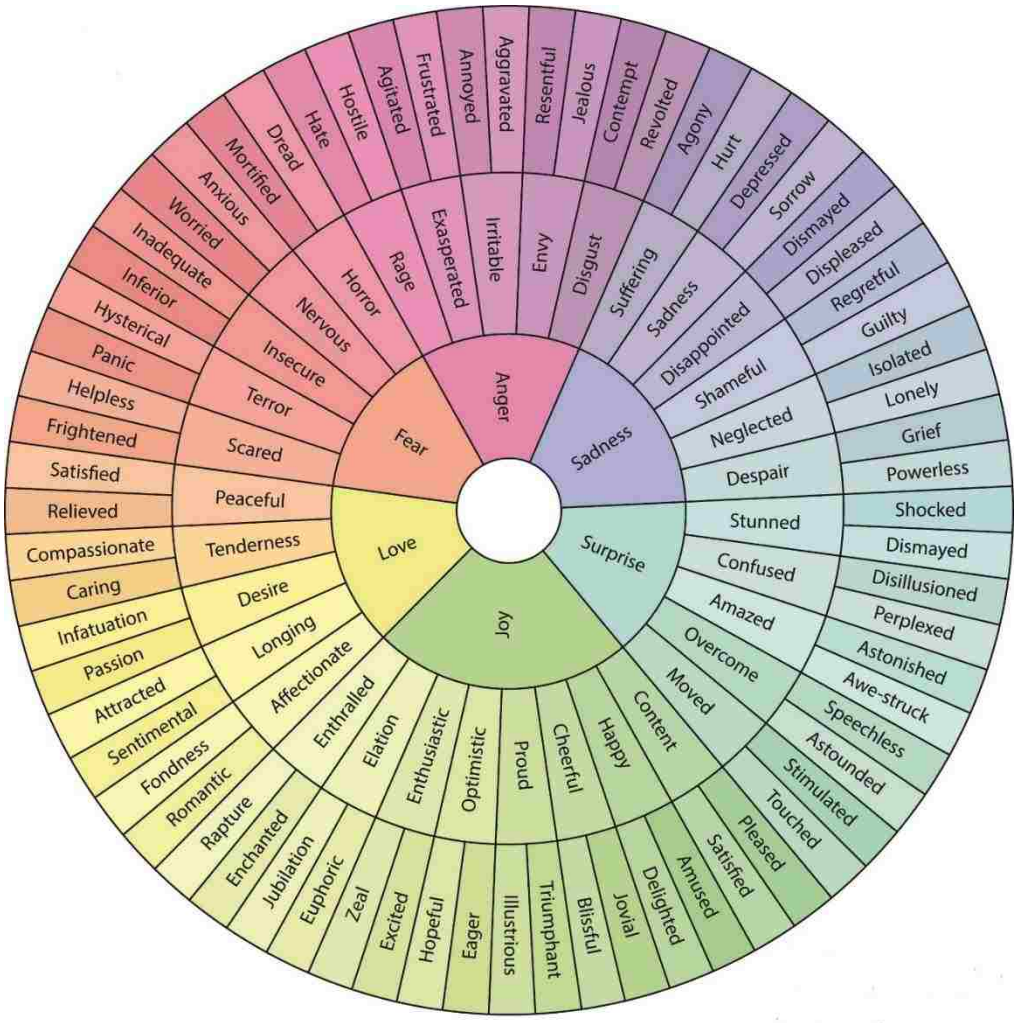


Figure 4.3: Emotion and feeling wheel. Adapted from Funto Institute of Entrepreneurial Leadership, Chadha 2016).

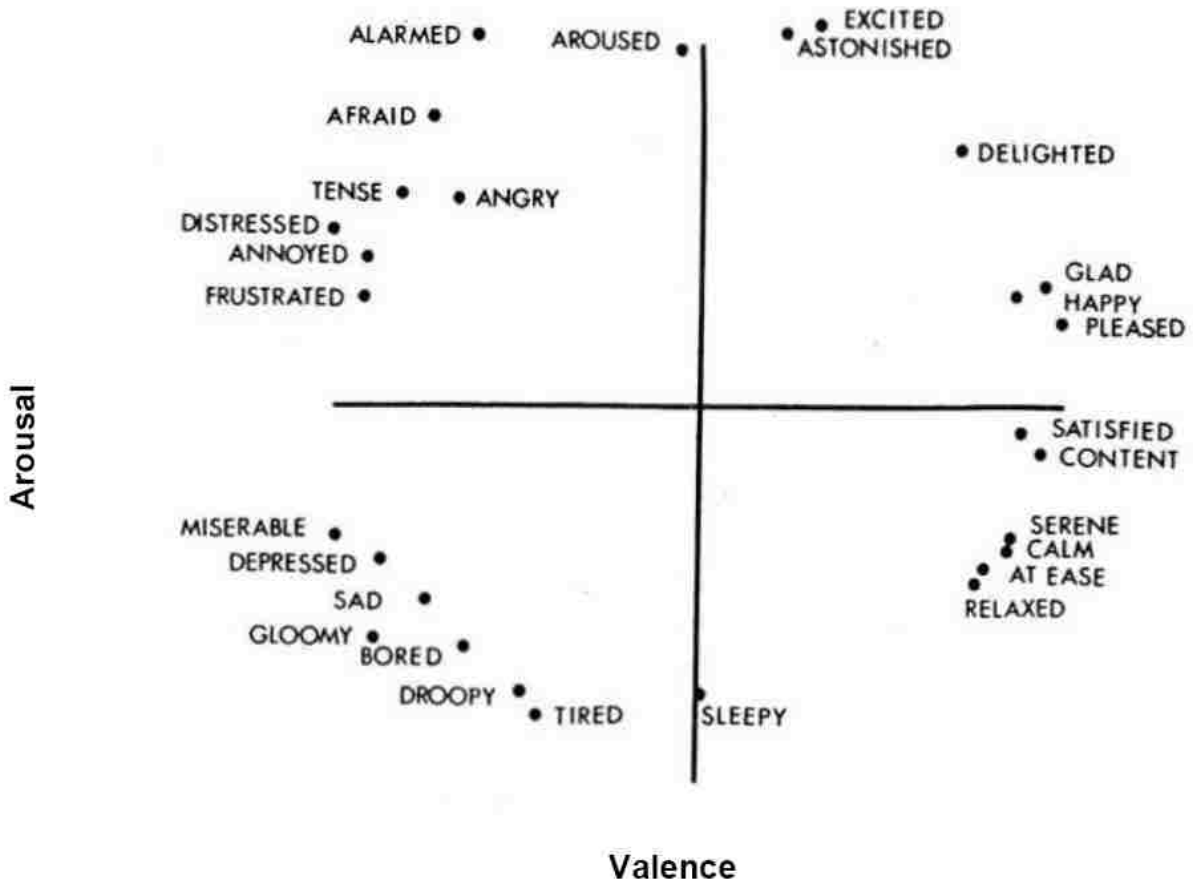


Figure 4.4: Mapping of various emotions and affective words into valence-arousal space. Adapted from Russel's circumplex affect model, Russell (1980).

Therefore, having thresholded the participants' ratings for each of the valence, arousal, dominance and liking scales along two classes, which are low/high by placing the threshold in the middle as also done in (Koelstra et al. 2012) such that all ratings above 4.5 score are respectively mapped to high arousal, high valence, high/in control dominance and like while ratings below 4.5 scores are mapped to the respective lower dimension/dislike emotion representations, the Distress Phase emotion representation is modelled as follows. The happy phase in the emotion model as shown in Table 4.3, will consist of physiological signals of all positive/high valence emotion scores as rated by the participants in the DEAP dataset. This is steered by the human emotions and the hedonism theory that people are motivated to seek pleasure and avoid pain (Franken 1994) as happy phase involves maximizing the positive effects of the various human

sensory system. Positive valence emotions include elation, joy, glad, delighted, satisfaction, pride, happiness, relief, amusement, love, satisfaction, pleased, satisfied, calm, serene, relaxed, at ease, and jubilation, as these are pleasurable and desired by people (Franken 1994). Individuals with these positive valence emotions are characterized with humour, self-confidence, optimism, cheerful, gratitude, sense of accomplishment, creativity and a sense of personal control traits. Therefore, in this study, it is concluded that whatever the arousal, dominance and liking scores are, the most important emotion representation index for the happy phase is the positive/high valence as shown in Table 4.2, which indicate pleasure, generally desired with no harm/threat and thus would not require any emergency assistance to be prompted/provided.

Table 4. 2: Emotion representation scores to determine Happy phase, Distress phase and Casualty phase.

Phase	Emotion Types	Emotion representation scores			
		Valence	Arousal	Dominance	Liking
Happy	Amusement, Joy, Love, Satisfaction, Happiness etc.	+ (positive/pleasant valence emotions)	+/- (low/high arousal)	+/- (low/high dominance)	+/- (like/dislike)
Distress	Fear, Anger, Distress, Annoyed, Alarmed etc.	- (negative/unpleasant valence emotions)	+ (high arousal)	+ (high dominance)	+/- (like/dislike
Casualty	Sadness, Shame, Grief, Sorrow, Depressed etc.	- (negative/unpleasant valence emotions)	- (low arousal)	- (low dominance)	+/- (like/dislike)

The other two phases to consider in the emotion model are the distress and the casualty phases. The intention is to build an affective model that can trigger emergency assistance once the danger is imminent, so as to avoid a casualty phase where a loss would be suffered. These two phases will both therefore have a low/negative valence score since humans are motivated to avoid pain (Franken 1994) thus requiring prompting an emergency services whenever there is a threat to pleasure.

Negative valence indicates unpleasantness, displeasure and aversiveness feelings of an emotion. They are characterised by suffering, insecure, pain - which may include physical and mental pain; panic, nervousness and helplessness (Chadha 2016). Negative valence discrete emotions include sadness, anger, fear, disgust, distressed, shameful, guilty, depressed, agony, hate, frustrated, distressed, contempt etc.

Since both the distress phase and casualty phase have low/negative valence emotions, the arousal and dominance ratings are utilised to separate these two phases. This is because the arousal rating quantifies heightened physiological activity and provide emotional responses in terms of “fight”, “freeze” or “flight” to daily life experiences being confronted.

High arousal rating indicates more will and determination to fight or flight, it connotes that an individual can still muster strength to face the impending threat and thus in control. The distress phase is therefore mapped with a high arousal rating as similarly done by Russell (1980) because it is exactly at this point that an emergency service is intended to be prompted not when the emotional arousal is very low, indicating an overpowered or helpless situation characterized with freeze, less willingness and determination to confront the threat thus possibly leading to a casualty phase.

Out of the negative valence emotions earlier listed, relying on Russell (1980) as well as the emotion wheel adopted in the DEAP dataset, the discrete emotions and affective words that could be mapped to the distress phase include fear, angry, distressed, tensed, hope, contempt, disgust, envy, aroused, alarmed and annoyed. The distress phase is characterised by hope, acknowledging that there is impending trouble but with a conviction that things can turn out fine and well, though fearing the worst but

always expecting the best. It is believed that with prompt emergency assistance and all depths of individual efforts, a positive outcome is expected to be achieved.

The casualty phase is what the distress phase emotion model wants to avert as it is not desired because, a tragedy has already occurred leading to loss of probably life, limb, valuables, successful kidnaping, rape, fire burns among others, thus necessitating the mapping of the casualty phase to low arousal and negative valence emotions/affective words. The casualty phase is characterized with not in control/overpowered, helplessness, irrevocable loss, dejection, defeat and sorrow. The discrete emotions and affective words mapped in this phase include sadness, regret, shame, hurt, disappointed, displeased, suffering, grief, despair, depress, miserable and droopy (Russell 1980; Chadha 2016).

Following the proposed Distress Phase emotion model earlier highlighted, the respective DEAP data set physiological signals that has been appropriately mapped to the happy phase, distress phase and casualty phase labels would then be passed as a tripartite labeling scheme under the Distress Phase model emotion representation for human emotion classification and compared with the results to be obtained from the binary classification problems of valence, arousal, dominance and liking respectively in order to determine the best obtainable result.

4.4 Feature Engineering

The feature-engineering concept is the process of using domain knowledge of a modality data set to generate features from it that are capable of enhancing the performance of machine learning classifiers. The classification results of a machine learning algorithm therefore rests among others, on the quality of features engineered from the raw data as well as how the features are presented to the model. Thus, if the feature engineering is done excellently the predictive power of the pattern classifier will be positively impacted.

Part of the steps involved in feature engineering, including transforming of raw data into discriminatory features for a better representation of the fundamental problem and

presentation of the predictive models for an enhanced model's accuracy are captured as feature extraction and feature selection, which are hereby discussed below.

4.4.1 Feature Extraction

Raw physiological data are often of higher dimensions and are accompanied by various errors and other artefacts. This makes directly analysing and processing the raw data for decision making not ideal as it may lead to unacceptable results and wrong interpretations. Therefore, anomalies have to be identified and removed. Feature extraction, thus involves converting the raw physiological data into a series of feature vectors that bears characteristic and useful information inherent in the raw data. These feature vectors are often of smaller dimensions than the raw physiological data, because associated errors and artefacts have been significantly removed.

Feature extraction in human emotion recognition and other pattern recognition studies such as human activity recognition, using physiological signals and accelerometers data respectively occur in time, frequency domains, sub-band spectra, multiscale entropy and geometric analysis (Picard, Vyzas and Healey 2001; Jonghwa and Andre 2008). These features are statistical in nature and include the mean absolute values of the first differences of the raw signals, RMS of the mean squared differences, mean of the absolute values of the first differences of the normalized signals, mean and standard deviation of the raw signals, mean square error, range and spectral power in the band (Picard, Vyzas and Healey 2001; Jonghwa and Andre 2008; Soleymani et al. 2012). The Fast Fourier Transform (FFT) technique is often utilised in calculating frequency domain features. This method reduces the processed signals into features that are discriminating for the required emotions or activities of interests.

However, the strength of statistical features is that they can easily be calculated in an online manner (Vyzas, and Picard 1999) but they have the disadvantage of not exploiting the knowledge about the physical sources of the physiological signals (Picard, Vyzas and Healey 2001). To circumvent the intrinsic drawbacks of statistical features, different researchers have used varying feature extraction techniques such as the

Empirical Mode Decomposition (EMD), Hilbert Huang Transform (HHT), Wavelet transforms, Fourier transforms and Robust Singular Spectrum Transform (RSST) (Cong and Chetouani 2009). In the literature, some feature extraction techniques that have been applied for human emotion recognition based on physiological signals are Fourier transform, Hilbert-Huang transform (fission and fusion), six scale Daubechies wavelet transform, thresholding peak detection, time domain statistical features, spectral power density, entropy, Fisher projection, wavelet transform, sequential floating forward search and Maximum a Posteriori (MAP) (Jonghwa and Andre 2008; Cong and Chetouani 2009; Maaoui and Pruski 2010).

Mathematically, the feature extraction problem can be defined as follows. If there exists a feature space $x_i \in R^N$, derive a mapping function that transforms R^N such that

$$Y = f(x): R^N \rightarrow R^M \qquad \textbf{4.9}$$

where M<N giving a transformed feature vector: $y_i \in R^M$ that retains a huge amount of the information or structure in R^N. Feature extraction is therefore limited to linear transformation of the type $y = W_x$ as shown in Equation 4.10.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{linear feature extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \\ y_M \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & & w_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \textbf{4.10}$$

4.4.2 Feature Selection

Contingent on the modality of the physiological signals, only a subset of the features extracted is required for human emotion recognition (Picard, Vyzas and Healey 2001; Soleymani et al. 2012), while the quality of the features employed significantly contributes to the recognition accuracy attained (Jonghwa and Andre 2008).

Thus, feature selection entails obtaining only the discriminatory features of the extracted features. This is because not all the extracted features significantly represent the inherent characteristics in the acquired signals.

This is achieved by selecting a subset of the extracted features without necessarily transforming them as shown in Equation 4.11. After the feature extraction process, applying dimensionality reduction algorithms, including the Principal Component Analysis (PCA) and the subsequent selection of the Eigenvectors as applied in this study constitute the feature selection procedure reported in this dissertation. The PCA is preferred among the other dimensionality reduction methods, including the Linear Discriminant Analysis (LDA) and the kernel PCA because it has been shown to be capable of yielding better result above the others (Goshvarpour, Abbasi and Goshvarpour 2017). Feature selection is essential because a high dimensional feature space can be computationally expensive, increases the curse of dimensionality as well as leading to poor classification results. Feature selection subsets are obtained through searches for a subset that minimizes some cost function such as test error. Feature selection involves reducing the data set dimension by analysing and understanding the impact of its features of the model.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{feature selection}} \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \\ x_{i_M} \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{feature extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \\ y_M \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) \qquad \textbf{4.11}$$

Mathematically, feature selection can be represented as:

Given a feature set:

$$x = \{xi \mid i = 1, \ldots, n\} \qquad \textbf{4.12}$$

finding a subset

$$x_m = \{x_{i1}, x_{i2}, \ldots, x_{im}\} \quad \text{with } m<n \qquad \textbf{4.13}$$

that could optimize an objective function (feature extraction function) is required on the set of segments containing feature vectors $X_{i.}$

The objective function here could use the k-fold cross-validation technique necessary in selecting the optimal parameters for modelling an activity or emotion classification problem and performance evaluation of the selected model.

In the study reported in this dissertation, the HOG, LBP and Histogram of Images (HIM) were extracted from the channel images obtained from the inverse fisher transformed physiological signals of DEAP dataset. These features are called ‘local’ features as they were extracted from each of the 40 channels of each sample, such that for each channel, the HOG descriptor has 81-feature vector size while the LBP and HIM descriptors have 256-feature vector size each, extracted per channel.

Therefore, to extract the whole features from the 40 channels for each of the trial sample, for the HOG descriptor, the feature vector of 81 dimensions for each channel is to be combined such that each sample of the physiological signal has a vector size of 81x40= 3,240. For the LBP and HIM descriptors as well, the feature vector size of 256 per channel gives a vector size of 256x40=10,240 for each sample after concatenating the 40 channels.

However, it has been shown that utilizing dimensionally reduced feature vectors of a transformed standardized data or features rather than on a non-standardized data gives better results (Raschka 2014) as the dimensionally reduced feature vectors are capable of improving classifier performance as reported in the literature (Adetiba and Olugbara 2015). This fact was therefore leveraged on as the Principal Component Analysis (PCA) was applied to the extracted features of each channel before the concatenation. The PCA is aimed at standardizing data using inherent relationship existing in the data, which are often linear or almost linear thus making the data responsive to the analysis. PCA rotates original data to new coordinates by finding a low-dimensional linear subspace such that when the data is projected, information loss is minimized while making the computation of eigenvalues and eigenvectors of the covariance matrix achievable. The process involved in obtaining the principal components (Ada and Kau 2013) in a feature include;

(a) Compute the mean and standard deviation of the feature data

(b) Centres and scaling of the feature by subtracting the sample mean from each observation in the feature and subsequently dividing by the sample standard deviation.

(c) The coefficients of the principal components are obtained as well as the associated variances by computing the Eigen function of the sample covariance matrix.

(d) The coefficients for the principal components is stored in the covariance matrix with the diagonal elements representing the variance of the respective principal components.

(e) The maximum variance obtained in the feature data contains the dominant information content, which is needed for improved recognition result.

Hence, a simple dimensionality reduction algorithm based on the Principal Component Analysis (PCA) with the associated eigenvectors indicating the most dominant or principal component elements of the feature vectors respectively was adopted.

For each feature vector I_i extracted by a pattern descriptor i.e. HOG, LBP and the HIM from N samples, I is an m x n matrix of each extracted feature vector where i = 1, 2, 3, …, N. The matrices for N samples are averaged and represented as I_{avg}, the covariance matrix C is thus computed (Abdelrahman and Abdelwahab 2018) as shown in Equation 4.14;

$$C = \sum_{I=1}^{N} \left(I_i - I_{avg}\right)' \left(I_i - I_{avg}\right) \qquad \textbf{4.14}$$

The covariance matrix C is evaluated to obtain the eigenvalues and the corresponding eigenvectors; y eigenvectors are selected based on the principal eigenvalues. The size of the matrix D containing the selected y eigenvectors is n x y such that the extracted feature F_i of each sample is obtained by projecting the feature vector I on the matrix D as shown in Equation 4.15;

$$F_i = I_i . D \quad \textbf{4.15}$$

After applying the PCA and obtaining the eigenvectors, the local HOG, LBP and HIM features of each channel for the EEG modality for instance, are reduced from 81, 256 and 256 feature vector sizes respectively to 10, 31 and 31 most dominant features respectively. By this step, for each sample of the physiological signals to be trained by the RBFNN pattern recognizer, as shown in Table 4.3, the feature vector sizes of the HOG, LBP and HIM features are now 320, 992 and 992 respectively after concatenating features for all the 32 channels available in each sample under the EEG modality for instance. By this procedure, the new feature size was realised with about 90% dimensionality reduction.

On the other hand, the peripheral physiological modality has 8 channels and the feature vector sizes of each channel after applying the PCA algorithm have 6, 7 and 7 dominant features respectively, for the HOG, LBP and HIM descriptor thus giving 48, 56 and 56 feature vector sizes respectively for each sample of the physiological signals. Also, the fused modality consisting of the EEG and peripheral physiological data have 4, 36 and 39 PCA dimensionally reduced feature vector sizes, which gives 160, 1440 and 1560 feature vector elements for each sample of the physiological signals for the HOG, LBP and HIM descriptors respectively.

Table 4. 3: Size of the original extracted and PCA dimensionally reduced feature vectors.

Modality	Number of channels	Size of original feature vectors of each sample for each pattern descriptor			Size of PCA dimensionally reduced feature vectors of each sample for each pattern descriptor		
		HOG	LBP	Histogram	HOG	LBP	Histogram
EEG	32	2592	8192	8192	320[A]	992[B]	992[C]
Peripherals	8	648	2048	2048	48[D]	56[E]	56[F]
EEG+Peripherals	40	3240	10240	10240	160[G]	1440[H]	1560[I]

In sum, the dimensionally reduced feature vectors of each channel were subsequently concatenated to form the respective HOG, LBP and HIM features of each sample. This feature selection and dimension reduction technique employed with the PCA

will significantly reduce the computation cost and time as the dimensions of the feature vectors have been greatly trimmed (even to about 95% with the HOG features of the fused modality) while also retaining the dominant and discriminatory properties inherent in the original features.

The nine different dimensionally reduced features extracted from the inverse Fisher transformed DEAP physiological data set namely the HOGPS, LBPPS, HIMPS; HOGPEPS, LBPPEPS, HIMPEPS; and HOGHES, LBPHES and HIMHES as stated in section 1.6 (b) are represented as A,B,C,D,E,F,G,H and I respectively in Table 4.3.

4.5 Feature Descriptors

The HOG and LBP feature extraction techniques of the digital image processing domain as well as the HIM features adopted for experimentations to be carried out and reported in this dissertation are hereby presented. These digital image processing techniques were adopted because of the success and high performances recorded by them in pattern recognition problems.

4.5.1 Histogram of Oriented Gradient

The Histogram of Oriented Gradient (HOG) is an image processing procedure developed for human recognition and object detection, considering that local object appearance and shape of an image can be represented by the distribution of intensity gradients or edge orientations (Dalal and Triggs 2005; Dipankar 2014). It was originally used for data compression as well as detecting edges around an object with a strong performance as an appearance, shape and feature extraction technique.

The implementation of HOG involves dividing an image into evenly sized overlapping cells and compiling the histogram of gradient directions for the pixels within the cells. The gradient vector is calculated by using the changes in pixel intensity values within a given cell from the x and y axis. Thus, the magnitude and direction of the gradient vector can also be computed from the fluctuation in the pixel intensity values. An

increased brightness or magnitude may alter the pixel values, however the relative difference between the pixels stand. This relative difference is utilised to distinguish the edges of objects within an image and it is invariant to the magnitude. The magnitude of a block of cells is computed and the value obtained is applied to all the pixels within the particular block. The aggregation of the histogram of gradient represents the HOG features. To compute the HOG feature from a given image, four essential steps are required, which are masking, orientation binning, local normalization and block normalization. Research work reported in Dalal and Triggs (2005) contains detailed information regarding the computations and characteristics of the HOG features.

A 3x3 block of cells and 9 bins were utilised in the original implementation of the HOG algorithm in generating a feature vector size of 81 elements from a greyscale image. In the research study being conducted and reported in this dissertation, the HOG feature was extracted from the inverse Fisher transformed physiological signals of the DEAP data set and passed to the RBFNN pattern classifier for human emotion recognition.

4.5.2 Local Binary Pattern

The Local Binary Pattern (LBP) descriptor was originally developed by Ojala, Pietikainen and Harwood (1996). It describes the texture of an image and has been widely applied in diverse applications (Ojala, Pietikainen and Harwood 1996; Ojala and Pietikäinen 1999; Ojala et al. 2001; Ojala, Pietikäinen and Mäenpää 2002). It assigns numeric label for the block of pixels of an image through a thresholding process that uses a 3x3 neighbourhood of the centre pixel value while treating the result obtained as a binary number.

Since the neighbourhoods to the centre pixel consist of 8 pixels, the texture descriptor is derived from the histogram of the $2^8 = 256$ different labels. The LBP operator offers impressive performance in unsupervised texture segmentation when used together with a simple local contrast measure. It was later revised and extended to a more generic form to accommodate neighbourhoods of various sizes including a circular neighbourhood.

The LBP value is computed following the steps described in (Ojala, Pietikainen and Harwood 1996). We chose the LBP descriptor because of its proficiency in appropriately describing the texture of an image (Ojala, Pietikainen and Harwood 1996; Garcia-Olalla et al. 2013). In addition, it has a modest theoretical definition, which is the foundation of its status as a computationally efficient image texture descriptor in the digital image processing field of study (Ahonen et al. 2006; Rahim et al. 2013).

4.5.3 Histogram of Images (HIM)

Digital image processing involves procedure of obtaining useful information from images by determining an image pixel property and variation for the purpose of analysis, classification and recognition/identification. Histogram of an image represents the histogram of the intensity values of pixels in the image. It is a graphical representation that covers all the various intensity values in the image.

Thus, after preprocessing and applying inverse Fisher transform on the DEAP dataset, the data obtained for each sample is converted and mapped into a greyscale image representation with pixel intensity values ranging from 0-255. As an image processing algorithm, the histogram features representing the pixel intensity values in the various greyscale images are computed using an automatic binning algorithm that yields bins with a uniform breadth which are selected to cover the range of pixel intensity elements thus revealing the underlying unique shape and patterns of the distribution.

Histogram as features has strong capabilities for identification and differentiation of patterns and was therefore employed as features for human emotion recognition along the valence, arousal, dominance, liking and distress phase class labels. Histogram based features have been used for image processing and in several pattern classification studies with promising results (Mohamad, Manaf and Chuprat 2011; Iman et al. 2017; Thamizhvani et al. 2018) thus necessitating the choice of the histogram feature as the third option to the HOG and LBP feature descriptors.

4.6 Feature Classification

Feature classification involves the cataloguing of the extracted features into appropriate classes/states using a pattern matching/classification algorithm. Such pattern classifier is always trained to learn the inherent characteristics in the different extracted features and attempt to match features with similar patterns in the same class. The radial basis function is the pattern classifier that was utilised for the various experiments conducted in this study and reported in this dissertation.

The Radial Basis Function Neural Network (RBFNN) is a feed forward artificial neural network for solving problems of pattern recognition and function approximation (Bors 2001; McCormick 2013). The concepts of RBF are ingrained in earlier pattern recognition techniques such as clustering, spline interpolation, mixture of models and function approximation (Bors 2001; Ugur 2004).

A typical RBF neural network as shown in Figure 4.5, consists of an input layer, one hidden layer consisting of RBF neurons and an output layer of artificial neuron/node for each class/category to be classified (Bors 2001; Ugur 2004; Xianhai 2011; McCormick 2013). Each neuron in the hidden layer implements a radial basis activation function that represents an arbitrary basis for the input vectors, while the network output is a linear combination of radial basis functions of the input and neuron parameters.

The classification task performed by a RBFNN measures the input similarity to samples from the training data set (McCormick 2013). A “prototype” representing one of the samples in the training data set is stored in each RBFNN neuron as classification of a new input involves each neuron computing the Euclidean distance between the new input and its prototype. The new input is classified as belonging to Class 1 prototypes if it resembles Class 1 than Class 2 prototypes. The prototypes are indeed cluster centres computed as the average of all the data points in the cluster.

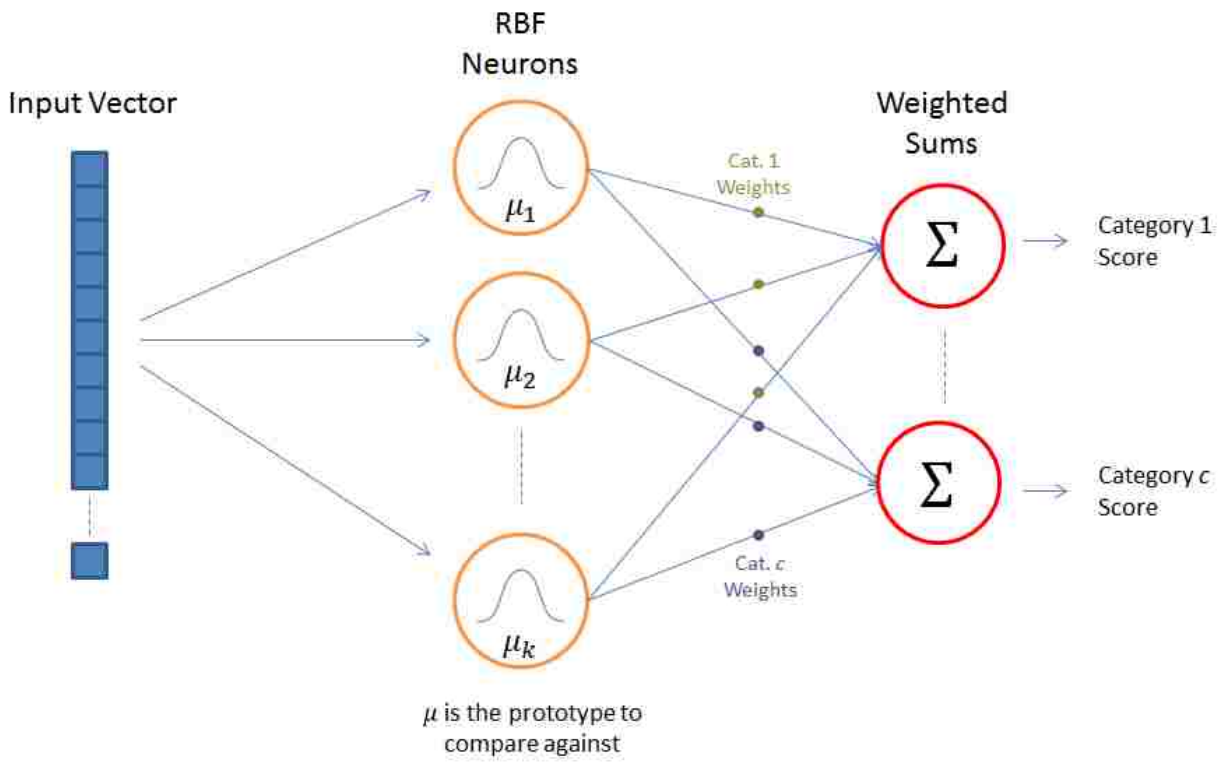


Figure 4.5: The architecture of Radial Basis Function Neural Network. Adapted from McCormick 2013).

Each of the n-dimensional feature vector, which are the HOGPS, LBPPS, HIMPS; HOGPEPS, LBPPEPS, HIMPEPS; and HOGHES, LBPHES and HIMHES extracted from the DEAP data set is respectively fed into the network through the input layer for classification. This n-dimensional feature vector is displayed for each of the RBF neurons in the hidden layer.

A 'prototype' feature vector obtained from one of the feature vectors in the training data set is stored in each RBF neuron which compares an input feature vector with its prototype and a measure of similarity with values of 0 or 1 is the output. If the new input matches the prototype, the output of that particular RBF neuron will be 1 otherwise 0. The RBF neuron's response has a bell curve shape as shown in Figure 4.5 and the value of the neuron's response is termed its activation value. The prototype feature vector is also called the neuron's centre because it is located at the centre of the bell curve.

There are various types of similarity functions with which the similarity between an input vector and its prototype can be computed by an RBF neuron. These include the Gaussian, thin plate spline, quadratic, inverse quadratic, shifted logarithm, linear and cubic functions. The most popular similarity/activation function is the Gaussian radial function which yields a bell curve with a mean value located at the centre of the curve. Because of its popularity and good performance, the Gaussian radial activation function was also employed in this study as a one dimensional input vector in the RBFNN classifier configurations for the various experimentations conducted. More details about the RBFNN can be obtained in the literature (Bors 2001; McCormick 2013).

Furthermore, in the experiments conducted and reported in this dissertation, the RBFNN classifier was utilised to recognize human emotion because of its ability to approximate continuous functions arbitrarily. In addition, it has faster training process because of its local mapping attribute as compared to other neural networks and it is very robust to noise (Leonard and Kramer 1991; Cha and Kassam 1995; Andina and Pham 2007) while it is also capable of yielding at least 10% higher accuracy than can be obtained by the traditional back propagation ANN algorithm (Chapman et al. 1991).

The various classification experiments conducted in this study are hereby presented in Chapter 5 of this dissertation.

CHAPTER FIVE

Experimentations

The personal computer utilised for implementing the experimental models contained an Intel core i7 processor with 3.4GHz CPU and 4GB of RAM running 64 bits Windows 8 operating system. The experimental setups were conducted on three modalities available in the DEAP data set namely the EEG, peripheral physiological and a fusion of EEG and peripheral physiological signals (Chanel et al. 2006; Chanel et al. 2009). Three experimental models were explored in this research study to determine which of the physiological signals extracted feature based on the digital image processing techniques and inverse Fisher transformation with the RBFNN pattern classifier would give the best result of human emotion recognition using the DEAP physiological signals dataset.

The RBFNN classifier was applied to the HOGPEPS, LBPPEPS, HIMPEPS; HOGPS, LBPPS, HIMPS; and HOGHES, LBPHES and HIMHES features extracted from the inverse Fisher transformed DEAP data set while comparing the results achieved to determine the combinations that would give the best performance in recognizing human emotional states across the valence, arousal, dominance, liking and distress phase labels.

In all, 45 different experiments were performed such that five groups each, of experiments based on the five emotion representation labels, which are arousal, valence, dominance, liking and distress phase, were performed using the nine highlighted different features from the three modalities. For instance, for the arousal class, three features, which are HOGPEPS, LBPPEPS, HIMPEPS extracted from the peripheral modality; another three features namely the HOGPS, LBPPS, HIMPS extracted from the EEG modality and three other features namely the HOGHES, LBPHES and HIMHES extracted from the fused (EEG+peripheral) modalities were experimented along the five emotion dimensions.

5.1 Experimental Models

The generic architecture as shown in Figure 3.2 is employed in the design and implementation of the procedures for the 3 experimental models consisting of 45 experiments in this study.

The first experimental model in this study was conducted on the 8 channels peripheral physiological data consisting of the electrooculogram, electromyogram, galvanic skin response, respiration, blood volume pulse and skin temperature data of the DEAP data set from which the HOG, LBP and HIM features namely the HOGPEPS, LBPPEPS and HIMPEPS respectively were extracted. There are 1,280 samples/instances in the data set which was obtained from the 32 participants' 63s (60 second trial and 3 second pre-trial) duration physiological signals of 40 trials per participant. The physiological signals were down sampled to 128Hz. These experiments were conducted to determine how best the combined peripheral physiological data could accurately recognize human emotions along the arousal, valence, dominance, liking and distress phase class labels. Therefore, under the peripheral physiological modality, 15 experiments in all were conducted using the HOGPEPS, LBPPEPS, HIMPEPS features for each of 5 emotion representation labels/classes.

The RBFNN classifier was first applied to the arousal class of the extracted HOGPEPS features, followed by the LBPPEPS features and then the HIMPEPS features in the MATLAB R2018a environment. For each sample of the 1,280-input dataset, as earlier indicated in Table 4.3, the HOGPEPS feature vector contains 48 elements per sample of the training data. This serves as the input data to the RBF network and therefore has 48 neurons in the input layer while the output layer has 2 neurons for the 2 emotion classes indicating the low and high arousal binary classes for classification.

To train an RBFNN network, determining the number of neurons in the hidden layer is very essential as this affects the result that can be obtained. According to McCormick (2013), the prototypes as well as the beta coefficient of the RBF neurons and the matrix of output weights between the RBF neurons and the output node are the parameters that must be carefully selected in the course of determining the number of neurons in the

hidden layer. There exists no strict rule in the literature for selecting the prototypes for the RBF neurons. One approach is to create an RBF neuron for each training sample (Adetiba and Olugbara 2015) such that for the problem at hand, we would have 1280 neurons; while the other is to randomly select *k* prototypes from the training samples (McCormick 2013). These requirements are slack because with adequate number of neurons, an RBFNN can outline any random complex decision boundary and recognition accuracy can always be improved upon by adding more RBF neurons in the hidden layer. However, a trade-off between the efficiency of the RBF network and the accuracy parameters should be considered because more RBF neurons will indicate more computation cost as it is essential that an excellent accuracy is obtained with the possible minimum number of RBF neurons.

A novel method for selecting the prototypes is to perform k-Means clustering of the training sample while selecting the cluster centres as the prototypes (McCormick 2013). The average of all the data points in the cluster is computed as the cluster centres. In addition, while utilizing the k-Means algorithm, the training samples are clustered according to classes such that samples from multiple classes are not included in the same cluster.

The RBFNN Matlab code provided by McCormick (2013) was adopted for the classification experimentations. In order to enhance the network's efficiency and reduce computation costs, instead of using all the available 1280 neurons in the hidden layer, the optimal number of neurons in the hidden layer of the RBFNN was determined by varying the number of clusters between 50-250 per emotion class. For instance, for the arousal class with 2 classes (high/low), and with 50 clusters per class, this translates to 100 neurons in the hidden layer for the 1280 training samples. A 50-250 number of clusters were chosen which indicate 100-500 neurons as similar to Adetiba and Olugbara (2015) where the same 100-500 neurons were chosen for each hidden layer in the MLP-ANN configuration consisting of 534 training samples, 2 hidden layers and 14 classes while the authors also utilised all the available 534 neurons in the hidden layer for the RBF network configuration in their study. The RBF network being hereby configured is therefore much simpler and more efficient than the MLP-ANN and RBF configurations reported in Adetiba

and Olugbara (2015) as a maximum fewer hidden neurons (39.06%) were utilised out of the available 1,280 neurons. Thus, the results obtained with the varied number of clusters for the HOGPEPS feature were chronicled. This experiment was further extended to using the LBPPEPS and HIMPEPS features of the peripheral physiological modality and classification was also done along the arousal classes (high/low).

The input feature vectors for the LBPPEPS as well as the HIMPEPS features have 56 elements each, representing the input neurons in the RBF network. The number of neurons in the hidden layer was experimentally determined as earlier done for the HOGPEPS features while the number of output neurons remains 2 representing high and low arousal classes. The differences between these three sets of experiments in this first model lie in the number of input neurons, the features employed and the optimal number of hidden neurons with which the best recognition results were attained.

The first experimental model was concluded by separately utilizing the three features HOGPEPS, LBPPEPS and HIMPEPS extracted from the peripheral physiological modality for classification of human emotion along the outstanding 4 emotion representation classes namely valence, dominance, liking and distress phase. The respective number of input neurons and the features employed are the same with those used for the arousal class, but the number of optimal neurons in the hidden layer vary as well as the number of neurons in the output layer which is 3 for the distress phase class representing the three classes of Happy, Distress and Casualty phases vary. In all, 15 experiments were conducted in the first experimental model mapped to the peripheral physiological modality data.

The second experimental model utilises the EEG modality data. The inverse Fisher transformed 32 EEG channels of the DEAP data set was used and the HOG, LBP and HIM feature descriptors was applied as explained in section 4.3 to extract corresponding features which is named Histogram of Oriented Gradient Physiological Signal (HOGPS), Local Binary Pattern Physiological Signal (LBPPS) and the Histogram of Images Physiological Signal (HIMPS) features respectively.

The first experiment conducted in this second experimental model is also on the arousal class just as was started with, in the first experimental model. The HOGPS features were utilised and the RBFNN classifier was applied in order to classify the extracted features into the High/Low arousal classes. The HOGPS features has 320 elements for each of the 1280 training samples that was fed into the RBF network. The number of neurons in the input layer is thus 320 while the output layer has 2 neurons, each one representing the High/Low arousal emotion class.

As also done with the experiments in the first experimental model, the optimal number of neurons in the hidden layer was experimentally determined as it was intended to reduce the complexity and computational costs of the RBF network by not utilizing all the available 1280 neurons of the total input samples. The number of clusters per class which eventually determines the number of neurons in the hidden layer was thus varied between 50-250 until the best recognition result was attained.

This experiment is different from the sets of experiments in the first experimental model in terms of the number of neurons in the input layer, the features employed, modality and possibly the optimal number of neurons in the hidden layer with which the best recognition result is attained. This experiment was extended to separately using the LBP features (LBPPS) and also the HIM features (HIMPS) of the EEG modality data. The total number of training samples is 1280 just as obtained in all the other experiments. The number of neurons in the input layer is 992 each for both the LBP features and the HIM features because the feature vectors of these descriptors have 992 elements in each of its 1280 training samples. This huge number of input neurons is as a result of the higher number of channels (32) and dominant components in the EEG modality data above that of the peripheral physiological data. However, the output layer has 2 neurons consisting of the High/Low arousal class as also obtained in the earlier experiments. Thus, three sets of experiments yielding three results were performed with the features extracted from EEG modality with the arousal emotion class (High/Low).

In concluding the second experimental model, the outstanding 4 emotion representation classes of valence, dominance, liking and distress phase were each and separately used with each of the HOGPS, LBPPS and HIMPS features of the EEG

modality data. The same parameters as utilised in the arousal class experiments in the second experimental model are adopted. In sum, 15 experiments were performed in the second experimental model mapped to the EEG modality data.

The third experimental model which is the last sets of experiment to be performed, utilised the fused modality (EEG+peripheral physiological) data of the DEAP data set from which the HOG, LBP and HIM features named HOGHES, LBPHES and HIMHES were extracted as explained in section 4.3, for the recognition task of human emotion. The fused modality has 40 channels per data sample and has a total number of 1280 samples.

The RBFNN was first applied to the extracted HOGHES features for classifying the arousal class. There are 160 elements in each feature vector indicating that the input layer of the RBF network has 160 neurons and the output layer has 2 neurons. The number of clusters per class which determines the number of optimal neurons in the hidden layer was varied between 50-250 until the best recognition result was attained.

This experiment as also done in the first and the second experimental model, was extended to the LBP and HIM descriptors by utilizing the LBPHES and HIMHES extracted features respectively for classifying human emotions along the arousal class. The feature vectors of the LBPHES and HIMHES features have 1440 and 1560 elements respectively. The RBF network, therefore, has 1440 input neurons for the LBPHES features and 1560 input neurons for the HIMHES features. The output neurons are 2 for the arousal label of each of the feature descriptors and the number of optimal neurons in the hidden layer is as experimentally determined as in the first and second experimental models.

The valence, dominance, liking and distress phase emotional representation classes are subsequently used for human emotion recognition under the fused modality data as the HOGHES, LBPHES and HIMHES features were each and separately used for classification along each of the four stated emotional classes. Moreover, in this third experimental model, a total of 15 experiments were performed with 3 experiments per each of the 5 classes using the HOGHES, LBPHES and HIMHES features.

The general differences in the three experimental models lie in the modality data used, hence the features extracted, the number of input neurons fed to the network and

the optimal number of neurons in the hidden layers. The number of output neurons is however the same for all the experimental models as 2 neurons were utilised for the arousal, valence, dominance and liking classes because each of them consist of 2 states (High/Low) while 3 neurons are used in the output layer for the distress phase representation consisting of the happy, distress and casualty phases.

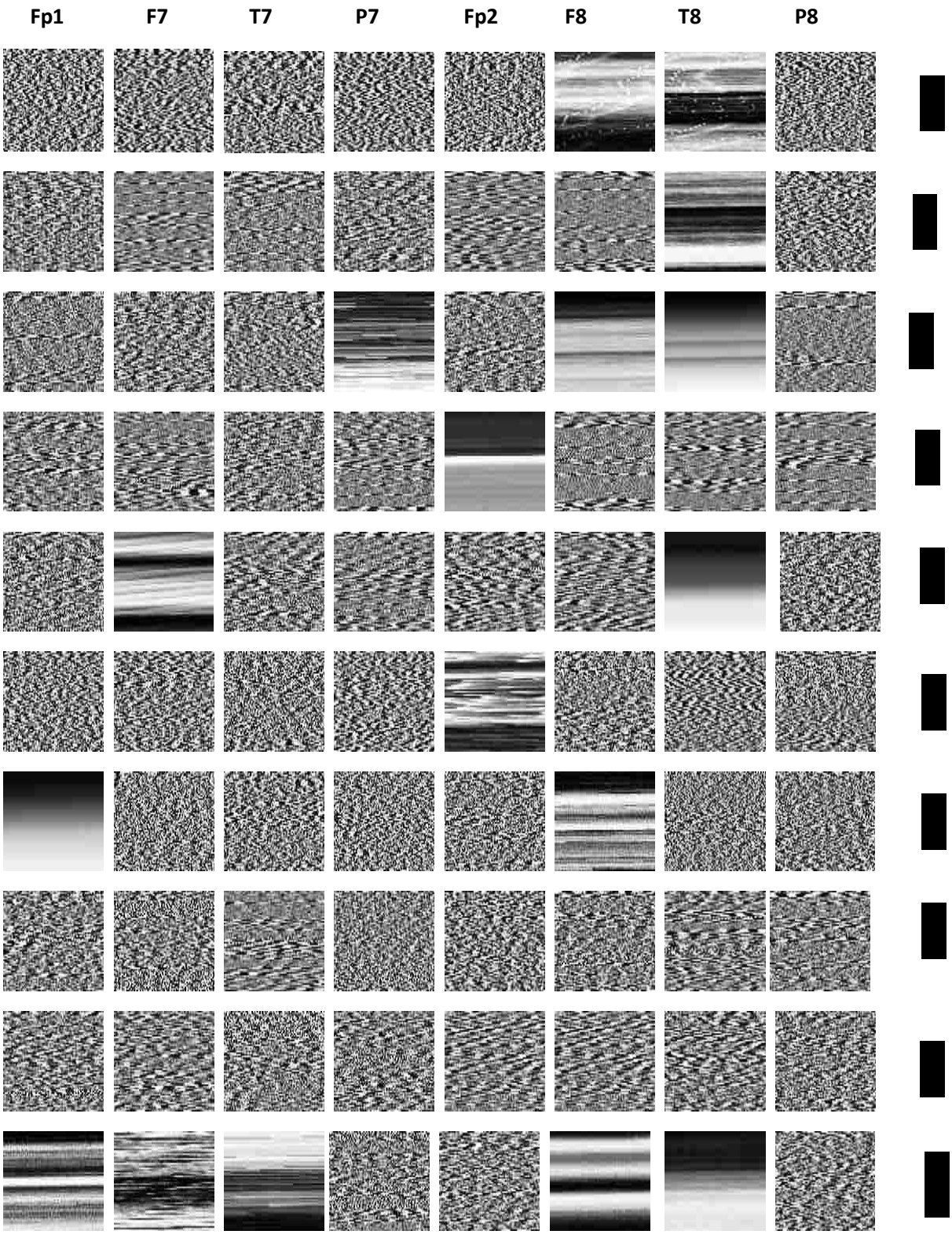
The results indicate the trend of performance of the various modalities, features and representation labels including the combination that posted the highest recognition performance in the 45 experiments conducted are presented and discussed in Chapter 6 of this dissertation.

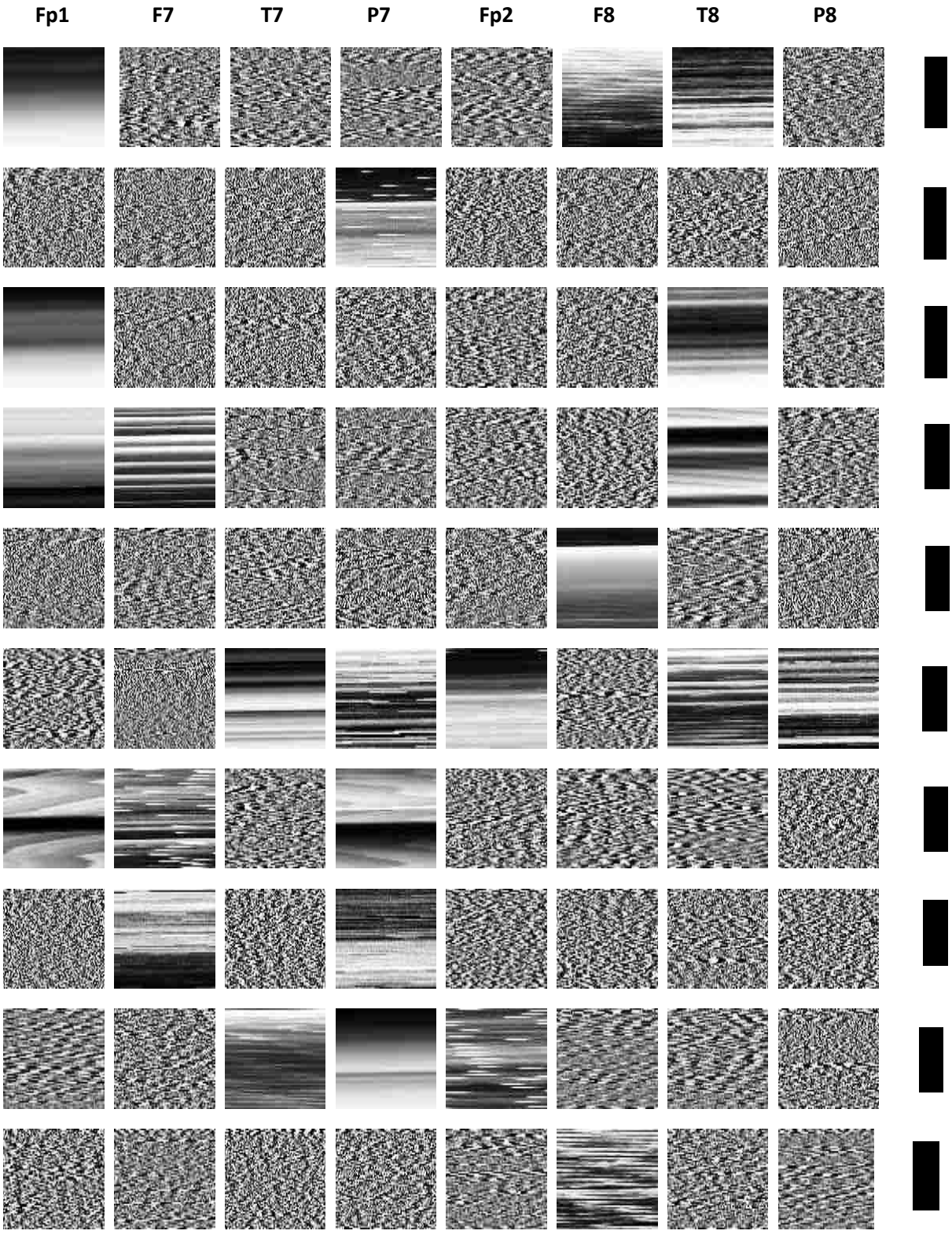
CHAPTER SIX

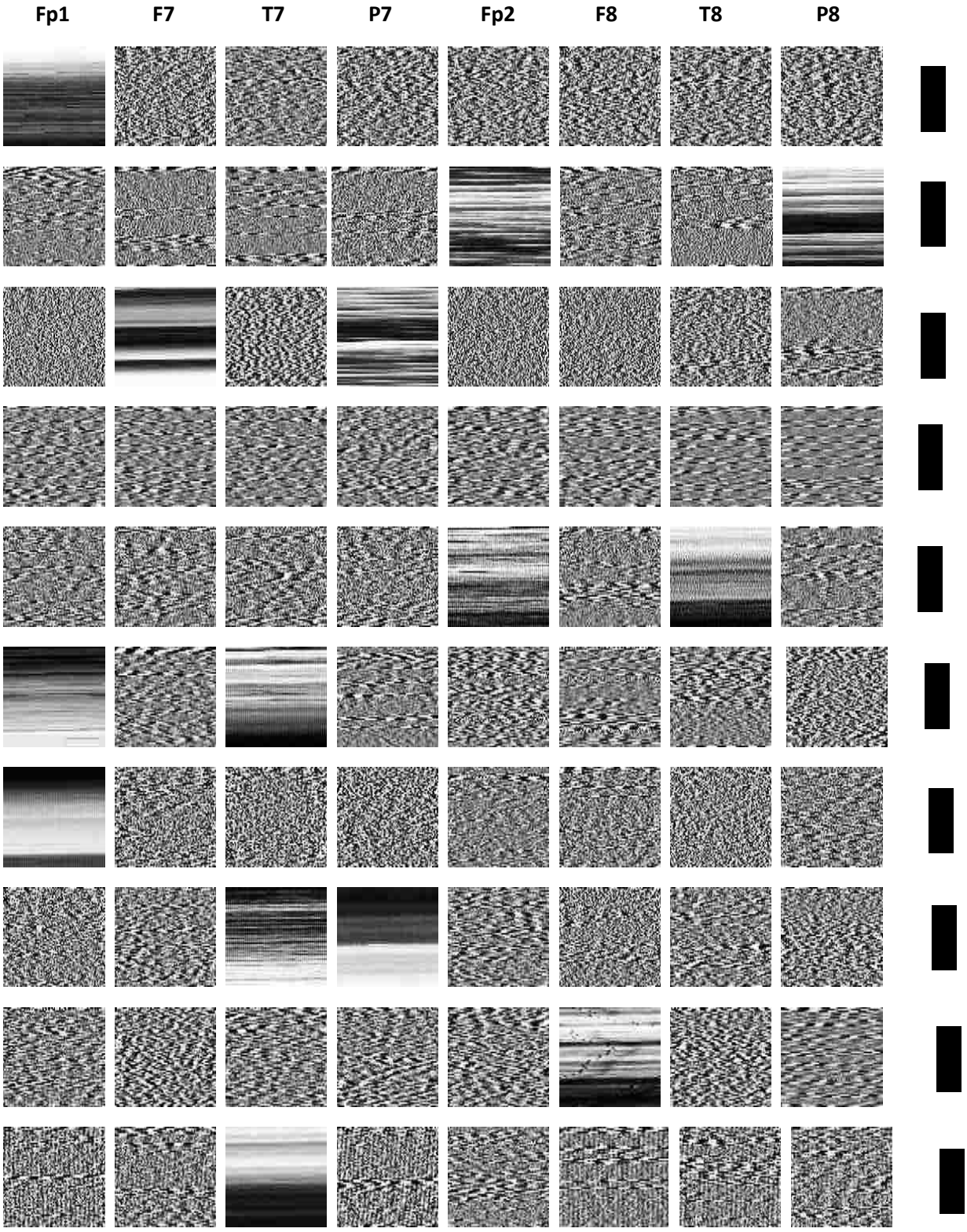
Experimental Results and Discussion

The results of the various emotion recognition experiments discussed in Chapter 5 are presented in this chapter, with some discussions. The results achieved with the various HOG, LBP and HIM features are presented in the order of peripheral physiological, EEG and fused (EEG+peripheral physiological) modalities. The various results were compared with one another within a specific modality and across modalities for all the four emotion dimensions of arousal, valence, dominance and liking as well as the development of the distress phase emotion model. The best results within and across the modalities for these dimensions and features were noted and recommended as an applicable combination for human emotion recognition in an affective system.

The qualitative comparative results of the inverse Fisher transformation of the emotion physiological data are first presented in order to ascertain the similarities across subjects' emotional responses despite the variation in individual's emotional experiences (Siemer, Mauss and Gross 2007). Out of the 40 experimental trials of each subject, 8 trials were randomly selected such that the trials selected for each of the 32 subjects (S1-S32) are unique to enable an appropriate trial mix towards enhancing the generalisation and reliability of inferences drawn and results obtained. From the 32 EEG channels, the Fp1, F7, T7, P7, Fp2, F8, T8 and P8 channels were identified and selected for the EEG modality are reported in literature to be directly related to human emotions (Davidson, Jackson and Kalin 2000; Niemic 2002; Petrantonakis and Hadjileontiadis 2010; Noppadon, Setha and Pasin 2013a; Noppadon, Setha and Pasin 2013b; Menezes et al. 2017; Nakisa et al. 2018). The greyscale image plots of the inverse Fisher transformed physiological data on each subject for 8 random trials and identified channels are shown in Figure 6.1 showing the similarities in the patterns of the images across the subjects. Figure 6.1 shows that there seems to be a similar pattern in the majority of the images across the subjects despite the individual variability that do exist in emotional experiences which necessitates the little variations noticed in the few image patterns of one or two channels per subject.


Fp1
F7
T7
P7
Fp2
F8
T8
P8


Fp1
F7
T7
P7
Fp2
F8
T8
P8


Fp1
F7
T7
P7
Fp2
F8
T8
P8

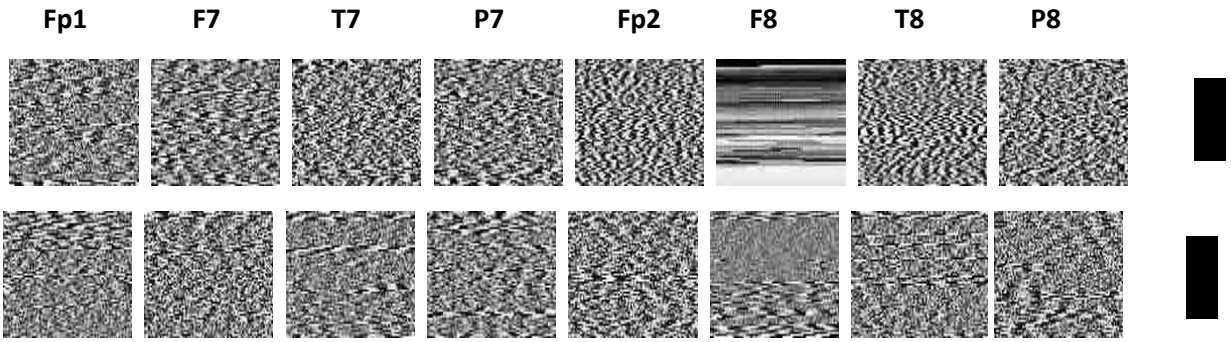
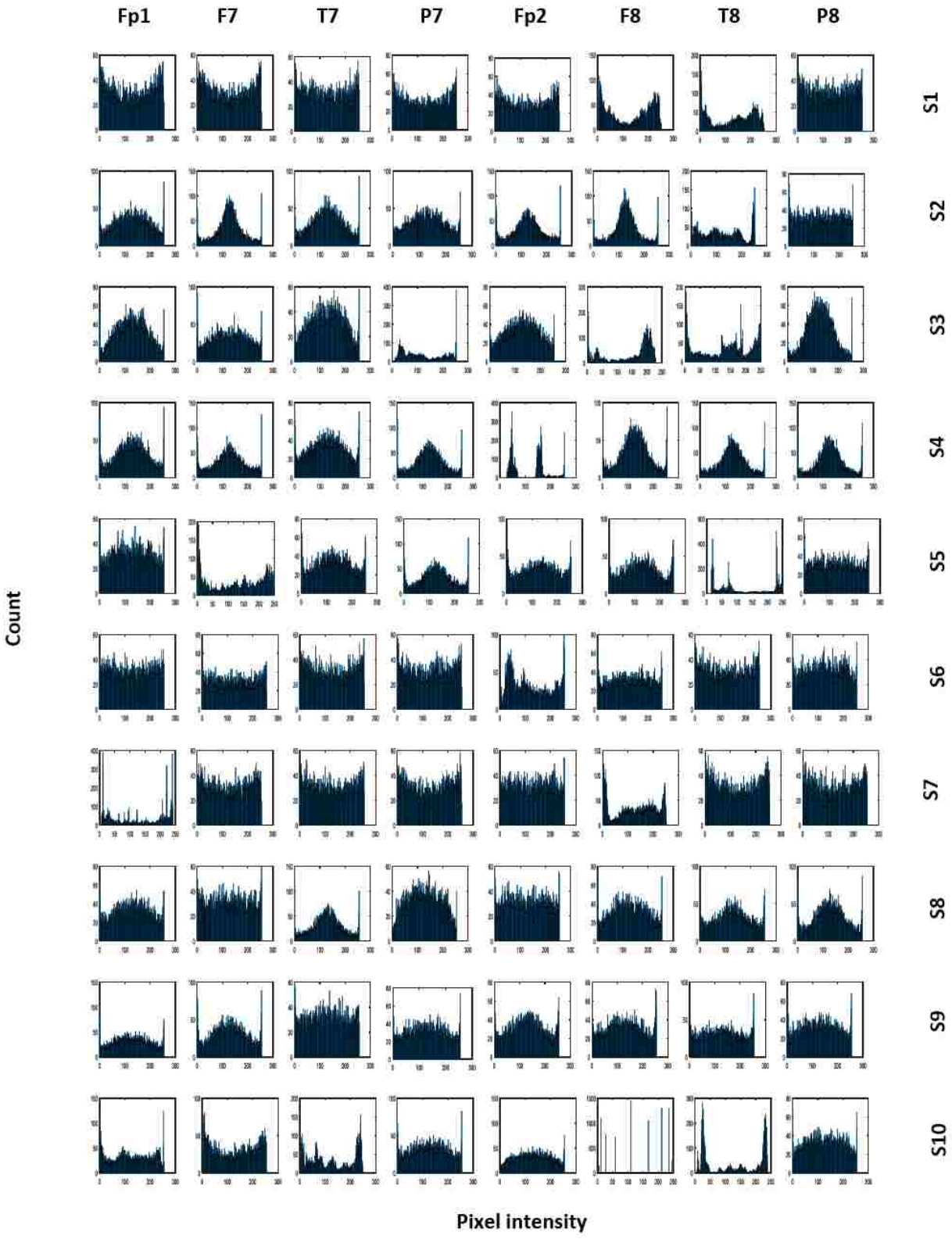
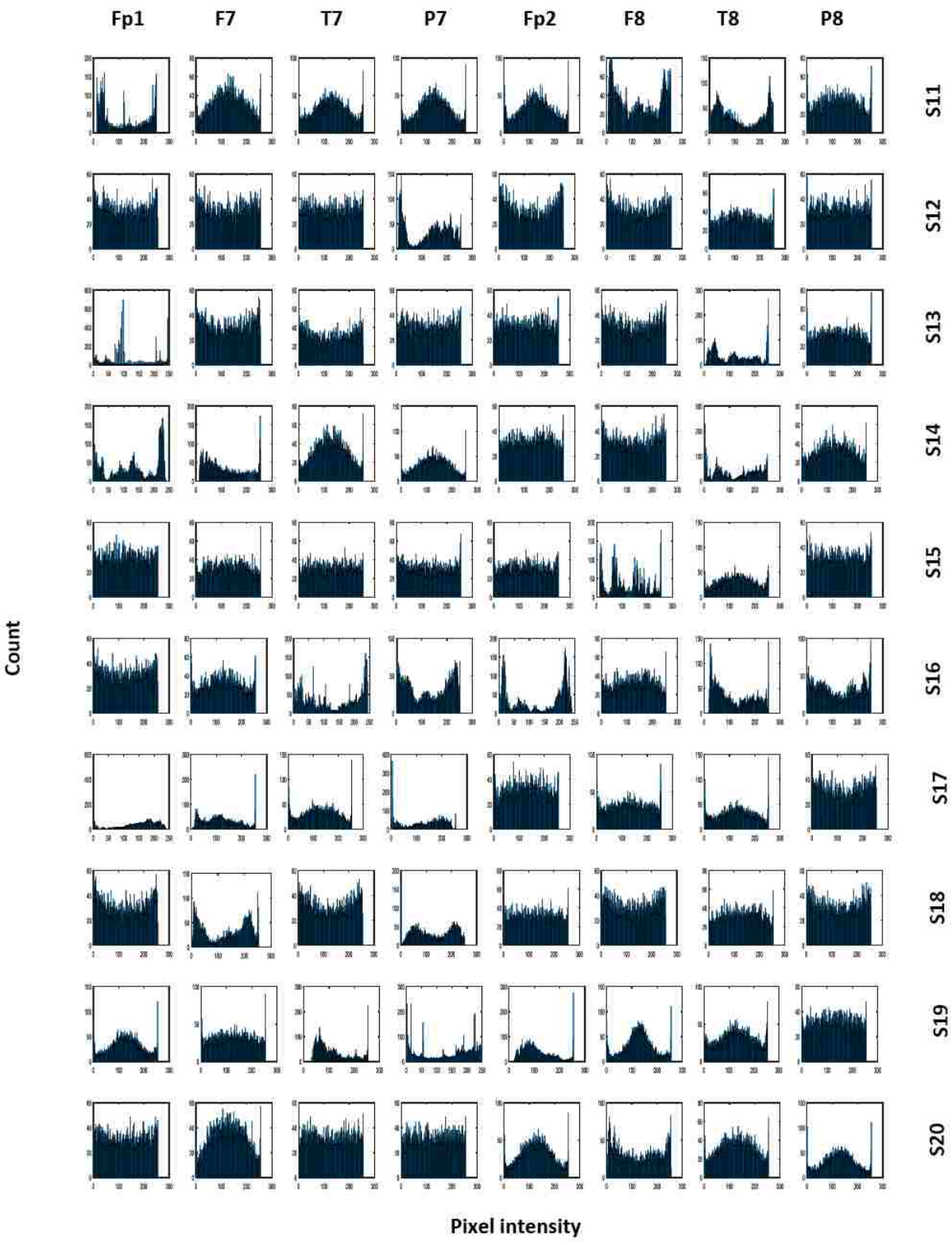


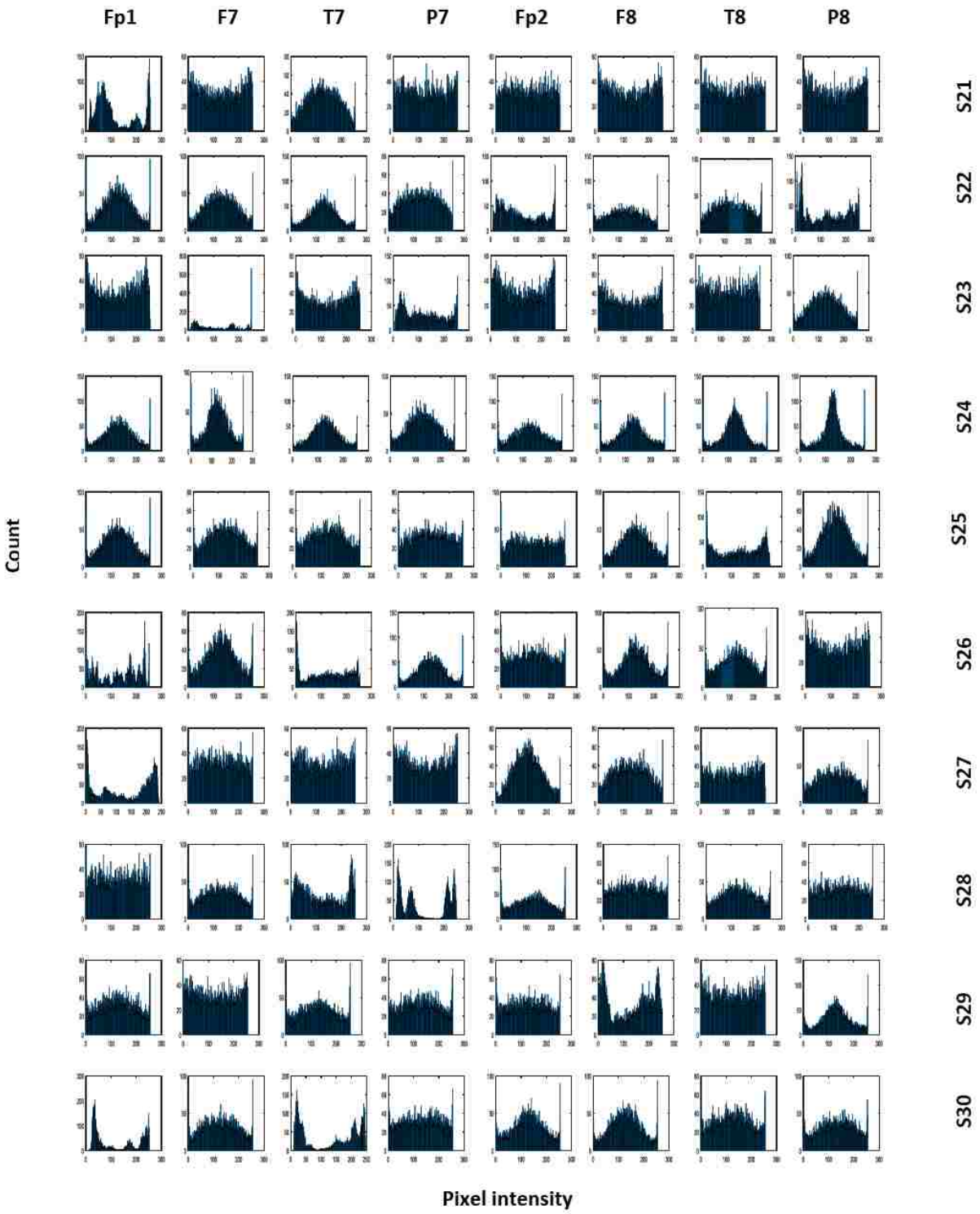
Figure 6.1: Images of inverse Fisher transformed EEG physiological data.

The texture similarities in the images across the subjects is observed in the dense and coarse patterns of the images with observed varied brightness/darkness as a result of the different pixel intensity values and distributions in each image. Thus the strength of the physiological data transformation method that was applied revealed the inherent similarity between the subjects' emotional experiences and responses towards ensuring a subject-independent based inferences and results.

In addition to the greyscale image plot of the transformed physiological data from which the respective features are extracted, the histograms representing accurate distributions of the pixel intensity values for each of the greyscale images in Figure 6.1 are presented in Figure 6.2. The range of pixel intensity values was binned to 256 that is 0-255, which represent the greyscale image pixel intensity distribution. The importance of the histogram data is that it reveals the density of the underlying distribution of the transformed physiological data and can be employed for probability density function estimation of the underlying variable especially by a pattern classifier. Histograms have been utilised in digital image processing for image analysis, brightness, equalisation, stretching and thresholding. There is no ideal image' histogram shape, but the notable patterns include the unimodal, multimodal, bimodal, skewed right, skewed left and symmetric. As observed, the multimodal pattern dominates the histogram plots in Figure 6.2. The multimodal pattern distribution has multiple peaks and could indicate that the physiological emotional responses of subjects have several patterns of responses and


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
Count
Pixel intensity


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
Count
Pixel intensity


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S21
S22
S23
S24
S25
S26
S27
S28
S29
S30
Count
Pixel intensity

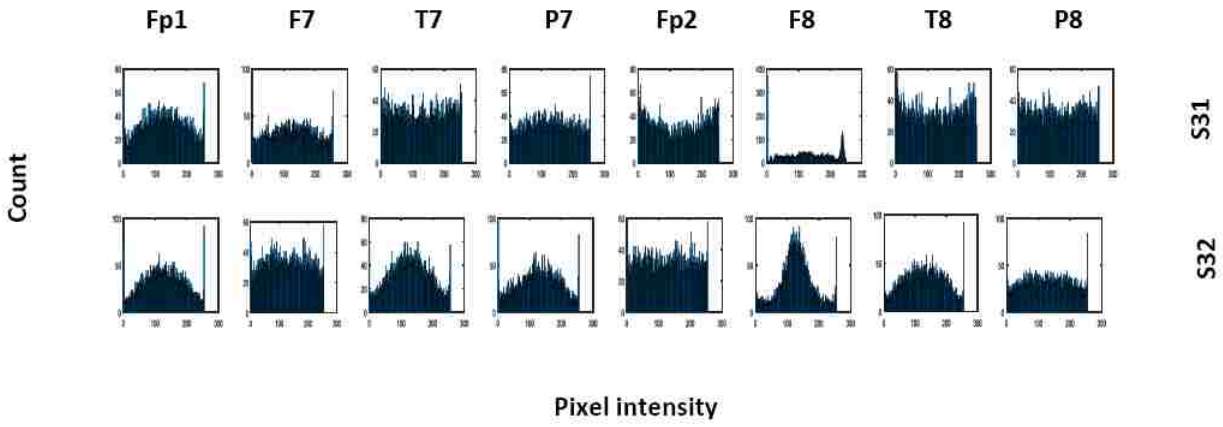


Figure 6.2: Histogram plot of the inverse Fisher transformed EEG physiological data.

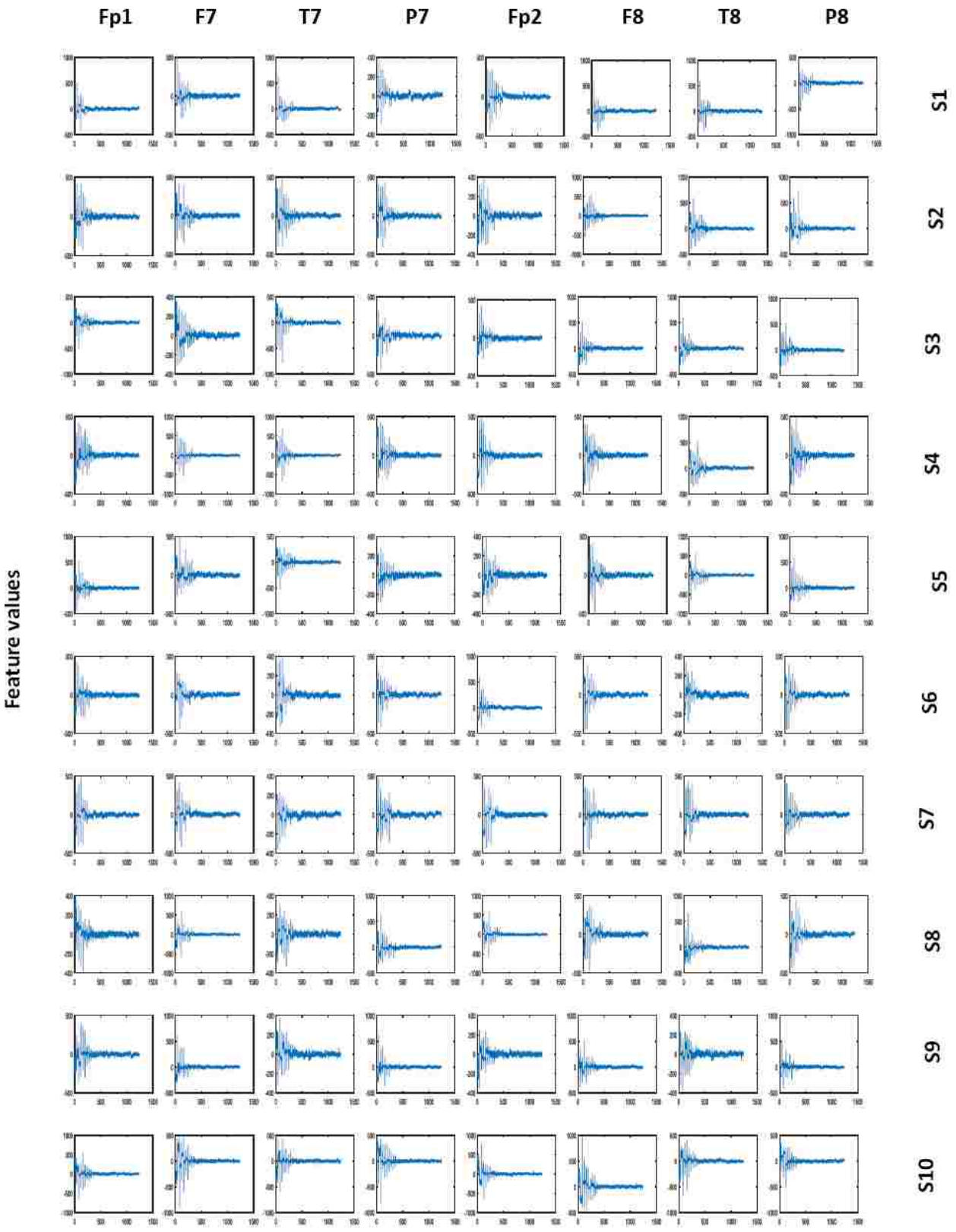
preferences that support the literature position that the emotional experiences and responses of different subjects may not be the same (Siemer, Mauss and Gross 2007).

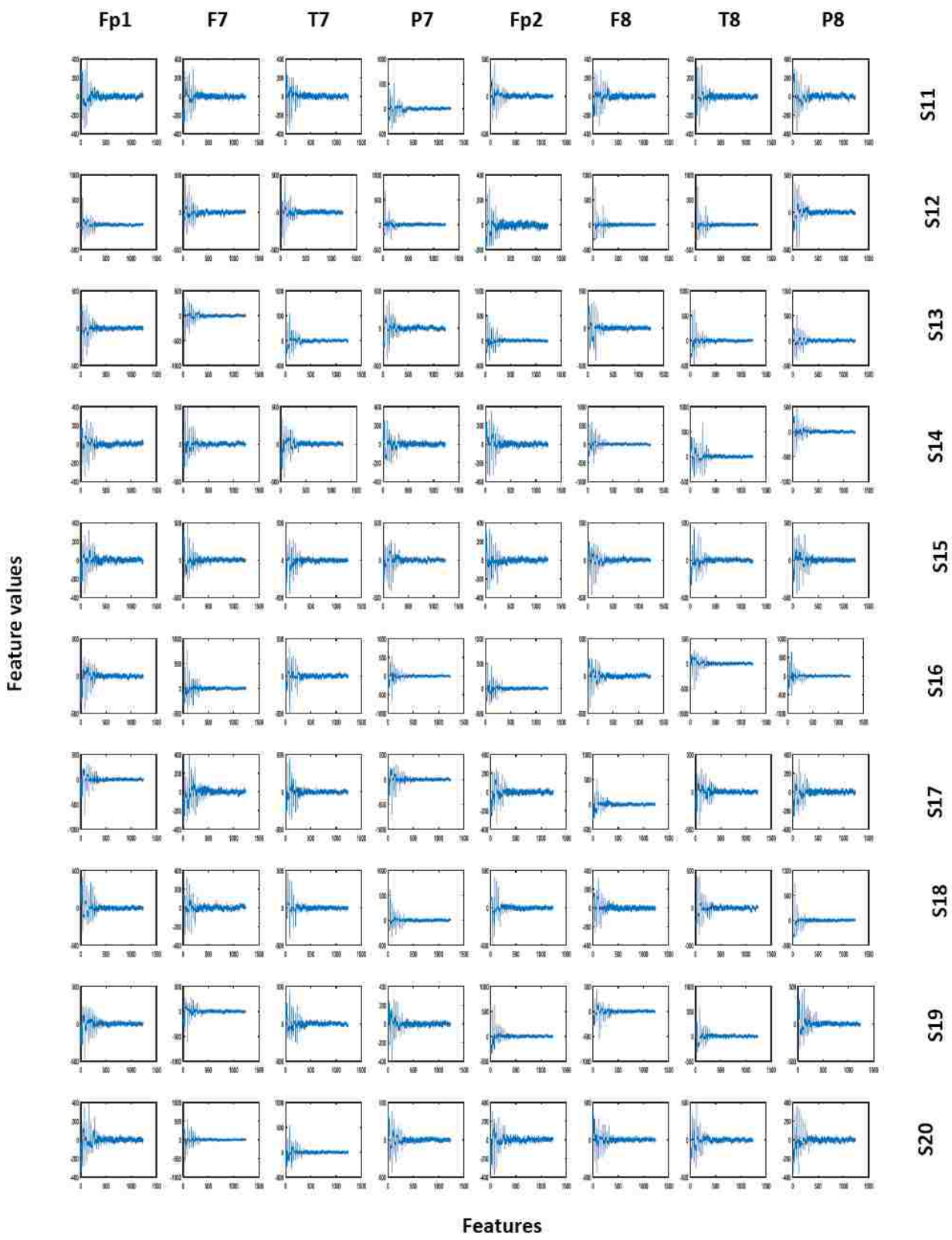
As shown in Figure 6.2, for all the selected channels and across subjects, the pixel count evenly covers a wide range of pixel intensity which indicate a good contrast properties of the image as well as a similarity in the contrast property among the subjects. In addition, the shape of the histogram plots for some of the channels are similar for each subject while some similarities can also be observed across many subjects. It is from these pixel intensity values that the respective features are computed for a subject-independent emotion recognition task that is employed in this dissertation.

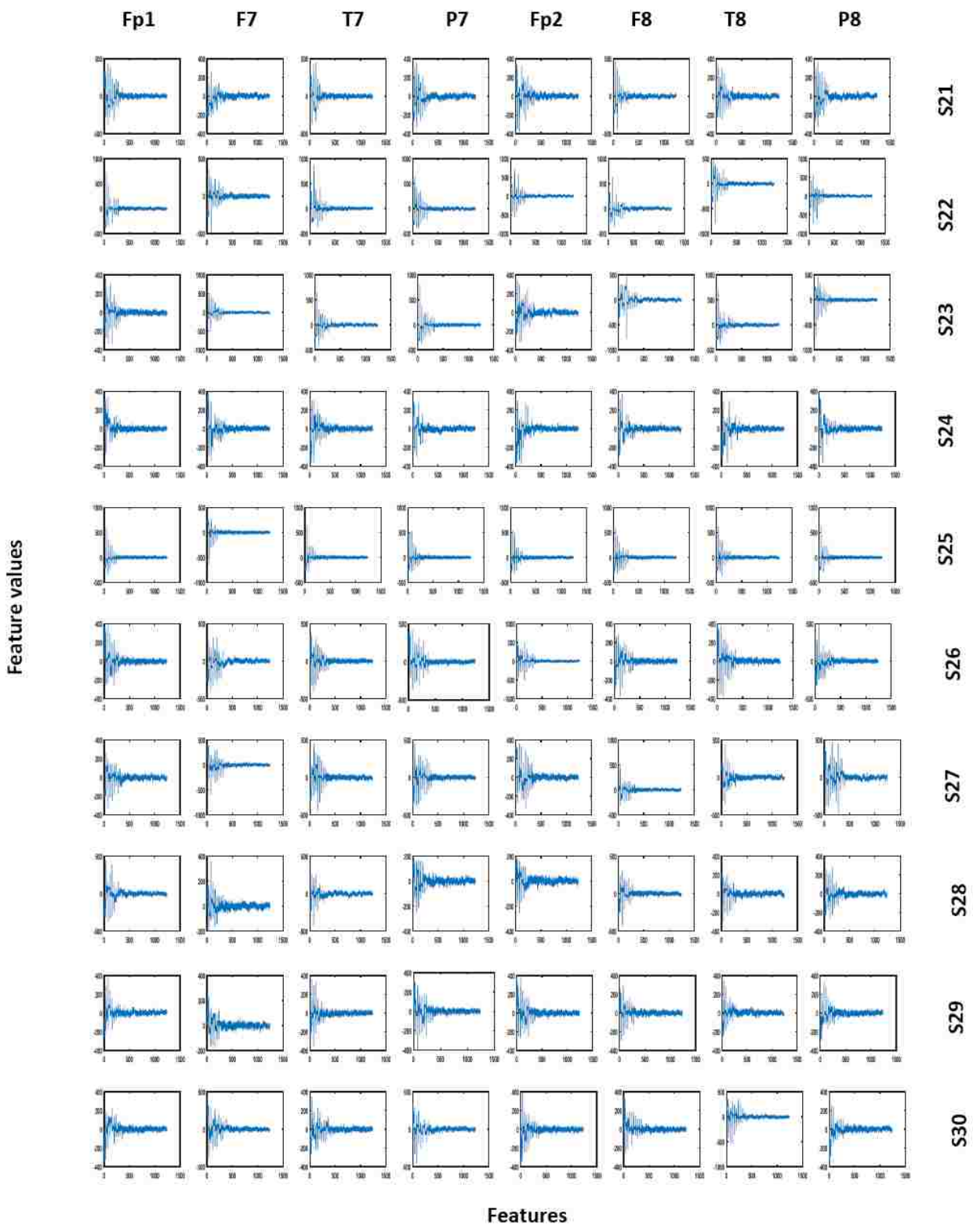
For instance, the Histogram features of the EEG signal modality of the dominance dimension for the 8 channels and trials are presented in Figure 6.3. Mathematically, the extracted features shown in Figure 6.3 are said to be similar across subjects as they have the same shape, but usually of varied sizes as a result of the varied feature value otherwise termed amplitude/intensity of the signal. This is an indication of the robustness of the inverse Fisher transformation method that was applied as well as the feature extraction algorithm in revealing the similarities inherent in the physiological data of the various subjects.

Furthermore, for the Distress Phase model emotion representation using the EEG modality data, the Histogram features for the 8 selected channels and random trials are also presented in Figure 6.4. Despite utilising the same EEG modality data and Histogram

features, the shape and patterns of the features of the Distress Phase model differ from that of the dominance dimension as observed in the dense nature of the Distress Phase model features with varied amplitude/intensity values. However, as observed in Figure 6.4, there exist a strong similarity in the patterns of the features across each subject, channel and trial thus indicating the robustness of the data transformation technique as well as the Histogram features employed by revealing the similar characteristics inherent in the signal across the subjects thereby enhancing classification performance.


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
Feature values
Features


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
Feature values
Features


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S21
S22
S23
S24
S25
S26
S27
S28
S29
S30
Feature values
Features

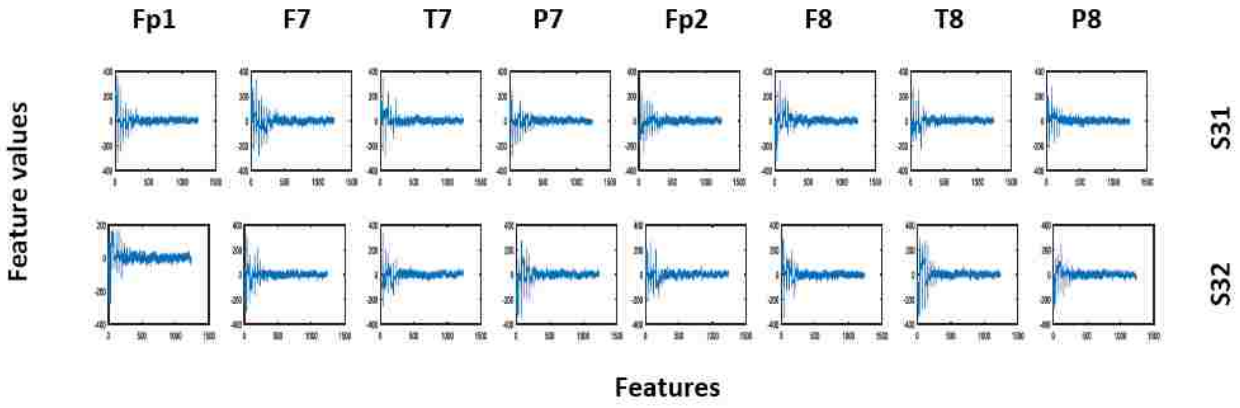
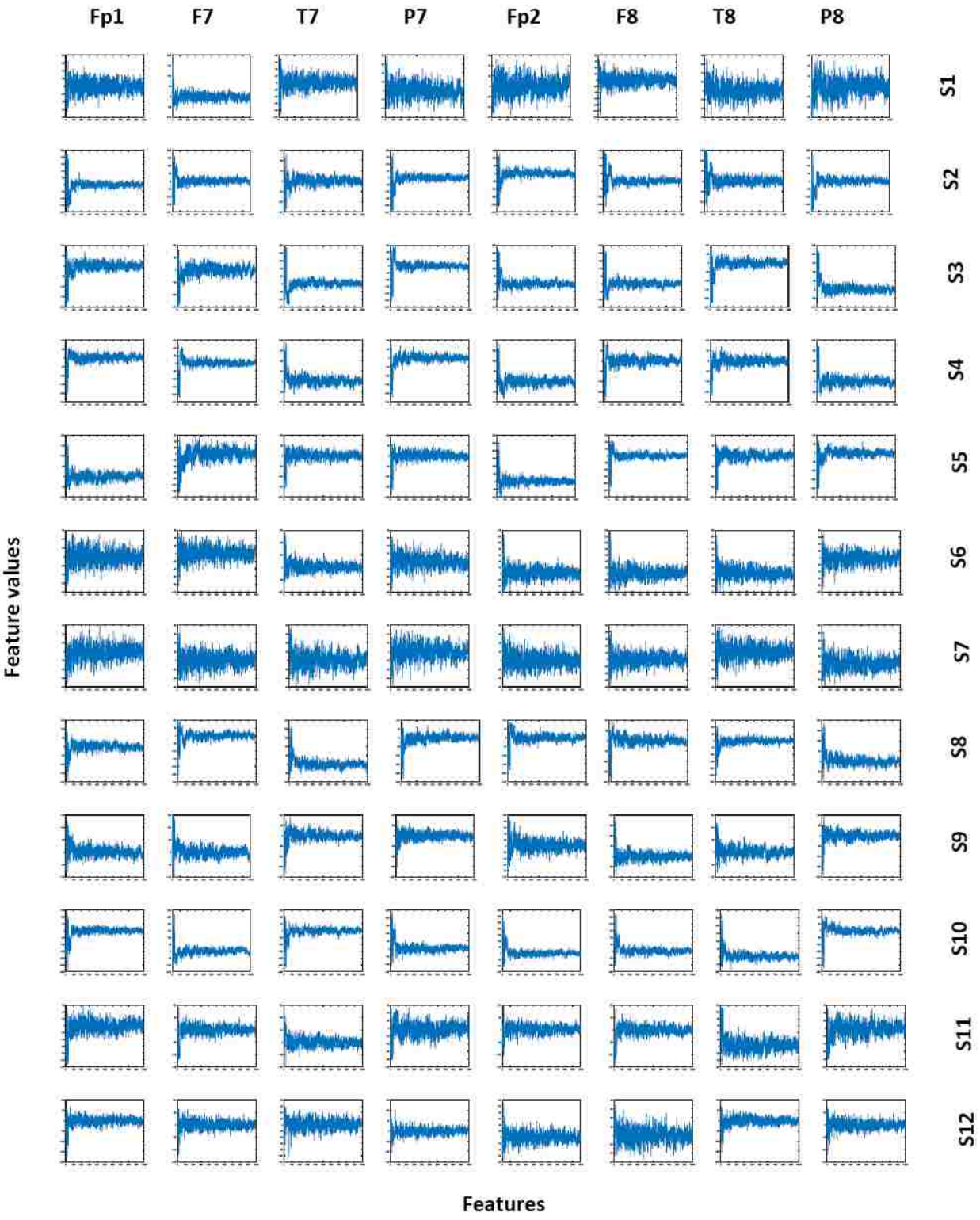
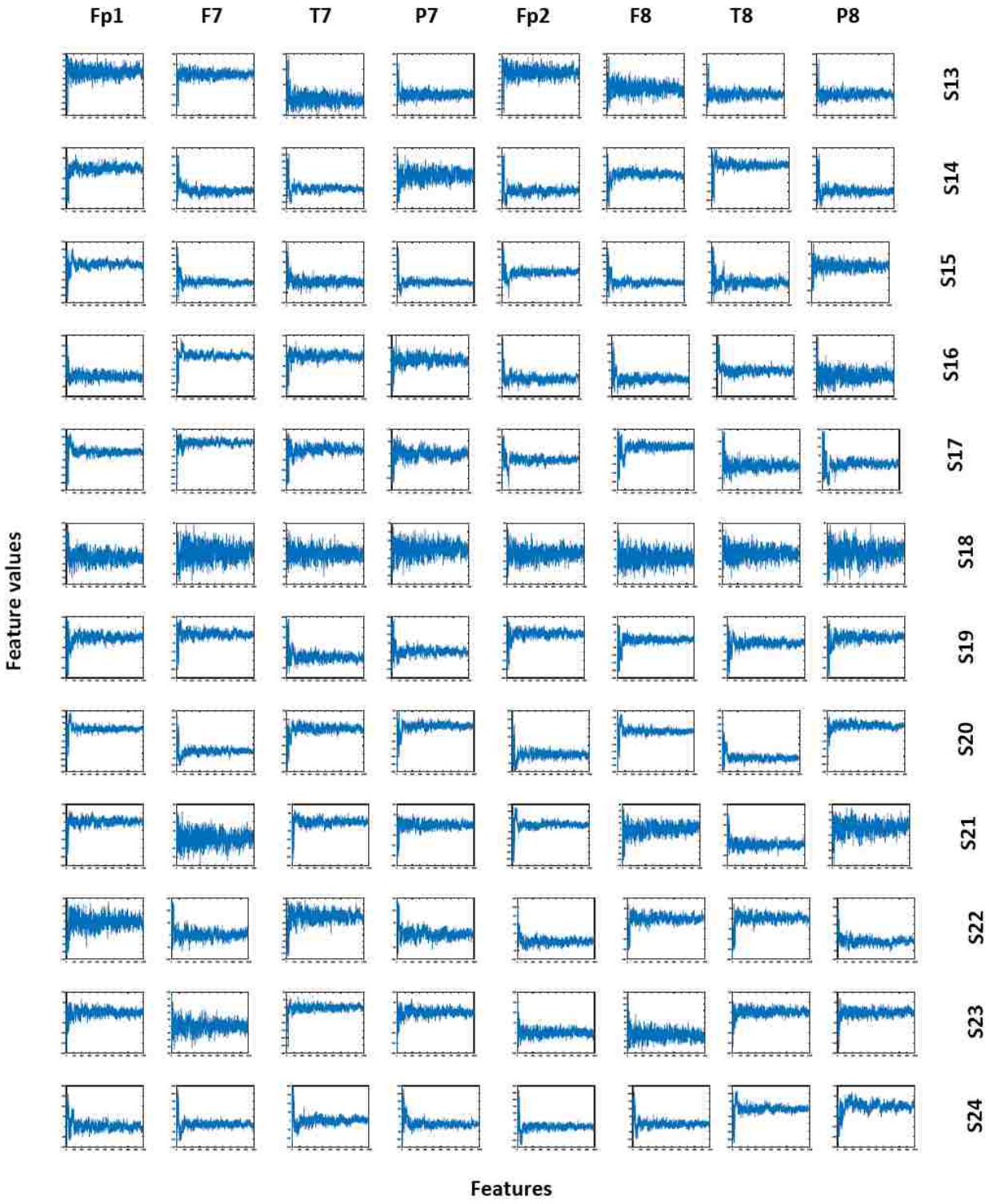


Figure 6.3: Sample Histogram features of dominance dimension of EEG modality data.


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
Feature values
Features


Fp1
F7
T7
P7
Fp2
F8
T8
P8
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
Feature values
Features

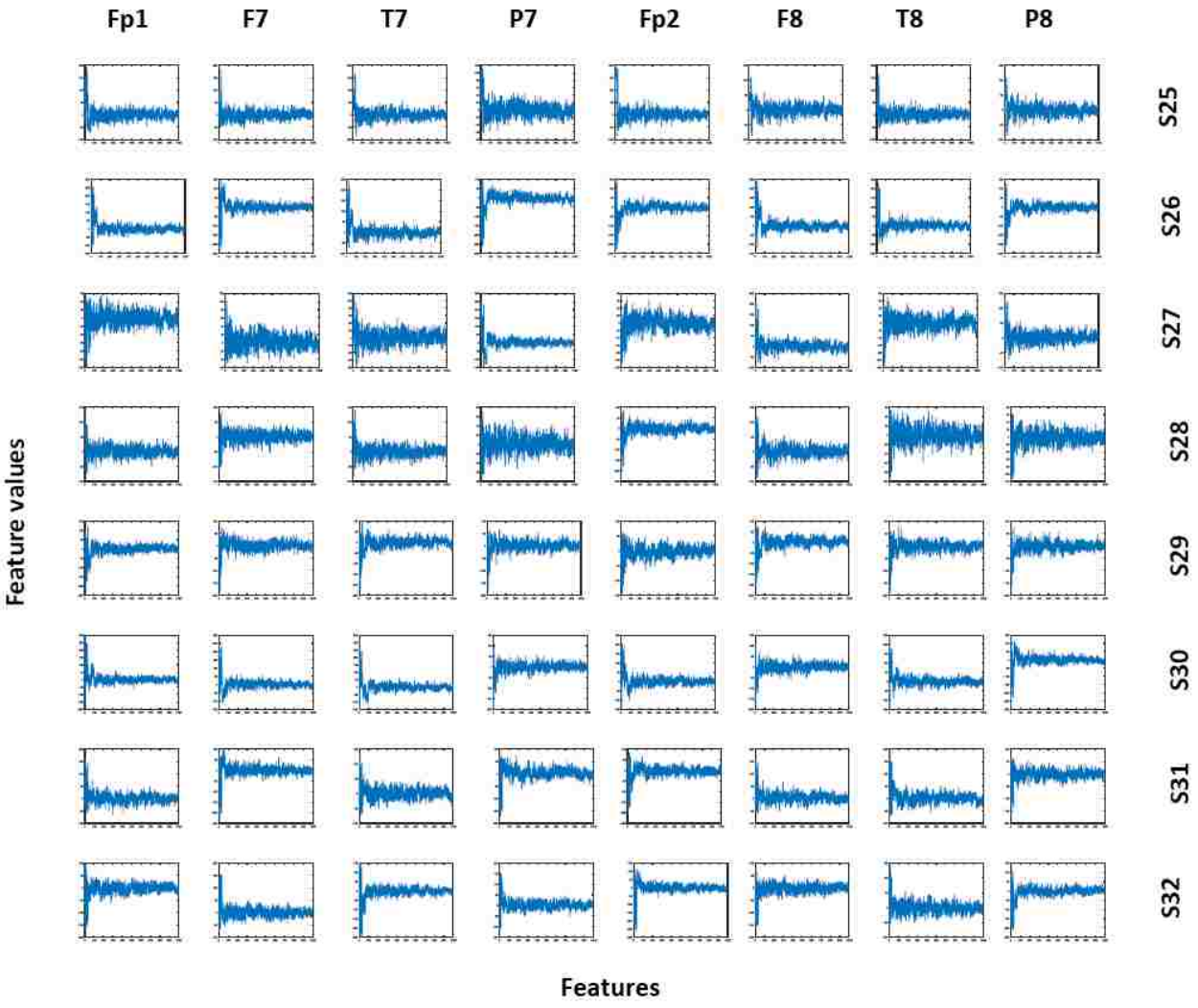
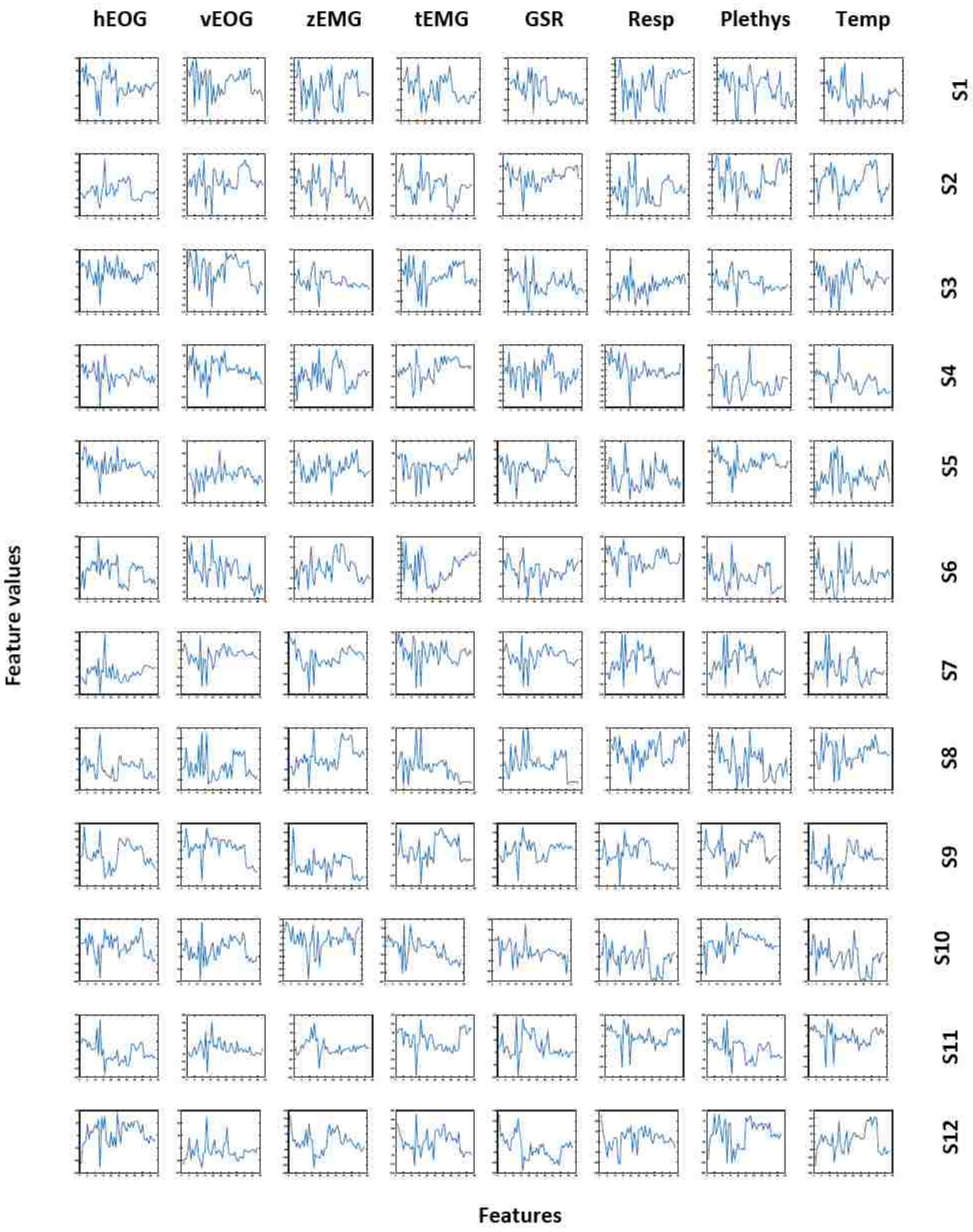
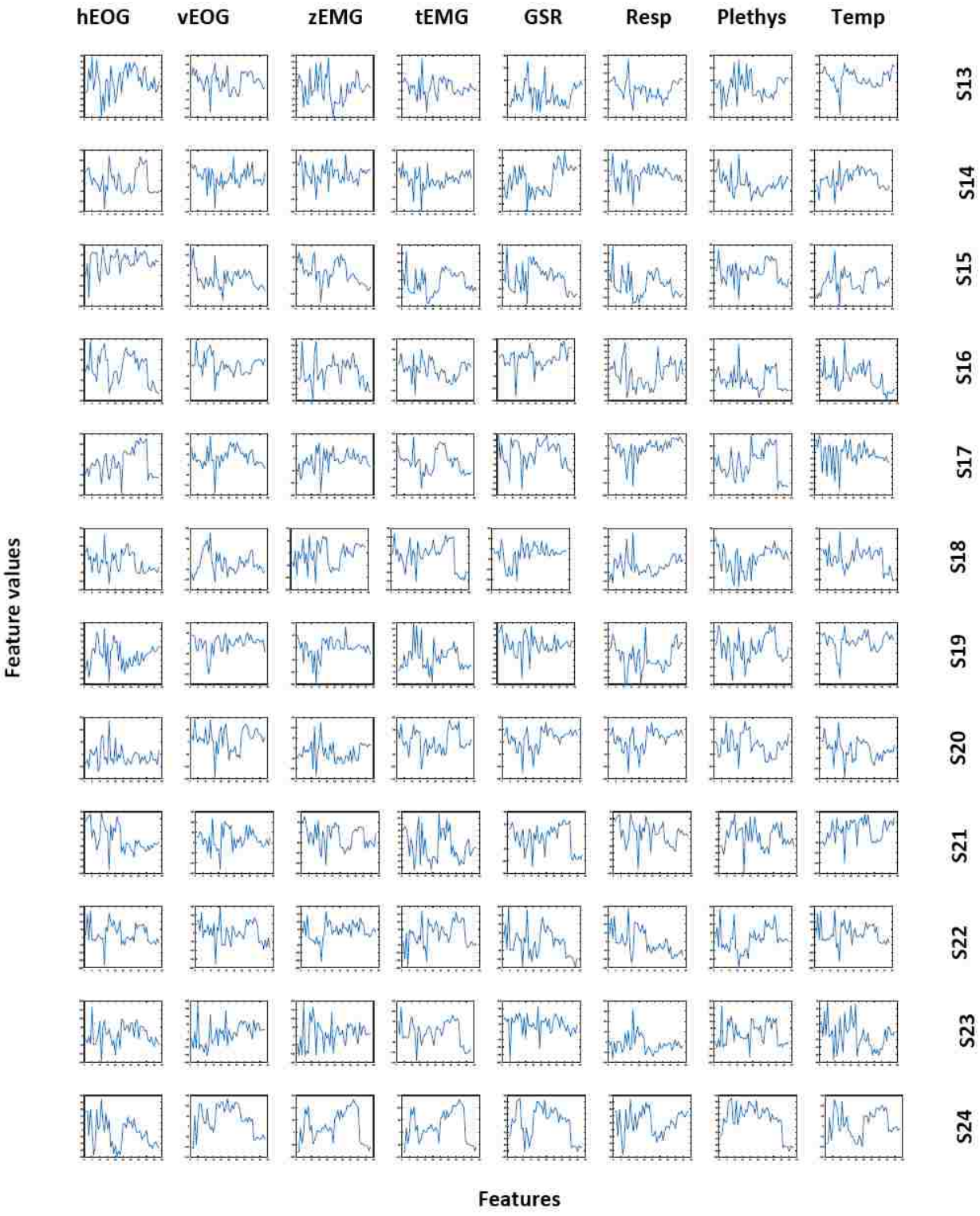


Figure 6.4: Sample Histogram features of distress phase model of EEG modality data.

On the other hand, using the peripheral physiological modality data comprising of the EOG, EMG, GSR, Respiration, Plethysmograph and temperature data, the HOG features extracted along the dominance dimension, the LBP features extracted along the liking dimension as well as the HOG features of the Distress Phase model emotion representation are shown in Figures 6.5, 6.6 and 6.7 respectively. The aim is to show the differences in shapes and patterns between the different modalities as well as the feature descriptors while also revealing the similarities across the subjects and channels within a particular feature. The strength of the data transformation technique and features employed thus confirm the various impressive results obtained with the peripheral


hEOG
vEOG
zEMG
tEMG
GSR
Resp
Plethys
Temp
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
Feature values
Features


hEOG
vEOG
zEMG
tEMG
GSR
Resp
Plethys
Temp
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
Feature values
Features

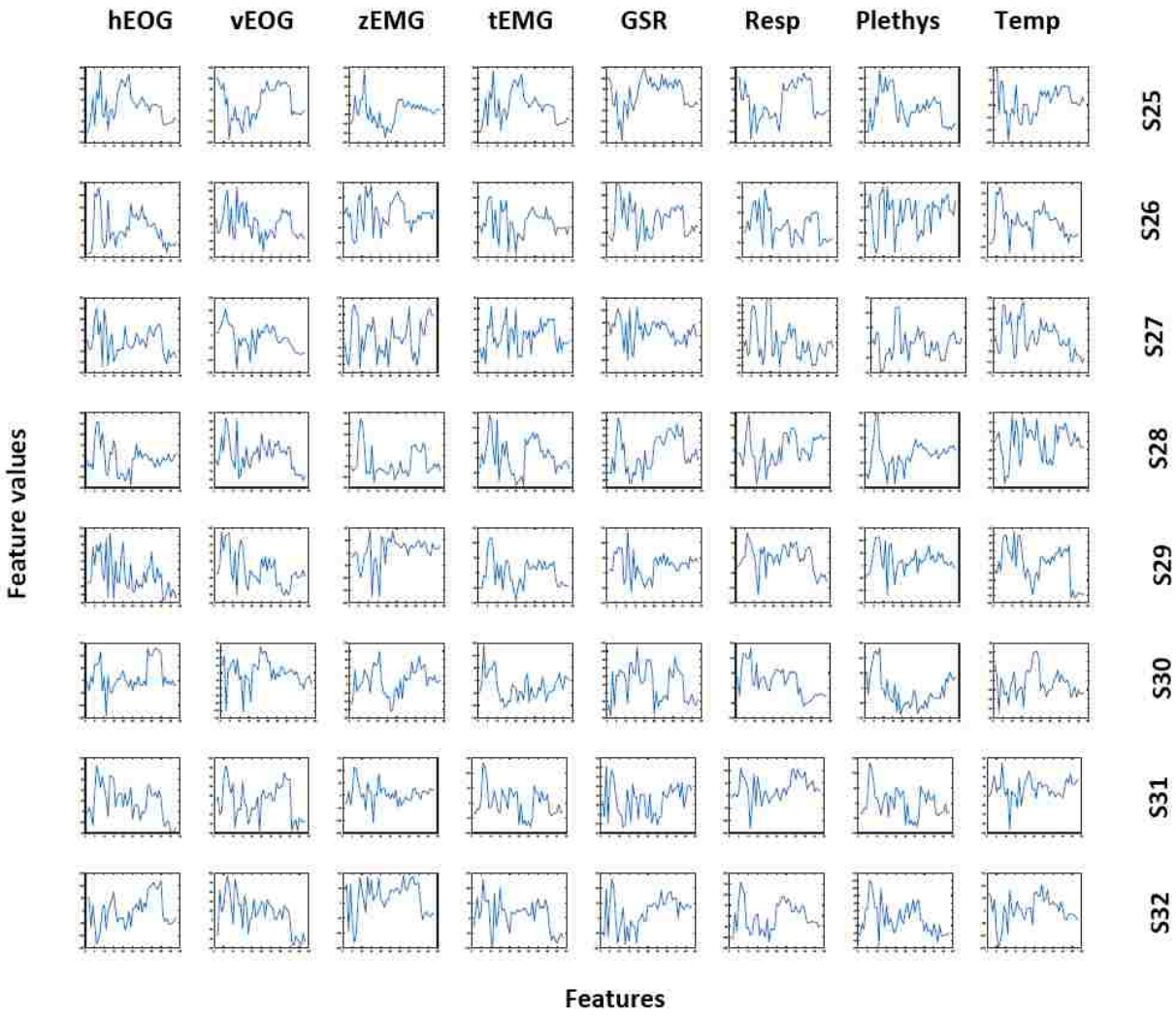
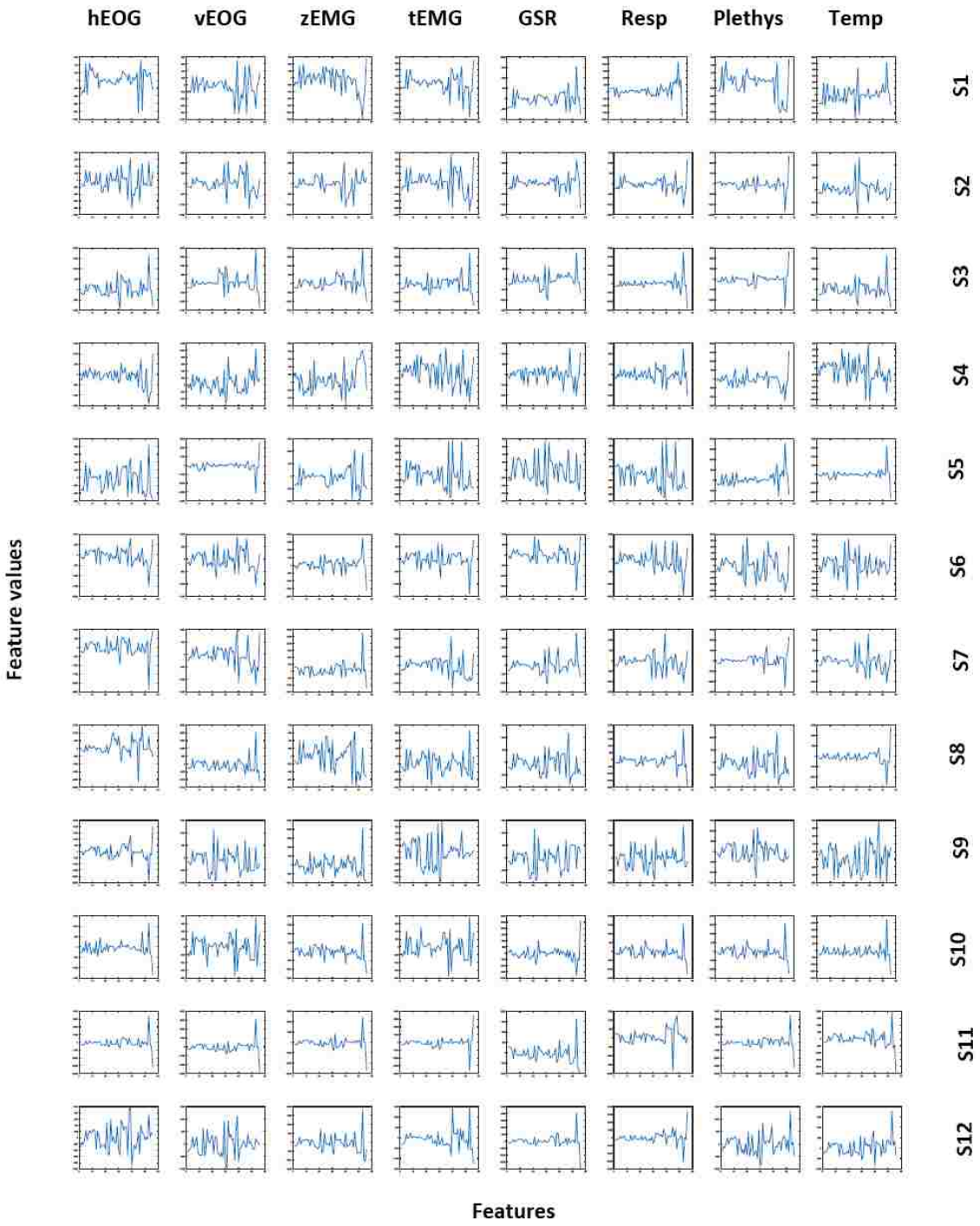
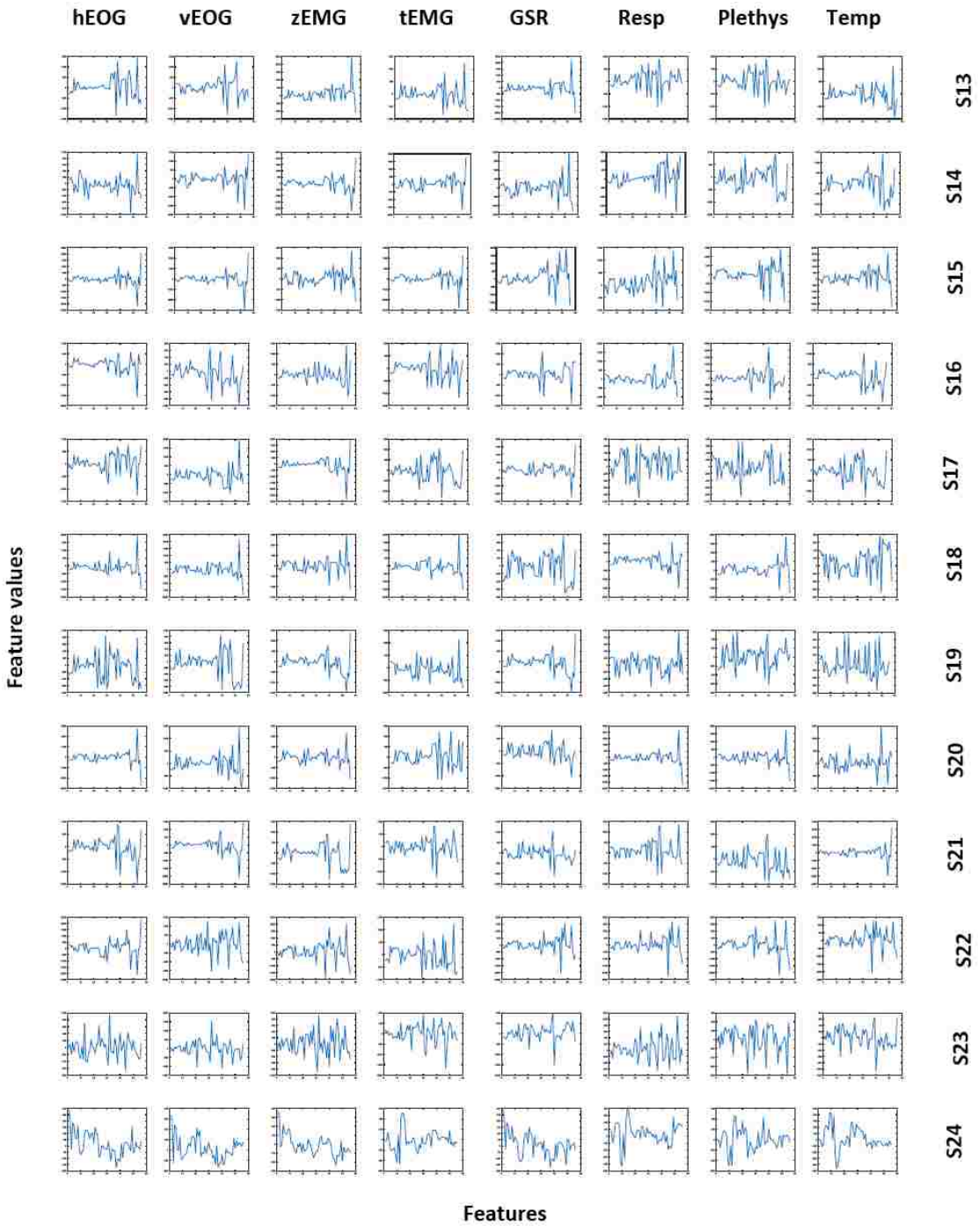


Figure 6.5: Sample HOG features of dominance dimension of peripheral modality data.


hEOG
vEOG
zEMG
tEMG
GSR
Resp
Plethys
Temp
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
Feature values
Features


hEOG
vEOG
zEMG
tEMG
GSR
Resp
Plethys
Temp
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
Feature values
Features

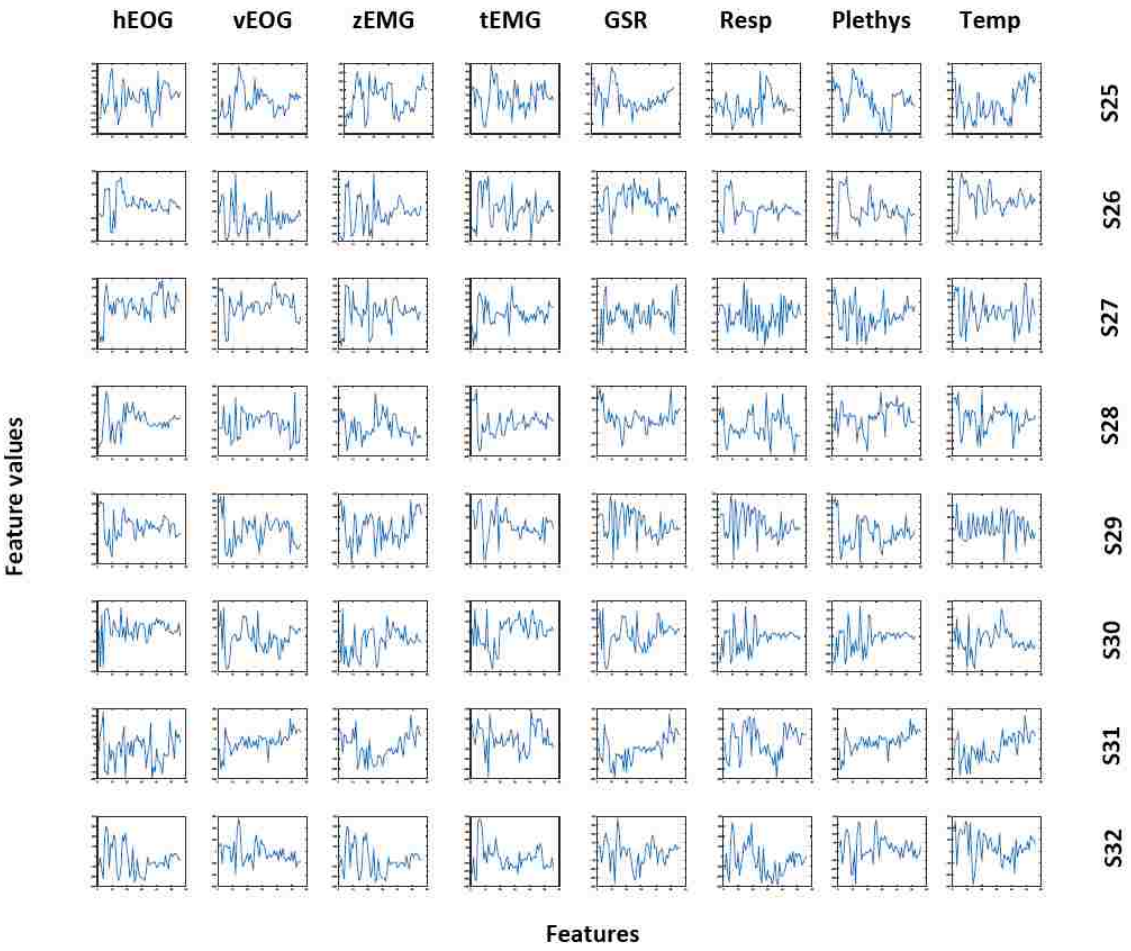
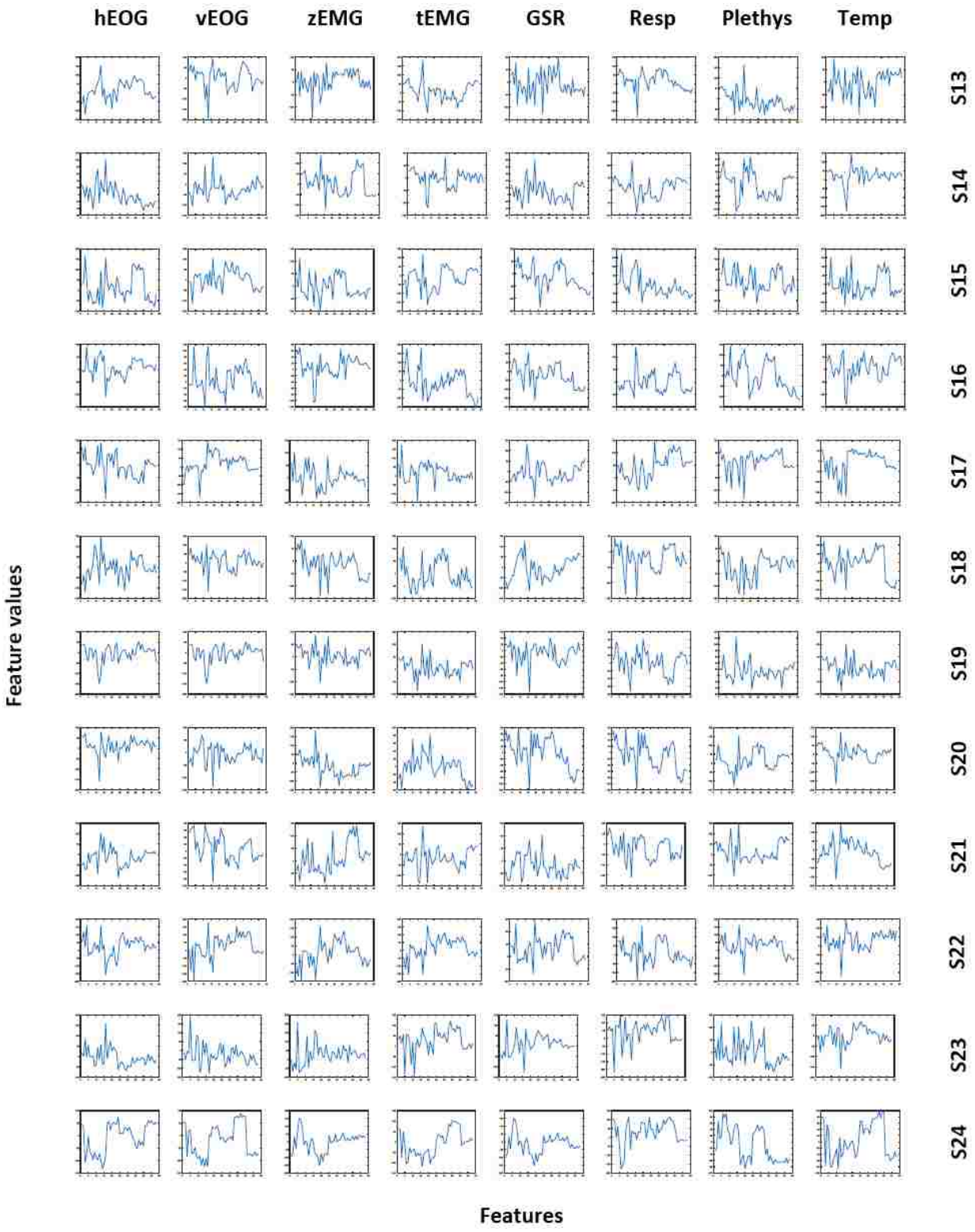


Figure 6.6: Sample LBP features of liking dimension of peripheral modality data.


hEOG
vEOG
zEMG
tEMG
GSR
Resp
Plethys
Temp
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
Feature values
Features


hEOG
vEOG
zEMG
tEMG
GSR
Resp
Plethys
Temp
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
Feature values
Features

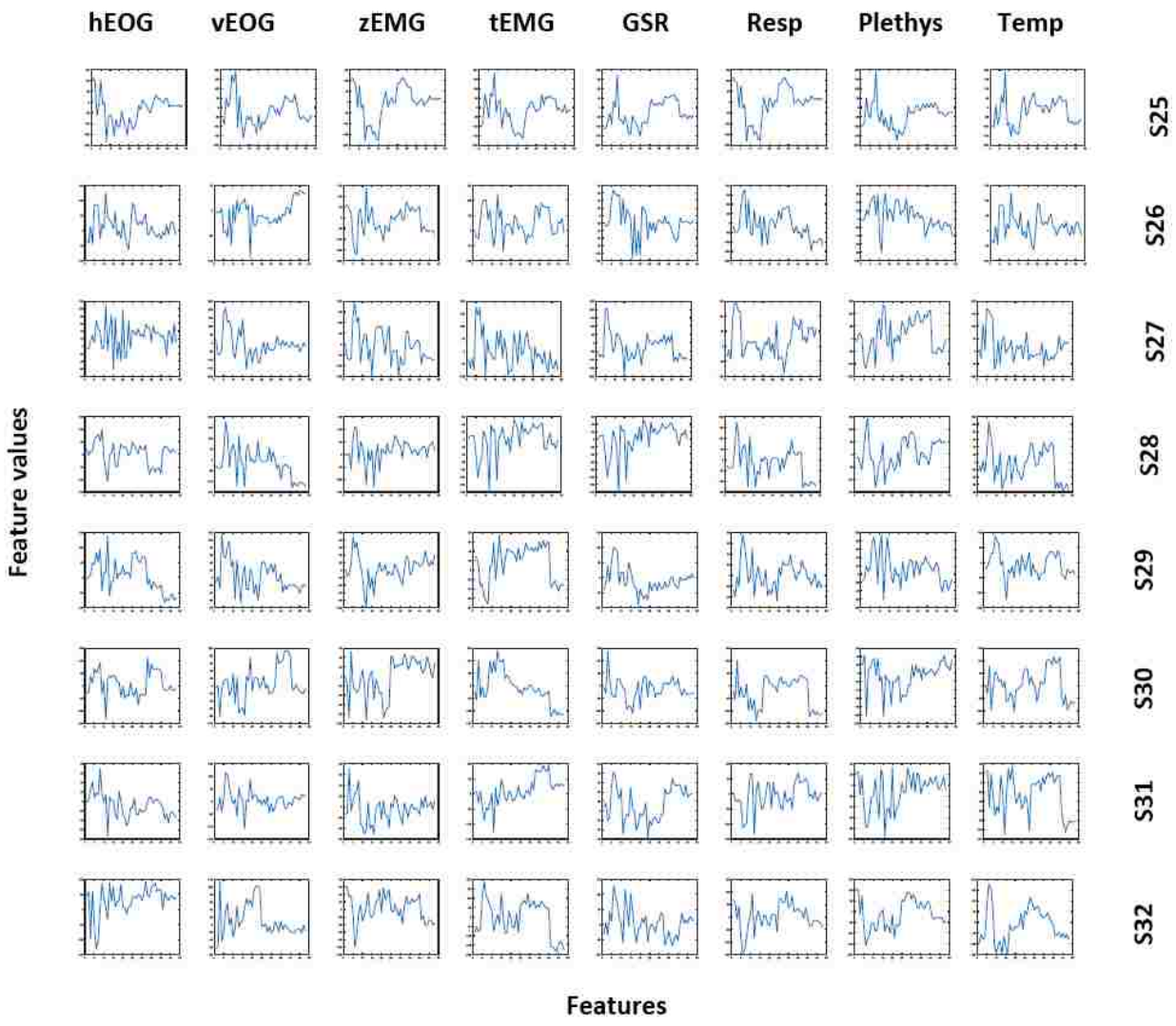


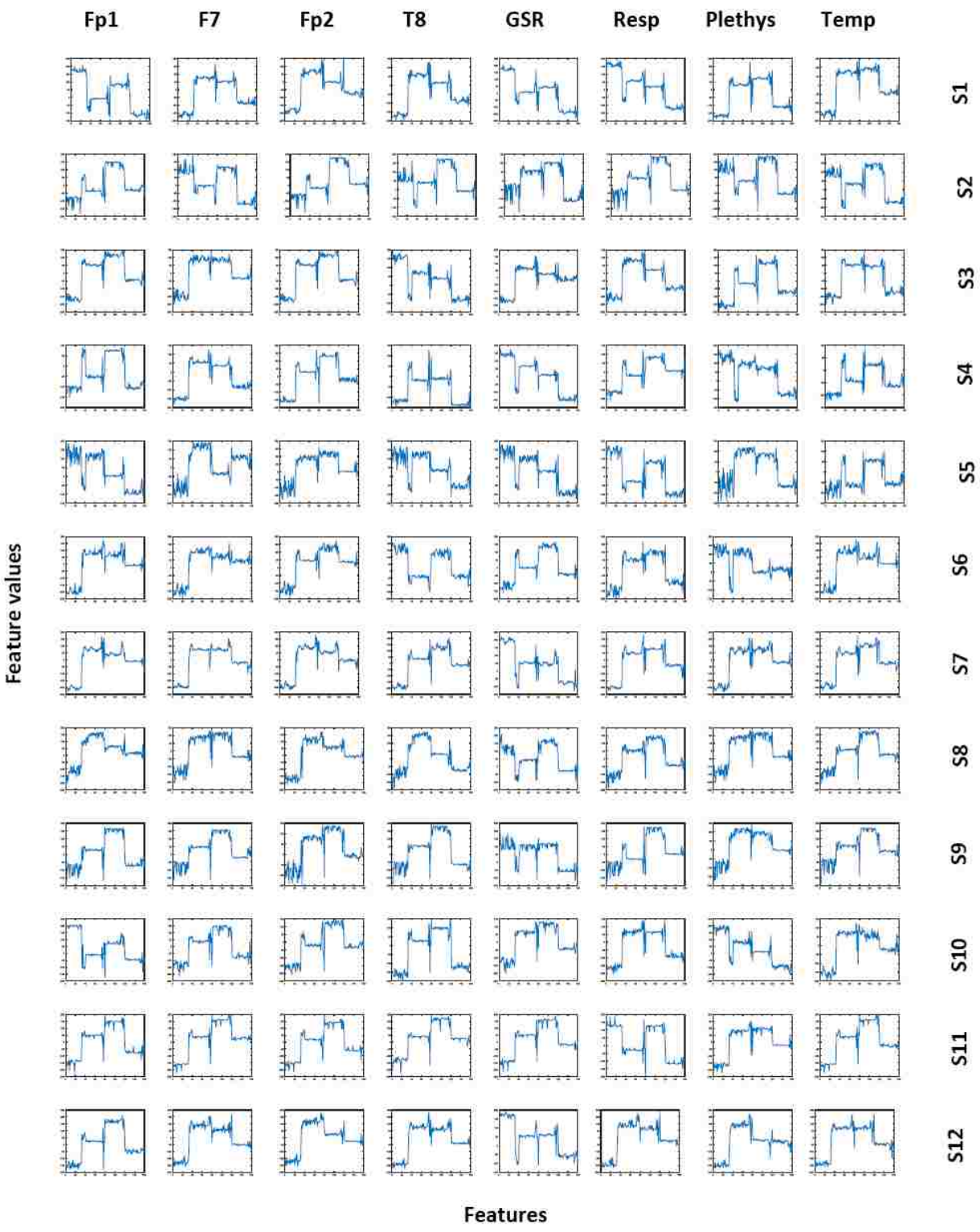
Figure 6.7: Sample HOG features of distress phase model of peripheral modality data.

physiological modality and the HOG features as revealed in the features' similarity across the subjects.

Lastly, the shapes and patterns of the HOG as well as the Histogram features of the fused modality data for the dominance dimension and the Distress Phase model are presented in Figures 6.8 and 6.9 respectively. Having been stated earlier to be connected to human emotions, only the feature plots of 8 channels comprising of 4 EEG channels – Fp1, F7, Fp2 and T8 and 4 peripheral physiological channels, which are the GSR, Respiration, Plethysmograph and temperature for the 8 random trials are shown for each subject. As observed in the figures, the shape and patterns are unique for each of the

dimensions. However, within each dimension, there is a huge similarity in the shape and pattern of the feature across each subject and channels, thus following the earlier identified trend capable of yielding impressive results as the strength of the data transformation method and feature descriptor employed revealed the inherent similarities thus enhancing the subject-independent classification task.

The various methods, including cross correlation coefficient, cross approximate entropy and mutual information have been applied in detecting similarities in signals' features. However, in this study, the respective features for each modality and dimension are fed directly into a RBFNN pattern classifier in order to determine the performances of each feature and modality by obtaining a classification result of the subject-independent based emotion recognition task.


Fp1
F7
Fp2
T8
GSR
Resp
Plethys
Temp
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
Feature values
Features


Fp1
F7
Fp2
T8
GSR
Resp
Plethys
Temp
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
Feature values
Features

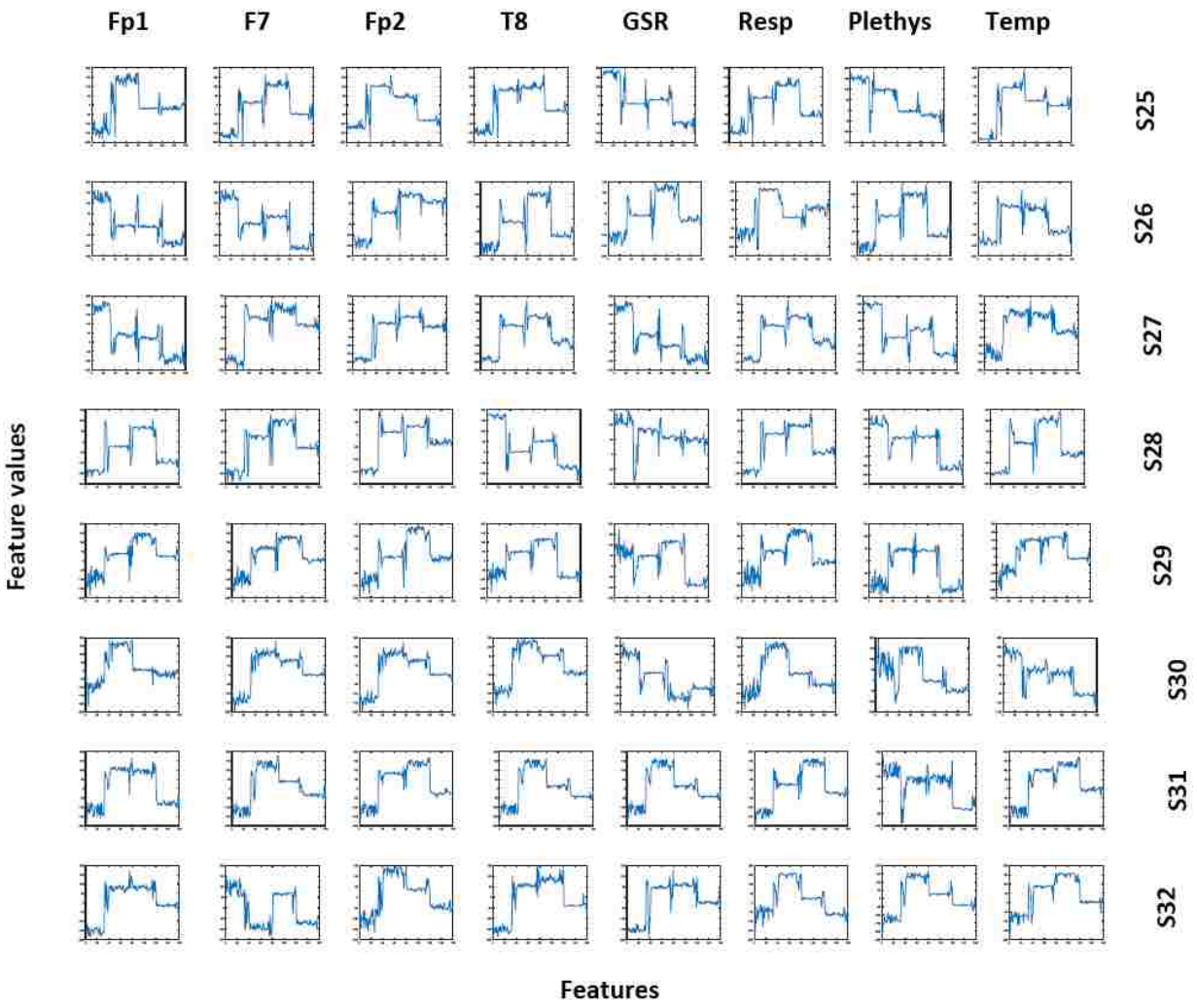
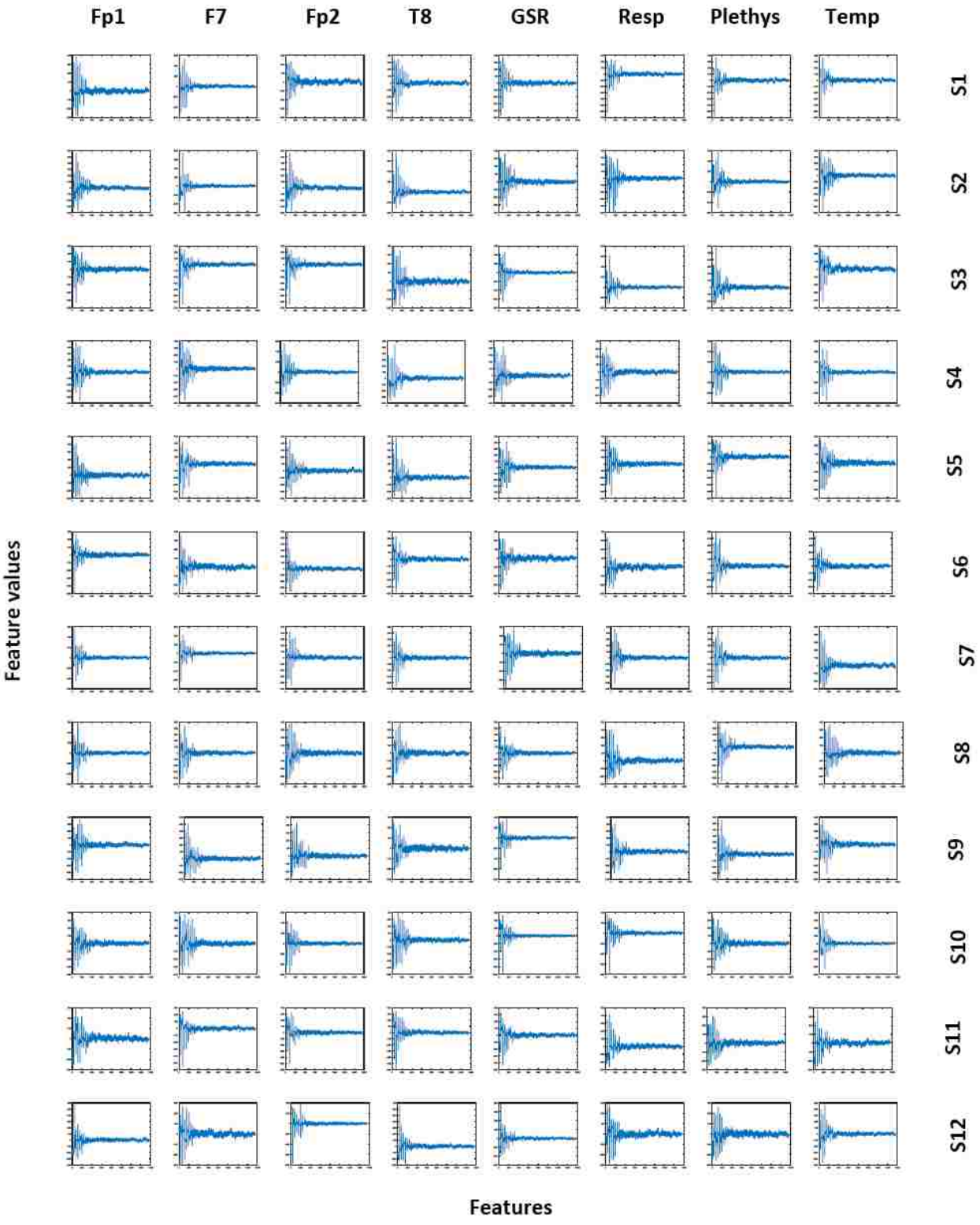
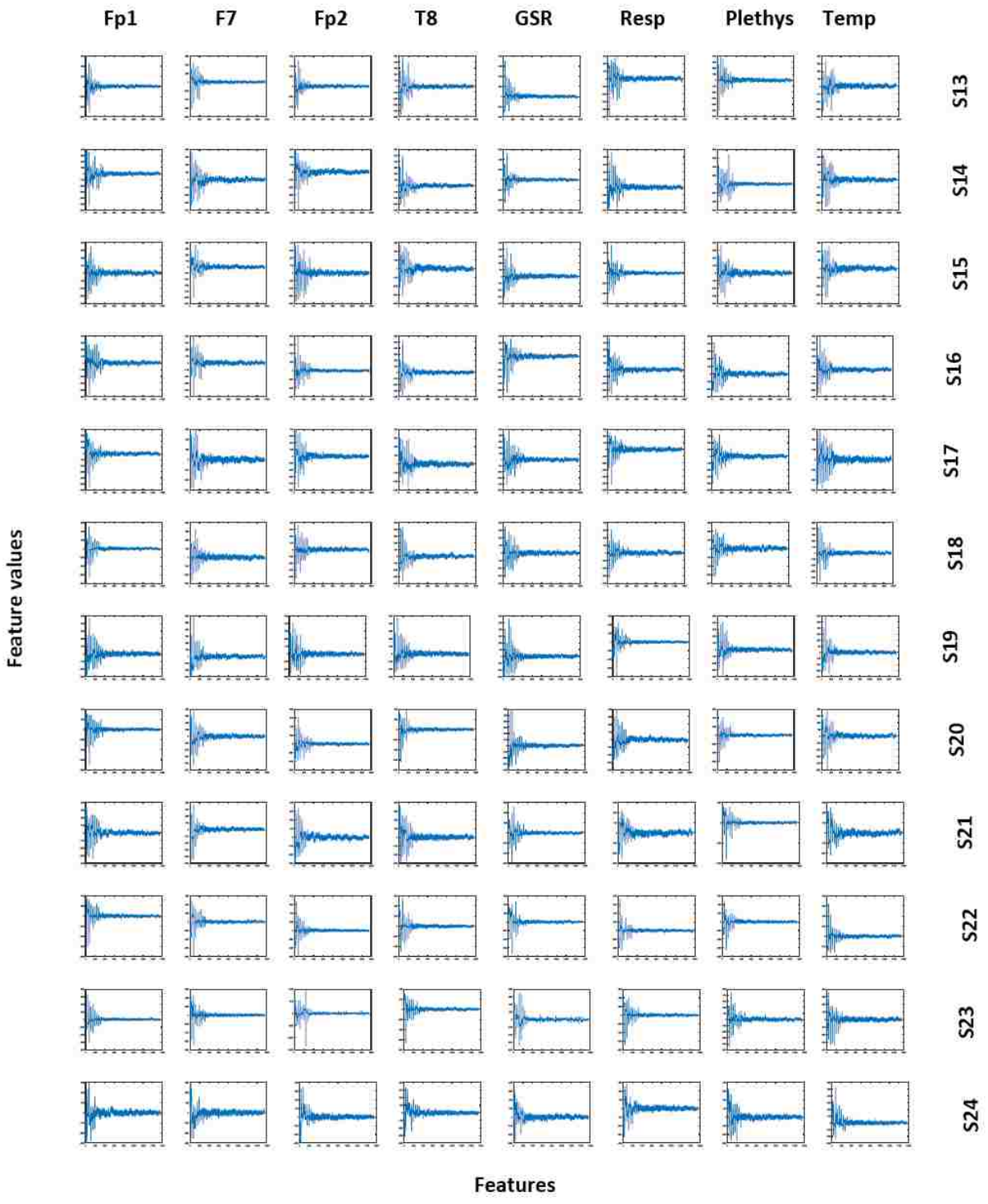


Figure 6.8: Sample HOG features of valence dimension of fused modality data.


Fp1
F7
Fp2
T8
GSR
Resp
Plethys
Temp
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
Feature values
Features


Fp1
F7
Fp2
T8
GSR
Resp
Plethys
Temp
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
Feature values
Features

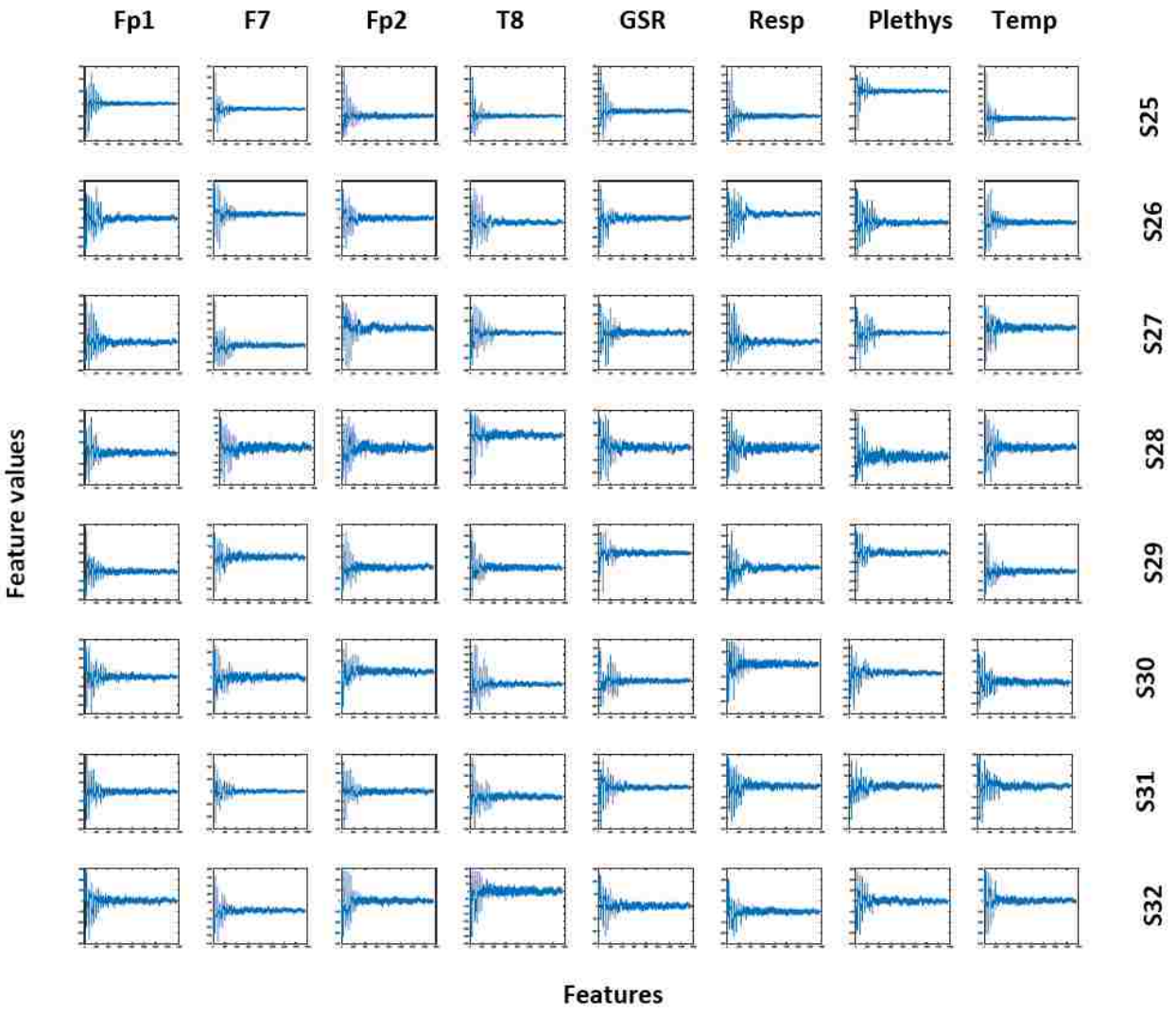


Figure 6.9: Sample Histogram features of distress phase model of fused modality data.

The subjective evaluations carried out are however, complemented as the quantitative analysis and results obtained using the RBFNN pattern classifier and the respective feature descriptors are hereby presented.

6.1 Experimental results of peripheral modality

For the arousal scale results, which consist of classifying human emotion along the high/low arousal classes, the RBFNN classifier using the HOGPEPS features with 50 clusters per class which represents 100 neurons in the hidden layer, achieved a recognition accuracy of 72.11% with a Mean Square Error (MSE) of 0.7187. Since the number of neurons utilised is far below the available and possible 1280 neurons that can be used, therefore the number of clusters which determines the number of neurons that is used in the hidden layer was varied, between 50-250 in a step of 50. It was noticed as shown in Table 6.1, that as more clusters/neurons are added, the recognition accuracy increases until it peaked at 85.16% with 250 clusters after which the accuracy starts to decline. Thus the best recognition accuracy attained using the HOG features of the peripheral physiological data (HOGPEPS) is 85.16% (MSE=0.0398). This result is indeed very promising and better than the best result of 69.2% obtained with the SVM classifier by Menezes et al. (2017) that used the band waves spectral power density features extracted from the DEAP physiological signals dataset.

The LBPPEPS features which represent the LBP features extracted from the peripheral physiological modality data recorded a recognition accuracy of 73.36% (MSE=0.3300) with 50 clusters per class or 100 neurons in the hidden layer. As the number of neurons is increasingly varied, the accuracy obtained increases.

Table 6. 1: Results of the arousal dimension for the peripheral modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	72.11	0.7187	73.36	0.3300	73.83	0.2854
100	78.75	0.5874	78.75	0.0630	79.84	0.3828
150	82.27	0.3343	84.22	0.2214	80.23	0.2987
200	84.92	0.1630	84.92	0.4973	82.50	0.2391
250	85.16	0.0398	81.72	0.3667	83.98	0.2089

The best recognition accuracy of 84.92% (MSE=0.4973) was however achieved with 200 clusters as the network performance suffers a decline with subsequent increase in the

number of hidden layer neurons as shown in Table 6.1. With the 73.36% (MSE=0.3300) accuracy achieved with just 50 clusters, as against the 72.11% (MSE=0.7187) recorded with the corresponding HOGPEPS features, it initially appeared as if the LBPPEPS will outperform the HOGPEPS because of its higher number of feature vector size which could connote more inherent discriminatory information. However, the best recognition accuracy of 84.92% (MSE=0.4973) achieved with the LBPPEPS is marginally lower by just 0.24% to the result of 85.16% (MSE=0.0398) posted by the HOGPEPS. In terms of network's efficiency, the LBPPEPS result is preferred because, it was achieved with a lower number of neurons in the hidden layer as an increase of 100 neurons yielding just 0.24% marginal accuracy is considered not efficient enough.

The HIMPEPS features which represent the Histogram of Images of the peripheral physiological modality data recorded the best recognition accuracy of 83.98% (MSE=0.2089) using 250 clusters or 500 neurons in the hidden layer as shown in Table 6.1. This result is lower by 1.18% and 0.94% respectively to the best results posted by the HOGPEPS and LBPPEPS indicating that for the peripheral physiological signal modality of the DEAP data set with the method proposed in this study, the pixel intensity values as features for classification is not better than the HOGPEPS and LBPPEPS features.

However, despite the seeming lower performance of the HIM features, the recognition accuracies achieved by the HOGPEPS, LBPPEPS and HIMPEPS are all better than the arousal class best results of 77.19% (Yin et al. 2017), 69.2% by the SVM and 74.0% by the random forest classifiers (Menezes et al. 2017), 71.99% (Zhuang et al. 2017), and 60.9% (Wang and Shang 2013) posted in these various research studies.

As noticed in Table 6.2, the classification results along the valence class for the peripheral physiological modality data follows a similar trend with that of the arousal class, thus indicating the similarity and consistency of the patterns in the features employed in these two classes. With the valence class, the HOGPEPS features achieved a recognition accuracy of 85.94% (MSE=0.4246) with 500 neurons in the hidden layer as shown in Table 6.2. This result is marginally better than the 85.16% (MSE=0.0398) posted by the same features under the arousal class label.

In the same vein, a recognition accuracy of 83.20% (MSE=0.3057) was recorded by the LBPPEPS features with the valence class of the peripheral physiological modality data. This was achieved with 400 neurons in the hidden layer of the RBFNN and the performance declined with more additional neurons. This result is lower than the results of the HOGPEPS features of the valence class, including also the corresponding LBPPEPS result posted under the arousal class as well as the results of the HOGPEPS and HIMPEPS features of the arousal class.

Also, as shown in Table 6.2, a recognition accuracy of 82.66% (MSE=0.4792) was achieved with the HIMPEPS features with the valence class using 500 neurons in the RBFNN hidden layer. The import of the various results posted by the various features for the valence class of the peripheral modality data is that, the HOGPEPS features have the most informative and discriminatory properties for this modality with which human emotion can be recognised using the DEAP dataset. This is because, the highest results are posted by the HOGPEPS features across the valence and arousal classes as shown in Tables 6.2 and 6.1 respectively. Though, the 84.92% (MSE=0.4973) recorded by the LBPPEPS features could be preferred, if efficiency is to be considered above accuracy as a 100 fewer neurons were expended by the RBFNN for the LBPPEPS features of the arousal class labels of the peripheral physiological modality.

Table 6. 2: Results of the valence dimension for the peripheral modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	72.58	0.5655	72.66	0.2560	71.88	0.4860
100	79.69	0.3468	79.53	0.4364	77.50	0.0226
150	82.50	0.2009	81.41	0.2704	80.23	0.4129
200	83.83	0.3798	83.20	0.3057	82.66	0.4792
250	85.94	0.4246	82.58	0.3633	82.27	0.4686

However, these best results achieved for the peripheral physiological modality using the HOGPEPS, LBPPEPS and HIMPEPS features for the valence class are better than the 69.10% (Zhuang et al. 2017), 76.17% (Yin et al. 2017), 51.2% (Wang and Shang

2013) and 58.2% (Li et al 2015) reported in various recent studies that utilised the DEAP data set with classification done along the valence class, thus indicating the robustness of the proposed data transformation method and the feature extractors applied in this study.

As contained in Table 6.3, the results obtained by the dominance emotion representation class with the peripheral physiological modality data are as follows. A classification accuracy of 87.34% (MSE=0.3775) was achieved with 500 neurons in the hidden layers of the RBFNN using the HOGPEPS features. This is the highest result achieved so far in this modality across all the features employed. However, with 100 neurons less in the hidden layer, a classification accuracy of 87.03% (MSE=0.1609) was recorded. In terms of comparing efficiency and accuracy, the 87.03% recognition results could be preferred because of its associated lesser network complexity and computation time.

From the result obtained, the respective HOG features of the high and low dominance classes was further investigated to determine some of the inherent characteristics in the features and classes that could have imparted this impressive result. As shown in Figure 6.10, 6 subjects (S4, S8, S12, S18, S22 and S27) were uniformly selected and, the plots of the HOG features for the high and low dominance classes revealed different feature values otherwise termed amplitude or intensity in this study. It was observed that the high dominance class has a higher amplitude than the low dominance class, thereby enhancing easy classification by the pattern recognizer. For instance, amplitude values of 197, 187, 179, 203 and 177 respectively, were recorded for the high dominance class for subjects S8, S12, S18, S22 and S27. Conversely, the low dominance class has amplitudes of 195, 173, 152, 198 and 157 respectively in the subjects mentioned and only subject S4 has a higher amplitude for the low dominance class than the high dominance class. However, the trend of a high amplitude value for high dominance dimension and low amplitude value for low dominance dimension is predominant among all the subjects and further contributes to the classification result obtained.

In emotional signal processing, including speech signal, an amplitude is otherwise referred to as intensity and can be computed in various ways which are a measure of the maximum change in a quantity that occurs when the signal is being transmitted (Glenn 2010) as a peak amplitude is a measure of emotional intensity. It has also been established in the literature that as a general rule, the larger the amplitude, the greater is the intensity of the signal (Glenn 2010) including emotional speech and physiological data. In addition, an increase in emotional intensity will trigger an increase in performance up to an optimal point (Nasoz et al. 2004). However, human emotions are associated with amplitude fluctuations and a highly significant effect of amplitude of various emotions does exist (Burkhardt 2005; Hammerschmidt and Jurgens 2007; Hartmut and Christian 2008). Emotions such as fear, sadness, disgust, joy and boredom have been detected with high, medium and low amplitudes expressed by various subjects (Hartmut and Christian 2008; Lech et al. 2018).

On the other hand, the LBPPEPS features recorded the best recognition accuracy of 84.92% (MSE=0.1271) using 400 neurons in the hidden layer. A decline in recognition performance was recorded as more neurons are added. However, both the results of 87.34% (MSE=0.3775) and 87.03% (MSE=0.1609) posted for the dominance class by the HOGPEPS features as shown in Table 6.3 is better than the 84.92% (MSE=0.1271) of the LBPPEPS features and also agrees with the trend earlier recorded in both the arousal and valence classes where the HOGPEPS features also outperforms the LBPPEPS features. This further confirms that the HOGPEPS features contain more useful information and discriminatory qualities for human emotion recognition using the DEAP physiological signal data set with classification done along valence, arousal and dominance labels.

With the Histogram of Image features, the best recognition accuracy of 84.14% (MSE=0.1700) was recorded by the dominance class using 500 neurons in the hidden layer of the RBFNN classifier. This performance, though falls short of the HOGPEPS and LBPPEPS features' results under the dominance class, but it is better than the 83.98% (MSE=0.2089) and 82.66% (MSE=0.4792) respectively recorded by the HIMPEPS features of the arousal and valence classes respectively. So far, the HOGPEPS features

and the dominance class are the best combination that yielded the highest recognition accuracy for the peripheral physiological modality using the DEAP dataset. This indicates clearly that the level of submissiveness/in-control of participants as contained in their physiological signals in response to emotional feelings using the 40 emotion elicitation music videos is high. In addition, this indicates the ability of the participants to correctly quantitatively report their emotional feelings. The various best recognition accuracies achieved by the HOGPEPS, LBPPEPS and HIMPEPS features of the peripheral physiological modality with classification done along the dominance dimension are all better than the 65.1% recorded for the dominance dimension by Wang and Shang (2013) who also used the DEAP dataset.

Table 6. 3: Results of the dominance dimension for the peripheral modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	73.98	0.4688	75.00	0.0185	74.84	0.1337
100	80.47	0.1096	82.27	0.1766	81.48	0.1377
150	86.48	0.4033	84.14	0.0904	83.52	0.1595
200	87.03	0.1609	84.92	0.1271	83.75	0.1082
250	87.34	0.3775	83.83	0.1528	84.14	0.1700

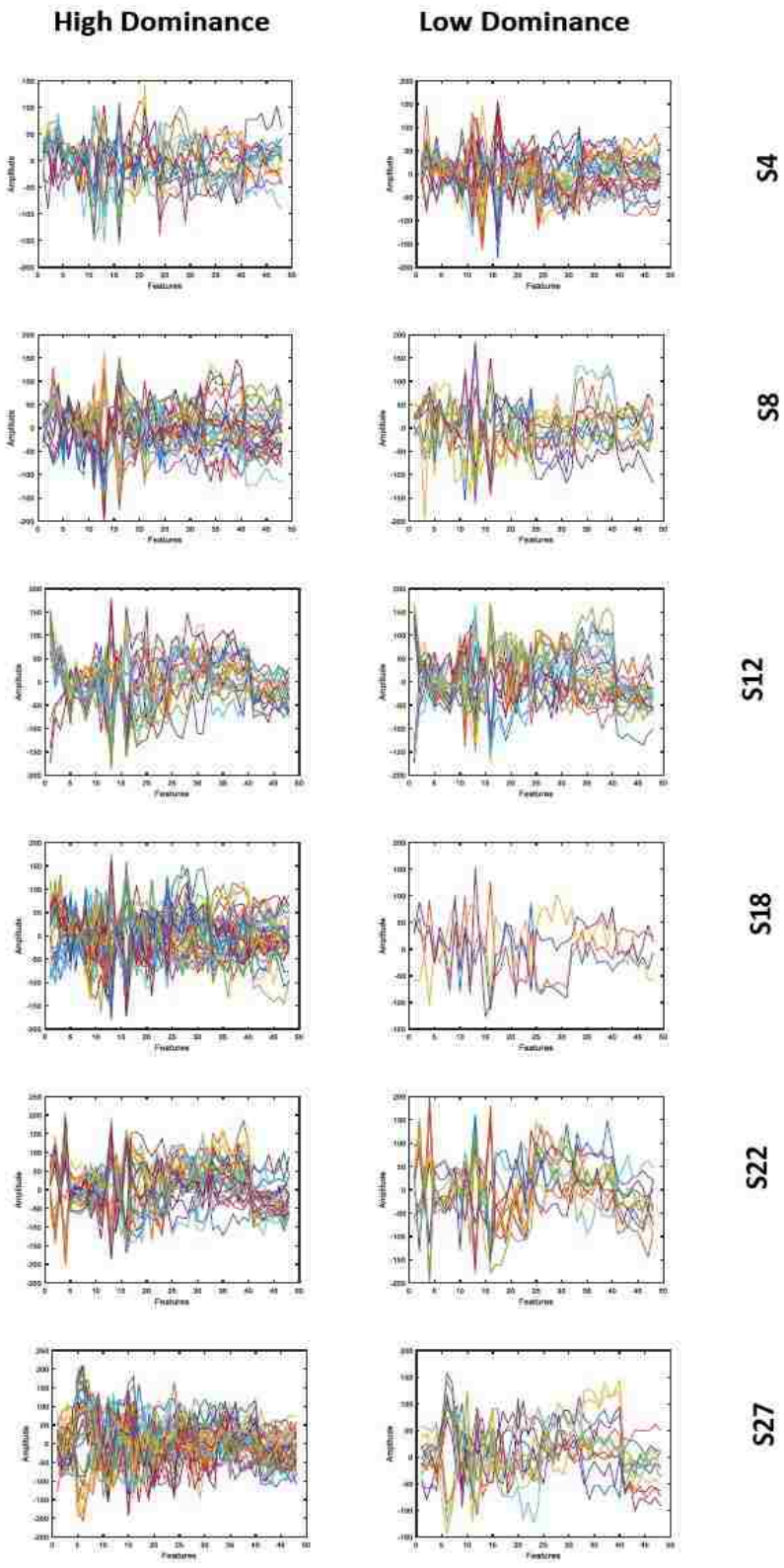


Figure 6.10: Images of sampled HOG features of dominance dimension of peripheral modality data.

Liking is the next emotion class that was considered. As shown in Table 6.4, for the HOGPEPS features, the best recognition accuracy of 83.52% (MSE=0.1647) was achieved by the RBFNN pattern classifier with 400 neurons in the hidden layer as a decline in performance was noticed as more neurons are added. This performance falls short of the results obtained with the same features under the arousal, valence and dominance labels. However, with a cluster size of 150, the LBPPEPS features posted 86.33% (MSE=0.0645) as its best recognition performance, which is more than all the results achieved with the respective features under the arousal, valence, dominance and liking class except the 87.34% (MSE=0.3775) result of the HOGPEPS features with classification applied along the dominance class. This might suggest that more textural discriminatory and useful information is inherent in this LBPPEPS feature set of the liking dimension for the peripheral modality.

This LBP result of the liking dimension was further investigated to determine the inherent distinctiveness in the features and class among the subjects that necessitated the recognition's performance. As shown in Figure 6.11 and as earlier done with the HOG result of the dominance dimension, the 6 subjects (S4, S8, S12, S18, S22 and S27) uniformly selected to post different feature values otherwise called amplitude for the like and dislike classes. For instance, subjects S4, S8, S12, S18 and S27 posted feature values of 1911, 4278, 4656, 4239 and 694 respectively for the like class. These are higher than the feature values of 1873, 2771, 3368, 4154 and 474 recorded respectively by the subjects for the dislike class, thus revealing the pattern classifier's prowess towards achieving the result obtained. The result aligns with the earlier findings in literature that the larger the amplitude, the greater is the intensity of the signal (Glenn 2010) including emotional speech and physiological data as the like class has a higher intensity than the dislike class.

Table 6. 4: Results of the liking dimension for the peripheral modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.94	0.5234	75.63	1.0002	74.30	0.6995
100	81.80	0.5475	82.03	0.1173	81.48	0.6488
150	82.58	0.5642	86.33	0.0645	83.20	0.7421
200	83.52	0.1647	82.50	0.3229	83.13	0.5425
250	82.81	0.1808	77.35	0.0155	80.23	0.6789

A recognition accuracy of 83.20% (MSE=0.7421) as shown in Table 6.4 was the best result posted by the HIMPEPS feature and this was achieved with 300 neurons in the RBFNN hidden layer. Furthermore, all the three best results of 83.52%, 86.33% and 83.20% achieved for the liking dimension of the peripheral physiological modality data by the HOGPEPS, LBPPEPS and HIMPEPS features respectively, are better than the 68.4% and 66.3% classification results achieved by Wang and Shang (2013) and Li et al. (2015) along the liking dimension using the DEAP dataset.

The Distress Phase emotion model developed as reported in this dissertation, while applying the RBFNN classifier recorded a recognition accuracy of 82.03% with 450 neurons in the hidden layer as human emotions were classified along happy, distress and casualty phases using the HOGPEPS features. The aim is to incorporate this emotion dimension model in an affective system such that emergency services can be prompted once a distress phase is detected in order to thwart progression into the casualty phase. The attained recognition accuracy of 82.03% shown in Table 6.5, competes with the results posted by the HOGPEPS features along the emotion representation labels of arousal, valence, dominance and liking since it has been proved in the literature that the performance of a pattern classification system is affected as the number of classes to classify increases (Nguyen, Nguyen and Shimazu 2007; Adetiba and Olugbara 2015).

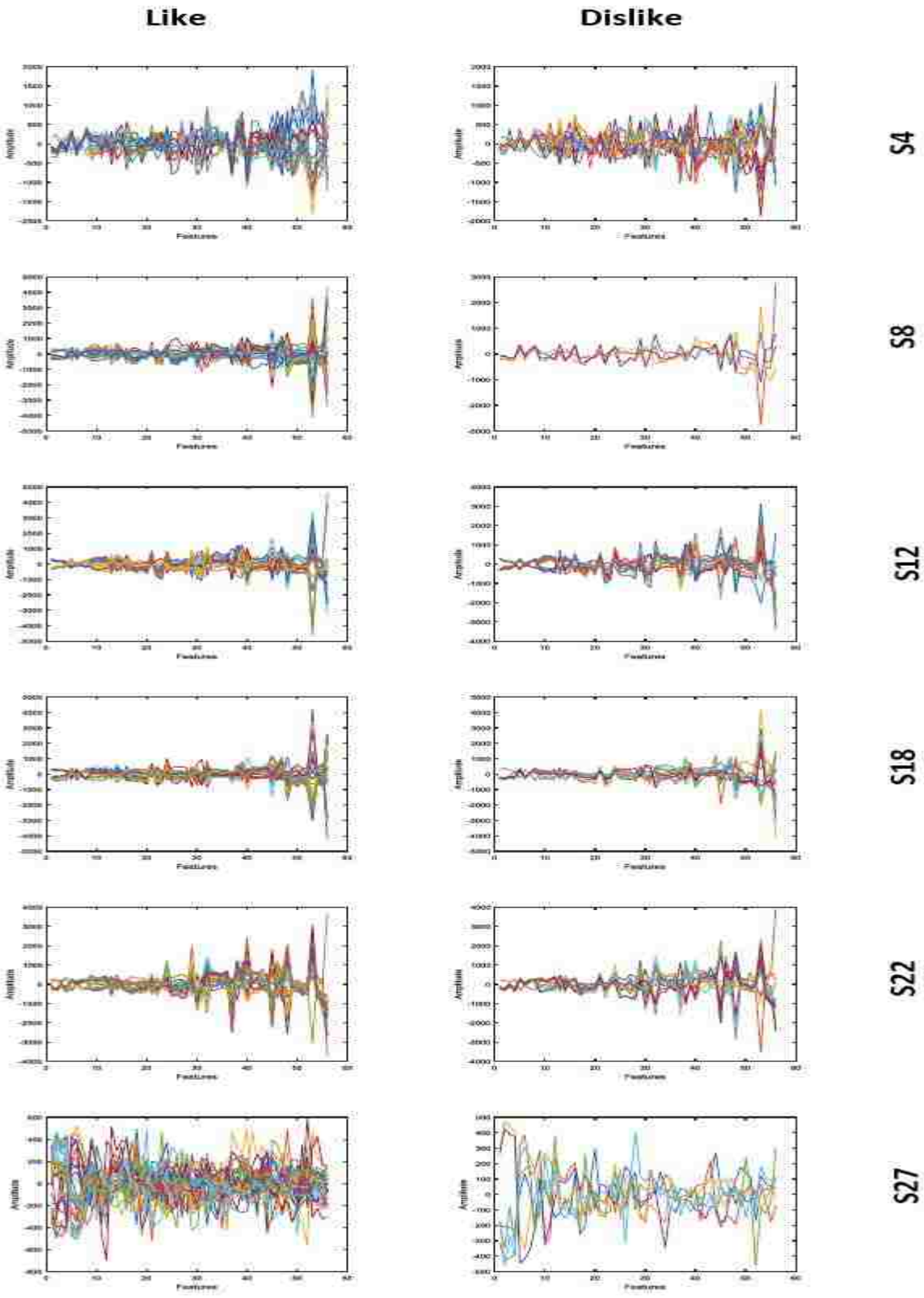


Figure 6.11: Images of sampled LBP features of liking dimension of peripheral modality data.

The Distress Phase dimension model has three classes as opposed to the two classes of the other four emotional dimensions of arousal, valence, dominance and liking.

Table 6. 5: Results of the distress phase dimension for the peripheral modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.08	0.5623	73.75	0.6683	73.67	0.7242
100	81.17	0.5960	79.45	0.6093	77.50	0.6411
150	82.03	0.4542	76.02	0.6510	76.09	0.8801

In addition, the Distress Phase emotion dimension model was developed by drawing from the inherent characteristics of all the other four emotion dimensions, thereby making the result much more meaningful and reliable for effective decision making regarding the emotional state of an individual and whether there is a need to prompt emergency assistance. For instance, the result of the arousal label alone asserting that a low/high arousal is detected is not sufficient to determine whether the individual is under threat or not and whether an emergency assistance is required. The distress phase model is thus able to extract knowledge from the assorted emotional dimensions and combine with features data to make informed decisions in an affective system.

This Distress Phase emotion model can be described as a feature fusion approach because the characteristics inherent in the four emotion dimensions are combined together and features are extracted from the pooled data to form a composite feature set which serves as the input to a pattern classifier.

As earlier done for the best results achieved by a particular feature descriptor across all dimensions for this modality, the image plots of the HOG features of the happy, distress and casualty classes of the distress phase emotion model under the peripheral physiological modality were also investigated. This is to examine the peculiarities in the features among the tripartite classes that necessitated the impressive result obtained by the HOG descriptor. As shown in Figure 6.12, unique patterns are noticed among the

classes. While a dense pattern and coarse texture are observed for the happy class, the distress and casualty classes have sparse patterns and soft textures and these observed characteristics permeate the 6 uniformly selected subjects as shown in Figure 6.12 indicating subject-independence. In addition, different amplitudes or intensities are also noticed among the three classes. It was observed that the happy class has the highest amplitude, followed by the distress class and then the casualty class thereby enhancing easy classification by the pattern recognizer. For instance, among the uniformly selected subjects, these was observed for subjects S8, S12, S18 and S22, while for S4 and S27, the amplitude values for the happy class was only able to surpass those of the distress class while those of the casualty classes are higher. This still confirm that the classes are separable and can be classified accordingly. Furthermore, this trend of decreasing succeeding amplitude/intensity values for happy, distress and casualty in that order for the HOGPEPS features of the Distress Phase model is predominant among the subjects investigated in this study.

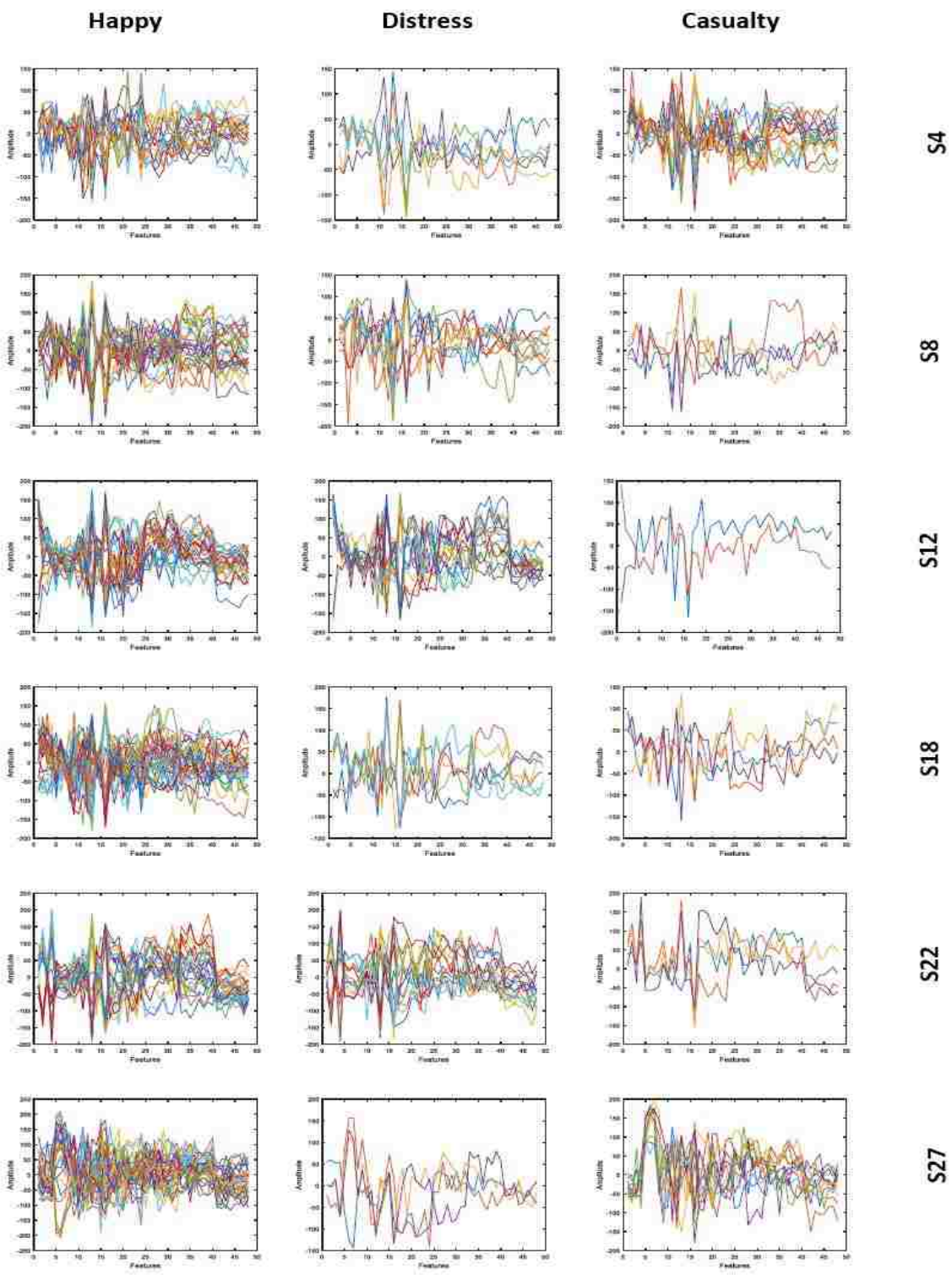


Figure 6.12: Images of sampled HOG features of distress phase model of peripheral modality data.

On the other hand, the best recognition result of 79.45% attained with the LBPPEPS features is shown in Table 6.5. This result falls below the result achieved with the HOGPEPS feature while the number of neurons used is 300 as the result also further confirms that despite its higher feature vector size as compared with the HOGPEPS, more useful and discriminatory information is contained in most of the HOGPEPS features.

The HIMPEPS features on its part, as shown in Table 6.5, yielded its best recognition result of 77.50% with 300 neurons in its hidden layer as classification is done along the Distress Phase emotion dimension label. The results achieved with the Distress Phase label with the peripheral physiological modality data are promising and are similarly trended with those of the arousal, valence, dominance and liking while the HOGPEPS features often produced the best results among the three features utilised.

In sum, for the peripheral physiological modality, the best recognition accuracy of 87.34% (MSE=0.3775) was achieved with the HOGPEPS features for the dominance class representation. The HOGPEPS features, thus contain more useful information than the other features utilised for the modality while the submissiveness/control feelings experienced and rated by the participants in the DEAP physiological data set were more significant than those of the other emotional dimensions.

The liking dimension and LBPPEPS features posted the next highest performance of 86.33% (MSE=0.0645). It is however, felt that apart from demonstrating the potential of the data transformation and feature extraction methods introduced in this study, this result is a subjective measure since the likeness scale rating is a measure of the taste of an individual about an emotion and not necessarily an emotional feeling (Koelstra et al. 2012). The combination of the HOGPEPS feature and dominance dimension or Distress Phase representation model for the peripheral physiological modality data of the DEAP data set is therefore recommended. This is because, the dominance dimension posted the best recognition result for this modality while the Distress Phase model possesses useful information and embedded knowledge that could be deployed as an emergency prompt while also posting a competitive result. The results of the EEG modality are the next to be presented and analysed.

6.2 Experimental results of EEG modality

As stated in Chapters 5, section 5.1 and 4, section 4.1.1, the EEG physiological modality data in the DEAP data set has 32 channels as against the 8 channels of the peripheral physiological modality data hence more data points from which feature sets were built. The features with which experiments were performed were named, Histogram of Oriented Gradient Physiological Signal (HOGPS), Local Binary Pattern Physiological Signal (LBPPS) and Histogram of Images Physiological Signal (HIMPS) as the results' analysis are done along the five emotion dimensions namely arousal, valence, dominance, liking and distress phase.

First, the arousal class is considered. As shown in Table 6.6, the HOGPS features on which the RBFNN pattern classifier was applied yielded a recognition accuracy of 74.77% (MSE=0.3817) with 100 neurons in the hidden layer. The number of clusters was also varied in a step of 50 between 50-250, as earlier done in the peripheral physiological modality experiments in order to determine the best result that would be posted with the varied range of number of clusters/neurons.

As observed in Table 6.6, the recognition accuracy obtained increases as more neurons are added until the best result of 88.28% (MSE = 0.1851) was achieved with 500 neurons. This result is better than all the results recorded by all the three different features along all the five emotion dimensions with the peripheral physiological modality data. This confirms the literature position that the EEG modality is capable of producing a better result than the peripheral physiological modality (Soleymani et al. 2012). This performance is also better than the results recently reported in the literature by other authors who also used the DEAP data set (Menezes et al. 2017, Nakisa et al. 2018).

Along the arousal dimension of the EEG modality, the LBPPS features recorded its best recognition result of 83.20% (MSE=0.4066) with the RBFNN using 400 neurons in its hidden layer as shown in Table 6.6. This performance matches the 83.20% also reported using the LBPPEPS features for the valence dimension as well as the HIMPEPS features of the liking dimension for the peripheral physiological modality data as shown in Tables 6.2 and 6.4 respectively. However, the result falls short of the performance of

88.28% (MSE=0.1851) posted by the HOGPS features indicating that despite its huge feature vector size of 992 elements, the LBPPS features do not contain as many useful and discriminatory data as available in the 320 elements feature vectors of the HOGPS features.

However, the HIMPS features which represent the pixel intensity values of images of the EEG modality data recorded its best recognition accuracy of 93.36% (MSE=0.2974) as shown in Table 6.6 while the RBFNN classifier utilised 500 neurons in its hidden layer. This result is very remarkable and surpasses all the results reported for the peripheral physiological modality and that of the HOGPS and LBPPS features of the EEG modality.

Table 6. 6: Results of the arousal dimension for the EEG modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	74.77	0.3817	72.97	0.3212	87.50	0.3590
100	81.56	0.4234	79.22	0.3567	90.70	0.2139
150	85.08	0.2790	81.41	0.4462	92.89	0.1913
200	87.66	0.0975	83.20	0.4066	93.13	0.4188
250	88.28	0.1851	82.73	0.0526	93.36	0.2974

This is an indication that with the proposed data preprocessing, inverse Fisher transformation and mapping to image technique, the histogram of images is a very potent feature for human emotion recognition as it has also yielded good performances in other field of studies (Mohamad, Manaf and Chuprat 2011; Iman et al. 2017; Thamizhvani et al. 2018). The 93.36% (MSE=0.2974) recognition result is also better than the results obtained in different recent research studies that have utilised the DEAP data set (Li et al. 2017; Yin et al. 2017; Menezes et al. 2017, Nakisa et al. 2018).

With this result obtained, as also done with the HOGPEPS and LBPPEPS result of the dominance, distress phase and liking dimensions for the peripheral physiological modality, the images of Histogram features of the low and high arousal dimensions was analysed with a view to determining some of the inherent characteristics in the features and classes that yielded this result. As shown in Figure 6.13, for the 6 subjects uniformly

selected, the plots of the Histogram features for the low and high arousal dimensions revealed different feature values or amplitude or intensity. It was observed that for all the uniformly selected subjects the high arousal dimension has a higher amplitude than the low arousal dimension for each of the subjects thereby enhancing easy classification by the pattern recognizer. For instance, amplitudes values of 246, 389, 232, 252, 234 and 460 were recorded for the high arousal class for subjects S4, S8, S12, S18, S22 and S27 respectively while the corresponding amplitude values for the low arousal class are 226, 296, 142, 243, 217 and 313 respectively. This trend of high amplitude for high arousal dimension and low amplitude for low arousal dimension is predominant among the subjects and also agrees with the findings reported for the HOGPEPS and LBPPEPS features of the dominance and liking dimensions of the peripheral modality data as shown in Figures 6.10 and 6.11.

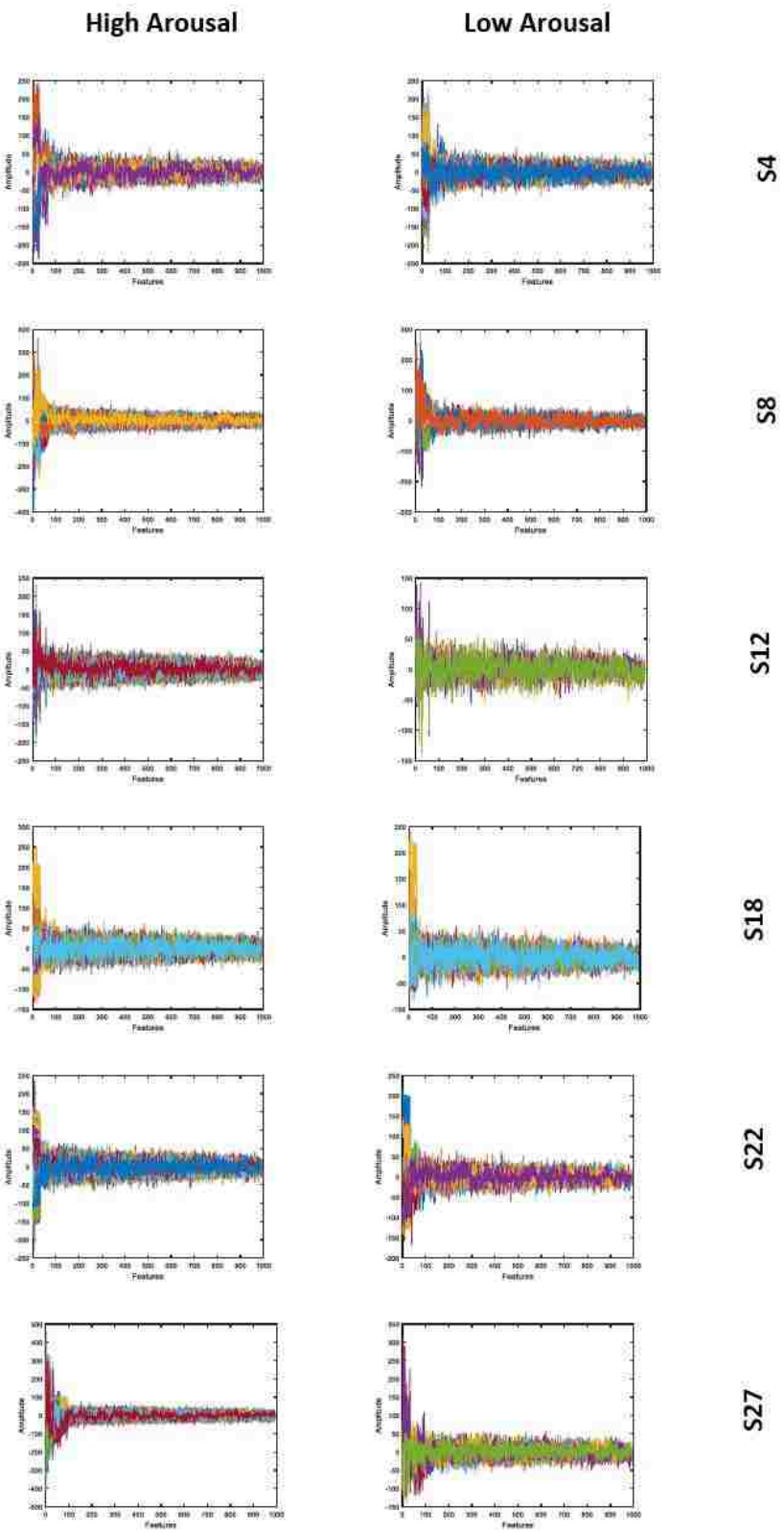


Figure 6.13: Images of sampled Histogram features of arousal dimension of EEG modality data.

The results of the valence dimension of the EEG modality with the HOGPS, LBPPS and HIMPS features are hereby presented. As shown in Table 6.7, the RBFNN classifier with the HOGPS features of the EEG modality posted its best recognition result of 88.05% (MSE=0.3022) along the valence dimension while utilizing 500 neurons in its hidden layer. This result is marginally lower than the 88.28% (MSE=0.1851) recorded for this modality by the HOGPS features with classification done along the arousal dimension but better than the LBPPS features' performance of the arousal dimension as well as all the results reported under the peripheral physiological modality.

On the other hand, the RBFNN classifier applied to the LBPPS features of the EEG modality physiological data recorded a recognition accuracy of 75.31% (MSE=0.4917) with 100 neurons in the hidden layer. As more neurons are added by varying the number of clusters up to 250 maximum as shown in Table 6.7, the recognition accuracy attained by the RBFNN classifier increases till the best recognition results of 85.31% (MSE=0.2523) was recorded using 500 neurons. Following a similar trend as noticed in the EEG results earlier reported, the 85.31% (MSE=0.2523) result is lower than the 88.05% (MSE=0.3022) achieved with the HOGPS features using the same number of clusters/neurons. This further confirms that the HOGPS features, though consisting of fewer elements as compared to the LBPPS features are more discriminating and have useful information as it readily outperforms the LBPPS features.

The last features that were experimented on with the EEG modality along the valence dimension is the HIMPS features. A recognition accuracy of 92.81% (MSE=0.3502) as shown in Table 6.7 was achieved with 400 neurons in the hidden layers of the RBFNN classifier. An additional 100 neurons were only able to marginally improve the result by 0.08% to 92.89% (MSE=0.0878). This was felt not significant enough, as the efficiency of the network is more relevant than the marginal accuracy obtained. Notwithstanding, the highest result obtained with the HIMPS features is better than the result of the HOGPS and LBPPS features of both the arousal and valence dimensions of the EEG modality. In addition, it is marginally lower by 0.47% to the highest result of 93.36% (MSE=0.2974) achieved with the HIMPS feature along the arousal class. Thus,

for the EEG modality, the HIMPS features yielded the best recognition result for the valence dimension as also noticed with the arousal dimension.

The dominance dimension is the next to be considered as the RBFNN pattern classifier was applied separately on the HOGPS, LBPPS and HIMPS features of the EEG modality physiological data. The HOGPS features were the first to be experimented on by the RBFNN pattern classifier. With a feature vector size of 320 and 1280 training samples, the best recognition accuracy of 89.53% (MSE=0.2345) was obtained along the dominance dimension with the EEG modality physiological data as shown in Table 6.8. This was achieved with 500 neurons as a decline was noticed in the RBFNN pattern classifier's performance with subsequent addition of neurons.

However, this result is better than those obtained by the HOGPS and LBPPS features for the arousal and valence dimensions of the EEG modality already considered. It is also observed that the results posted by any of the HOG, LBP and HIM features extracted from the EEG modality is often better than the corresponding results obtained with the peripheral physiological modality as the ability of the features extracted from the EEG modality to outperform the features extracted from the peripheral physiological modality is further proven and aligns with the literature (Soleymani et al. 2012).

Table 6. 7: Results of the valence dimension for the EEG modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	72.58	0.6499	75.31	0.4917	86.64	0.2818
100	81.09	0.4594	81.33	0.2990	91.25	0.1138
150	84.92	0.3973	81.56	0.3877	92.03	0.0685
200	86.48	0.1809	84.84	0.2622	92.81	0.3502
250	88.05	0.3022	85.31	0.2523	92.89	0.0878

With the LBPPS features of the EEG modality classified along the dominance dimension, a recognition accuracy of 87.89% (MSE=0.0430) shown in Table 6.8 and achieved with 400 neurons of the RBFNN pattern classifier performs less than the corresponding HOGPS features. Though, this result is better than the results of the LBPPS features for the arousal and valence classes, it still confirms that the EEG modality of the DEAP physiological data is more responsive to the HOG descriptor than the LBP descriptor. It also shows that the edges and corners of the images from which these features are extracted are better captured than the local patterns of the images thereby resulting in better HOGPS features' performances than the LBPPS features.

While still classifying along the dominance scale, the HIMPS features of the EEG modality when experimented with the RBFNN pattern classifier achieved its best recognition accuracy of 93.36% (MSE=0.2623) using 300 neurons in its hidden layer as indicated in Table 6.8. This result is better than the 92.89% (MSE=0.0878) recorded by the same features when classified along the valence dimension while it also matches the 93.36% (MSE=0.2974) obtained with the arousal dimension. Though the arousal dimension result was achieved while using 500 neurons, the result of the HIMS features with classification done along dominance dimension is preferred as it was achieved with just 300 neurons.

Table 6. 8: Results of the dominance dimension for the EEG modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.72	0.6589	76.80	0.8443	86.64	0.6510
100	82.34	0.1717	80.78	0.0454	91.88	0.2910
150	85.86	0.3414	85.47	0.0849	93.36	0.2623
200	88.05	0.2387	87.89	0.0430	93.20	0.3694
250	89.53	0.2345	85.31	0.1099	93.20	0.3068

This is also because; the cost, efficiency and simplicity of the RBF neural network in a human recognition system, especially for real-time deployment should often be considered above the accuracy. Again, the result of the HIMPS features is the best among

the three features utilised for the dominance class and it aligns with the trend earlier reported with the EEG modality physiological signals where the HIMPS, HOGPS and LBPPS features are listed in order of their performances attained.

As done for the arousal dimension in Figure 6.13, the high performance of the Histogram features of the EEG modality for the dominance dimension was also investigated. The Histogram features of the 6 subjects uniformly selected for the low and high dominance dimension were examined as shown in Figure 6.14 to determine if there are similar properties inherent in the features and classes among these same uniformly selected subjects, which could indicate a subject-independent result as also shown in the arousal dimension of the EEG modality.

The amplitude of the high dominance dimension for subject S12 is 119 while the low dominance dimension is 96. Similarly, subjects S18, S22 and S27 have amplitudes of 87, 84 and 107 respectively, for the high dominance dimension with 77, 47 and 88 respectively for the low dominance dimension. With marginal values, only subjects S8 and S4 did not align with this observation. However, the observed trend is predominant across the subjects and the same as also noticed in the Histogram features of arousal dimension signals of the EEG modality as the large amplitudes are associated with high dominance dimension and low amplitude values are attached to the low dominance dimension. The few signals that do not follow this trend may be due to some emotions and emotional responses of some subjects which may cover any amplitude range including high, medium and low (Lech et al. 2018) but these subjects are few out of the 32 subjects considered in the DEAP data set that is utilised in this study.

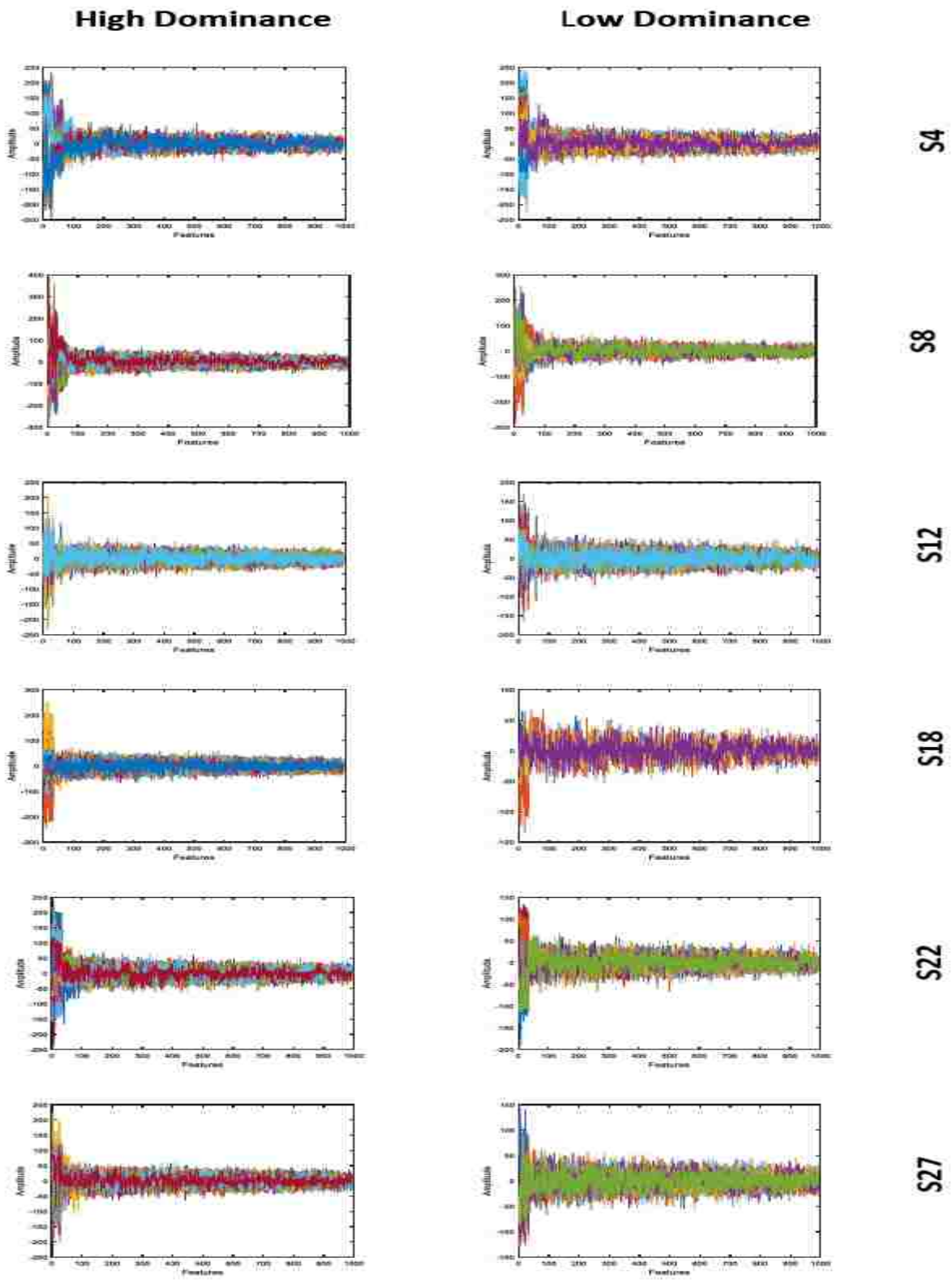


Figure 6.14: Images of sampled Histogram features of dominance dimension of EEG modality data.

In another vein, the liking dimension of emotion is utilised with the HOGPS, LBPPS and HIMPS features extracted from the EEG modality data set and RBFNN pattern classifier applied for classification. According to the various results shown in Table 6.9, the HOGPS recorded its highest recognition accuracy of 88.83% (MSE=0.1101) with 400 neurons in the hidden layer of the RBFNN classifier. This result is better than all the results posted by the LBPPS for the arousal, valence and dominance dimensions as well as marginally better than the HOGPS results of the arousal and valence dimensions.

However, the best recognition result of 85.23% (MSE=0.1002) was achieved with the LBPPS features using 400 neurons and with this performance, the HOGPS features are more responsive to and yielded better performance with the RBFNN algorithm than the LBPPS features with the DEAP data set and data transformation method proposed.

As the experiment was extended to the HIMPS features and classification done along the liking dimension, a recognition accuracy of 86.17% (MSE=0.4480) was recorded with 100 neurons. The recognition performance increases to its best result of 93.13% (MSE=0.1136) which was achieved using 300 neurons after which performance declines with further neuron addition. Out of the HIMPS features' results of the EEG modality data for the four emotional dimensions considered so far, this HIMPS features' result is only marginally better than the 92.89% (MSE=0.0878) obtained with the valence dimension. The fact is also reiterated that the results obtained along the liking dimensions are not necessarily emotional feelings, but subjective ratings of participants indicating their tastes of like/dislike of the various emotions/elicitation materials. Thus, the liking dimension ratings aside from assisting in the tagging of elicitation materials for human emotion recognition can also be used for predictive analytics, including addiction management, suicidal thoughts, crime control as well as marketing, advertisement and sales of materials with emotional contents.

To conclude the experimental results of the HOGPS, LBPPS and HIMPS features with the EEG modality physiological data, the results obtained by the designed RBFNN with classification done along the Distress Phase emotion dimension scheme are hereby presented.

Table 6. 9: Results of the liking dimension for the EEG modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.17	0.2599	77.03	0.4403	86.17	0.4480
100	83.28	0.0190	84.22	0.1092	92.19	0.2335
150	86.64	0.0825	84.30	0.3361	93.13	0.1136
200	88.83	0.1101	85.23	0.1002	92.50	0.1253
250	87.89	0.0855	84.84	0.2150	91.95	0.1401

The Distress Phase emotion scheme was modelled as earlier mentioned by utilising the characteristics extracted from the four emotional dimensions, which are arousal, valence, dominance and liking while mapping along three classes, which are happy, distress and casualty. The results obtained using the various features are shown in Table 6.10.

The HOGPS features were the first to be experimented on with the EEG modality, as the RBFNN classifier was applied to classify the features along the Distress Phase model dimension. The recognition accuracy of 76.09% (MSE=0.6922) which was achieved using 50 clusters which represents 150 neurons here since three classes are involved in this dimension is shown in Table 6.10. With an increase in the number of neurons up to a maximum of 450 neurons, the best recognition accuracy of 84.14% (MSE=0.6721) was achieved using 300 neurons. This is obviously the lowest results recorded by the HOGPS with the EEG modality. However, as earlier mentioned, the performance of a pattern classification system is affected with higher number of classes, but this performance is still better than most of the results posted under the peripheral physiological modality across the three feature sets and emotion dimensions. This result is also more meaningful as the characteristics of all the other four emotion dimensions have been incorporated in an effective decision making in a distress phase situation, which involve an emergency assistance to be sought and responded to.

The LBPPS features also yielded its best recognition result of 81.25% (MSE=0.6725) as shown in Table 6.10 under the Distress Phase dimension with the

RBFNN using 300 neurons in the hidden layer. This result also aligns with the trend earlier observed in our various experiments that the LBPPS results often fall below those of the HOGPS indicating its less responsiveness to the RBFNN algorithm.

Table 6. 10: Results of the distress phase dimension for the EEG modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.09	0.6922	75.86	0.4546	89.77	0.2978
100	84.14	0.6721	81.25	0.4583	91.41	0.5248
150	83.59	0.8660	79.30	0.4688	88.05	0.4408

The HIMPS features of the EEG modality data on the other hand recorded its best recognition result of 91.41% (MSE=0.5248) as shown in Table 6.10, using 300 neurons in the hidden layer of the RBFNN as its performance declines when more neurons are added. This result outperforms those of the HOGPS and LBPPS features under the Distress Phase dimension of the EEG modality, but falls below the results posted by the same features across each of the other four emotion dimensions.

The images of the Histogram features of the three classes catalogued under the Distress Phase model, which are happy, distress and casualty for sampled uniformly selected subjects, are presented in Figure 6.15 to examine some inherent characteristics in these features among the subjects and the classes necessitating the results obtained.

As observed with the HOG features of the peripheral physiological modality of the Distress Phase model, it was also observed in Figure 6.15 that different amplitudes or intensities are also noticed among the three classes. The happy class also has the highest amplitude, followed by the distress class and then the casualty class, thus enhancing easy classification by the pattern recognizer. Among the uniformly selected subjects, these were observed for subjects S4, S12 and S22, while for S8, S18 and S27, the amplitude values of the distress class is the lowest among the three. This confirms that the classes are separable and can be classified accordingly.

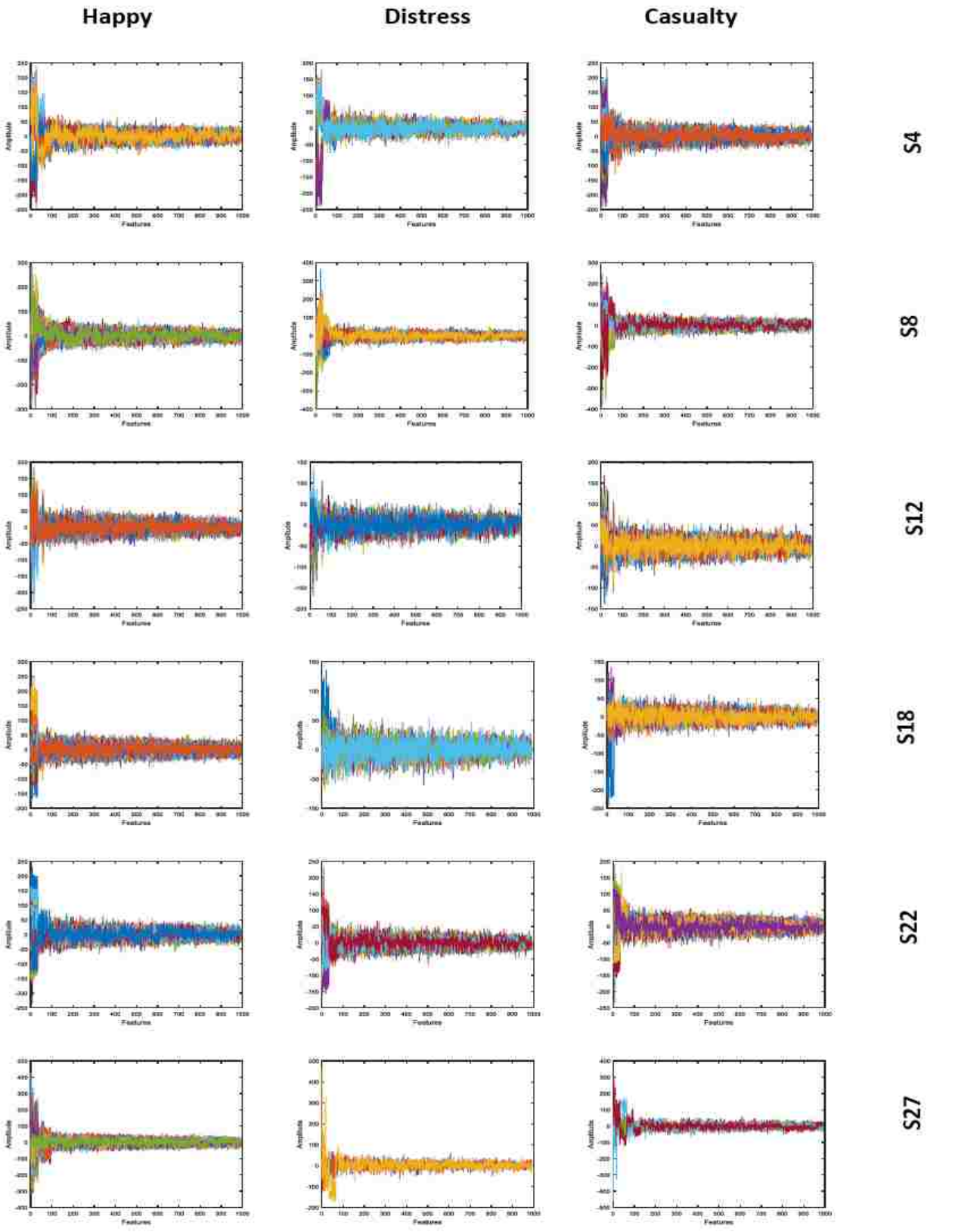


Figure 6.15: Images of sampled Histogram features of distress phase model of EEG modality data.

In summary, for all the EEG modality experiments conducted using the three features across the five different emotion dimensions, the highest recognition accuracy of 93.36% was recorded by the HIMPS features for both the arousal and dominance dimensions. However, the results of the dominance dimension are preferred having been attained with a lower number of neurons of 300 as against the 500 neurons of the arousal dimension. The HOGPS and LBPPS features also posted their best results of 89.53% (MSE=0.2345) and 87.89% (MSE=0.0430) respectively with the dominance dimension. Therefore, because they yielded the highest recognition results, for the EEG modality, the HIMPS features with classification done along the dominance and the distress phase model dimensions are recommended followed by the HOGPS and the LBPPS features in this order with respect to the high performances obtained.

6.3 Experimental results of fused modality

The fused modality is the last experiment conducted using the HOGHES, LBPHES and the HIMHES features extracted by the HOG, LBP and Histogram of Images descriptors when applied to the fused modality physiological data of the DEAP dataset. It has been shown in the literature that modality fusion is capable of yielding an improved classification result in a human emotion recognition system (Koelstra et al. 2012; Soleymani et al. 2012). This is because complementary characteristics in the different modalities are exploited with a view to detecting a unique pattern from which features are extracted with the primary aim of obtaining a higher performance.

The RBFNN pattern classifier using the HOGHES features for classification along the arousal class achieved a recognition accuracy of 75.47% (MSE=0.7774) while using 100 neurons in its hidden layer as shown in Table 6.11. With more neurons added, the recognition accuracy increases until the highest result of 87.11% (MSE=0.2134) was attained with 400 neurons in the hidden layer. The corresponding results of 85.16% (MSE =0.0398) in Table 6.1 and 88.28% (MSE=0.1851) in Table 6.6 achieved with the HOG features of the peripheral physiological and EEG modality data respectively under the arousal dimension are compared. This result (87.11%) is only better than the 85.16%

(MSE =0.0398) accuracy of the peripheral physiological modality and falls short of the 88.28% accuracy (MSE=0.1851) of the EEG modality.

This result therefore implies that the fused modality did not always lead to a higher performance. This might be traced to the somewhat lower performance of the peripheral physiological data which negatively affect the 87.11% (MSE=0.2134) result that was obtained. However, this trend was also shown in the liking dimension F1-score result reported by Koelstra et al. (2012). According to the author, a fusion of EEG, peripheral and Multimedia Content Analysis (MCA) features yielded an F1-score of 61.8%, which is more than the 50.2% and 53.8% attained by the EEG and peripheral modalities respectively but falls short of the 63.4% obtained with the MCA features. In addition, the fused peripheral and MCA modalities only yielded an F1-score of 62.2%, which is also less than the 63.4% obtained by only the MCA modality.

Furthermore, in the same study, with an F1-score of 58.3%, 53.3% and 61.8% recorded in the EEG, peripheral and MCA modalities respectively obtained with classification done along arousal dimension, a marginally low 61.6% F1-score was obtained when the three modalities were fused. These results are lower than the best single modality's results recorded as noted under the arousal and liking dimensions, thus agreeing with the findings in this fused modality experiment using the HOGHES features as well as literature position that classification results might not always necessarily improve with fused modalities. Notwithstanding this fact, as also shown in the same study (Koelstra et al. 2012), the F1-score posted by the fused best two modalities for the arousal and valence dimensions respectively, are better than the result of any of the single modality of these dimensions.

The LBPHES is the next features of the fused modality data that was experimented on by applying the RBFNN classifier for classification along the arousal dimension scheme. The best recognition accuracy of 81.80% (MSE=0.1968) obtained using 300 neurons in the hidden layer is shown in Table 6.11. This result is lesser than the 83.20% (MSE=0.4066) and 84.92% (MSE=0.4973) achieved respectively, with the LBP features extracted from the EEG and peripheral single modalities as shown in Tables 6.2 and 6.7. Hence the fused modality result achieved does not indicate its superiority to those of the

single modalities and it is also more computationally expensive as it contains a huge feature vector size of 1440 elements.

Table 6. 11: Results of the arousal dimension for the fused modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.47	0.7774	75.00	0.2645	78.52	0.2749
100	82.50	0.2621	76.80	0.2713	82.03	0.4684
150	85.47	0.3005	81.80	0.1968	86.25	0.4769
200	87.11	0.2134	80.16	0.4179	88.13	0.3302
250	87.03	0.0572			88.36	0.3363

The HIMHES result of the hand recorded its best recognition result of 88.36% (MSE=0.3363) with 500 neurons and aligns with the trend earlier noticed in the HOGHES and LBPHES features wherein the fused modality data do not yield a better result than all of the results of the single modalities.

The valence dimension is the next scheme that was considered for the fused modality data. The RBFNN classifier was applied to the HOGHES features and 90.08% (MSE=0.4309) was obtained as the best result as shown in Table 6.12. The number of neurons utilised in the hidden layer to achieve this best result was 500. This result is better than the 88.05% (MSE=0.3022) and 85.94% (MSE=0.4246) obtained by the corresponding HOG features of the EEG and peripheral physiological modality data respectively as shown in Tables 6.7 and 6.2. This clearly shows that the HOGHES features with classification done along the valence dimension confirms and aligns with the fact in the literature that modality fusion can indeed improve performance of a pattern classifier as the feature set is enriched with more useful and discriminatory information.

The image plots of the HOG features for the high and low valence classes of some uniformly sampled subjects are presented in Figure 6.16 to examine the characteristics among these subjects and classes that resulted in this impressive performance of the HOGHES features. A 4-window signal shape representing HOG features of the fused

modality data is common among the uniformly selected subjects, suggesting a subject-independent behaviour as the amplitude or feature values between the high and low valence classes varies. The high amplitude figures are mostly associated with high/positive valence class while corresponding lower amplitude are predominantly associated with the low/negative valence class thus enhancing the classification task. For instance, subjects S18, S22 and S27 have 221, 218 and 225 respectively as amplitude values for the high valence class while 213, 212 and 217 are amplitude values for the low valence class thus agreeing with the earlier findings.

However, a best recognition result of 80.39% (MSE=0.2964) was achieved by the RBFNN with the LBPHES features with 400 neurons as shown in Table 6.12. This performance is lower than the 85.31% (MSE=0.2523) and 83.20% (MSE=0.3057) results respectively obtained with the corresponding LBP features of the EEG and peripheral modality data as shown in Tables 6.7 and 6.2. This similar trend is also recorded by the

Table 6. 12: Results of the valence dimension for the fused modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.47	0.6702	75.16	0.7735	80.31	0.4552
100	80.78	0.3733	78.91	0.7226	84.06	0.4287
150	86.72	0.3506	79.77	0.3857	85.31	0.4795
200	88.28	0.3864	80.39	0.2964	88.28	0.4678
250	90.08	0.4309	79.84	0.3807	88.36	0.3970

HIMHES features where its best recognition accuracy of 88.36% (MSE=0.3970) shown in Table 6.12 is only able to surpass the 82.66% (MSE=0.4792) of the peripheral modality data, but falls short of the 92.89% (MSE=0.0878) attained by the EEG modality data. Thus, it is only with the HOGHES features of the valence dimension that the modality fusion approach improved the classification result obtained by the RBFNN pattern classifier.

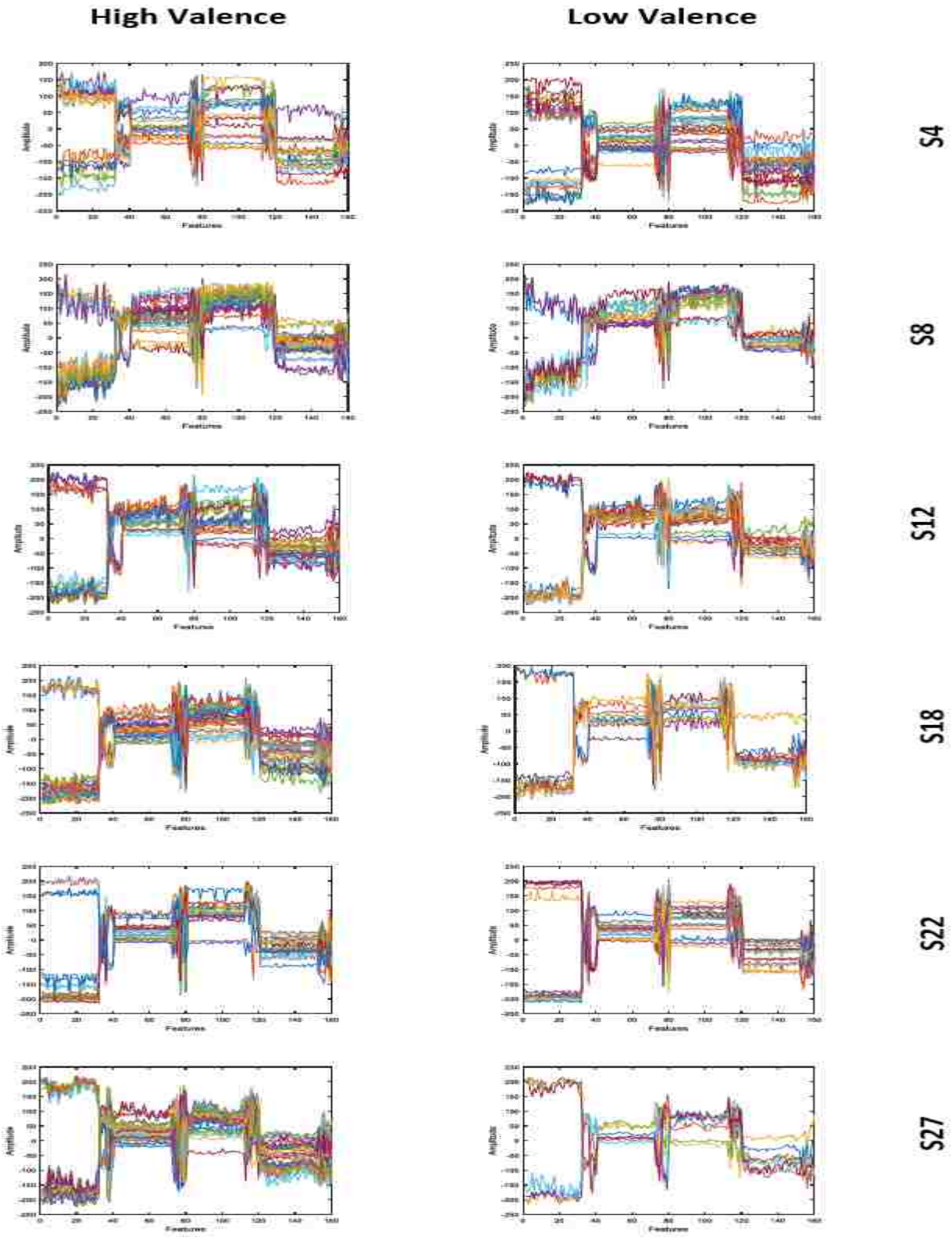


Figure 6.16: Images of sampled HOG features of valence dimension of fused modality data.

As observed in Table 6.13, the dominance emotion dimension, using the HOGHES features of the fused modality data by the RBFNN yielded the best recognition accuracy of 88.36% (MSE=0.1396) using 400 neurons in the hidden layer. The 89.53% accuracy (MSE=0.2345) in Table 6.8 and 87.34% accuracy (MSE=0.3775) in Table 6.3 as best results obtained with the EEG and peripheral physiological modality respectively, with 500 neurons each indicates that the fused modality features performance cannot be a replacement to the single EEG modality. This trend is also demonstrated by the LBPHES features' performance wherein the best recognition accuracy of 81.33% (MSE=0.2092) shown in Table 6.13 achieved with 200 neurons is below the 87.89% (MSE=0.0430) and 84.92% (MSE=0.1271) achieved with the corresponding LBP features of the EEG and peripheral modalities. In the same vein, the HIMHES features with 89.97% best recognition accuracy and MSE of 0.1108 as shown in Table 6.13, was only able to surpass the 84.14% recognition accuracy and MSE of 0.1700 as shown in Table 6.3 of the peripheral modality. However, it falls short of the 93.36% (MSE=0.2623) of the EEG modality as shown in Table 6.8 and; thus cannot replace this single modality performance beside having more feature vector elements capable of increasing computation costs.

In summary, for the dominance dimension, the best recognition performance of 89.97% (MSE=0.1108) recorded with the fused modality data was achieved with the HIMHES features and as earlier explained, this result is not better than its corresponding EEG single modality result and cannot be recommended as a replacement.

Table 6. 13: Results of the dominance dimension for the fused modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.09	1.1032	76.95	0.2945	82.19	0.0179
100	79.69	1.1365	81.33	0.2092	86.17	0.0648
150	85.39	0.1590	80.86	0.3424	89.77	0.1183
200	88.36	0.1396	80.78	0.3559	87.97	0.1108
250	86.88	0.0843	78.36	0.4133		

The fused modality data from which the HOGHES, LBPHES and HIMHES features were extracted and classification done along the liking dimension is the next scheme that is considered and the various results are shown in Table 6.14. The HOGHES features has its best recognition result of 88.59% (MSE=0.3303) which was attained by RBFNN using 400 neurons in its hidden layer. This performance was compared with the corresponding HOG features results of the EEG and peripheral physiological single modalities that recorded 88.83% (MSE=0.1101) and 83.52% (MSE=0.1647) respectively as shown in Tables 6.9 and 6.4. However, only the peripheral physiological modality was surpassed by the fused modality result, while it falls marginally below the EEG modality result by 0.24%, hence not a replacement to the single modality result.

Also, with the LBPHES features, a similar trend is noticed where the best recognition result of 81.41% (MSE=0.0683) as shown in Table 6.14, with the fused modality data cannot match the 85.23% (MSE=0.1002) and 86.33% accuracy (MSE=0.0645) of the EEG and peripheral physiological single modalities shown in Tables 6.9 and 6.4. In addition, the HIMHES features has its best result of 89.38% (MSE=0.0826). This is below the highest performance of 93.13% (MSE=0.1136) of the EEG modality as shown in Table 6.9 and, only better than the 83.20% of the peripheral modality as shown in Table 6.4, which aligns with the earlier position that the fused modality scheme cannot always replace the single modalities scheme in terms of classification performance.

Table 6. 14: Results of the liking dimension for the fused modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.80	0.2948	79.77	0.3613	82.58	0.8574
100	83.36	0.3549	79.60	0.0486	84.61	0.2087
150	86.72	0.3729	81.17	0.0252	87.03	0.2613
200	88.59	0.3303	81.41	0.0683	89.22	0.0795
250	86.64	0.2224	80.31	1.0126	89.38	0.0826

The developed and proposed Distress Phase emotion model reported in this dissertation was also used for experimentations with the fused modality data. The four emotion dimensions, including arousal, valance, dominance and liking were utilised in developing the model. The HOGHES representing the HOG features extracted from the fused modality data with the RBFNN classifier applied while classification was done along the Distress Phase dimension consisting the tripartite happy, distress and casualty phases produced 81.41% (MSE=0.6991) as its best recognition accuracy as observed in Table 6.15. This was achieved with 450 neurons in the hidden layer. As compared with the results of the corresponding HOG features of the single modalities of EEG and peripheral physiological signals shown in Tables 6.10 and 6.5 respectively, this performance is lower than both the 82.03% (MSE=0.4542) of the peripheral modality and the 84.14% (MSE=0.6721) of the EEG modality. With this result, the fused modality performance cannot replace those of the single modalities.

In addition, the RBFNN classifier when applied to the LBPHES features with 300 neurons in the hidden layer recorded its best result of 78.98% (MSE=0.8827) as shown in Table 6.15 which is below the 81.41% (MSE=0.6991) result achieved with the HOGHES features of the same Distress Phase emotion dimension. This result followed the already established trend in the experimental results reported in this dissertation wherein the performance of the LBP features falls below those of the corresponding HOG features, thus indicating the stronger discriminatory capability of the HOG descriptor with respect to the method employed. Also, this 78.98% (MSE=0.8827) result does not match the 81.25% (MSE=0.4583) and 79.45% (MSE=0.6093) classification results of the EEG and peripheral single modalities respectively, which was achieved with the same number of hidden neurons as the results of the single modalities shown in Tables 6.10 and 6.5 are therefore recommended over the fused modality's result.

Table 6. 15: Results of the distress phase dimension for the fused modality.

Features / Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.78	0.5872	76.17	0.9176	80.00	0.4072
100	81.25	0.6788	78.98	0.8827	86.02	0.3796
150	81.41	0.6991	74.69	0.8223	84.22	0.6391

The last experiment to be conducted in this study as well as with the fused modality data utilised the HIMHES features on which the RBFNN classifier was applied and classification was done along the happy, distress and casualty phases as modeled in the Distress Phase dimension. The best recognition accuracy of 86.02% (MSE=0.3796) was recorded with 300 neurons utilised in the hidden layer as indicated in Table 6.15. This result is only better than the corresponding HIMS features results of 77.50% (MSE=0.6411) of the peripheral modality in Table 6.5 but falls below the 91.41% (MSE=0.5248) of the EEG modality in Table 6.10. This further affirms the literature position and the data processing, feature extraction, emotion modelling as well as classification methods employed, that the fused modality does not repeatedly outperform the single modality results.

It was also noticed that the best human emotion recognition result of 86.02% (MSE=0.3796) with the Distress Phase dimension and fused modality data was attained by the HIMHES features. The overall best result of 90.08% (MSE=0.4309) for the fused modality data in this study was achieved with the valence dimension and the HOGHES features as shown in Table 6.12 as this combination is recommended for apparent inclusion in an emotion recognition system for the fused modality data. However, the combination lacks the knowledge representation and multidimensional characteristics embedded in the Distress Phase dimension which becomes handy and relevant for decision making in a distress phase situation that requires prompting of emergency services. Therefore, the HIMHES features with the best classification result of 86.02%

(MSE=0.3796) and Distress Phase dimension are recommended for the distress phase dimension using fused modality data.

The trend observed with the HIM and HOG features of the EEG and peripheral physiological modalities respectively, for the Distress Phase model was also repeated for the HIM features of the fused modality using the Distress Phase model data. As shown in Figure 6.17, different amplitudes or intensities are also noticed among the three classes. The happy class also has the highest amplitude, followed by the distress class and then the casualty class, thus enhancing easy classification by the pattern recognizer among the three. This confirms that the classes are separable and can be classified accordingly.

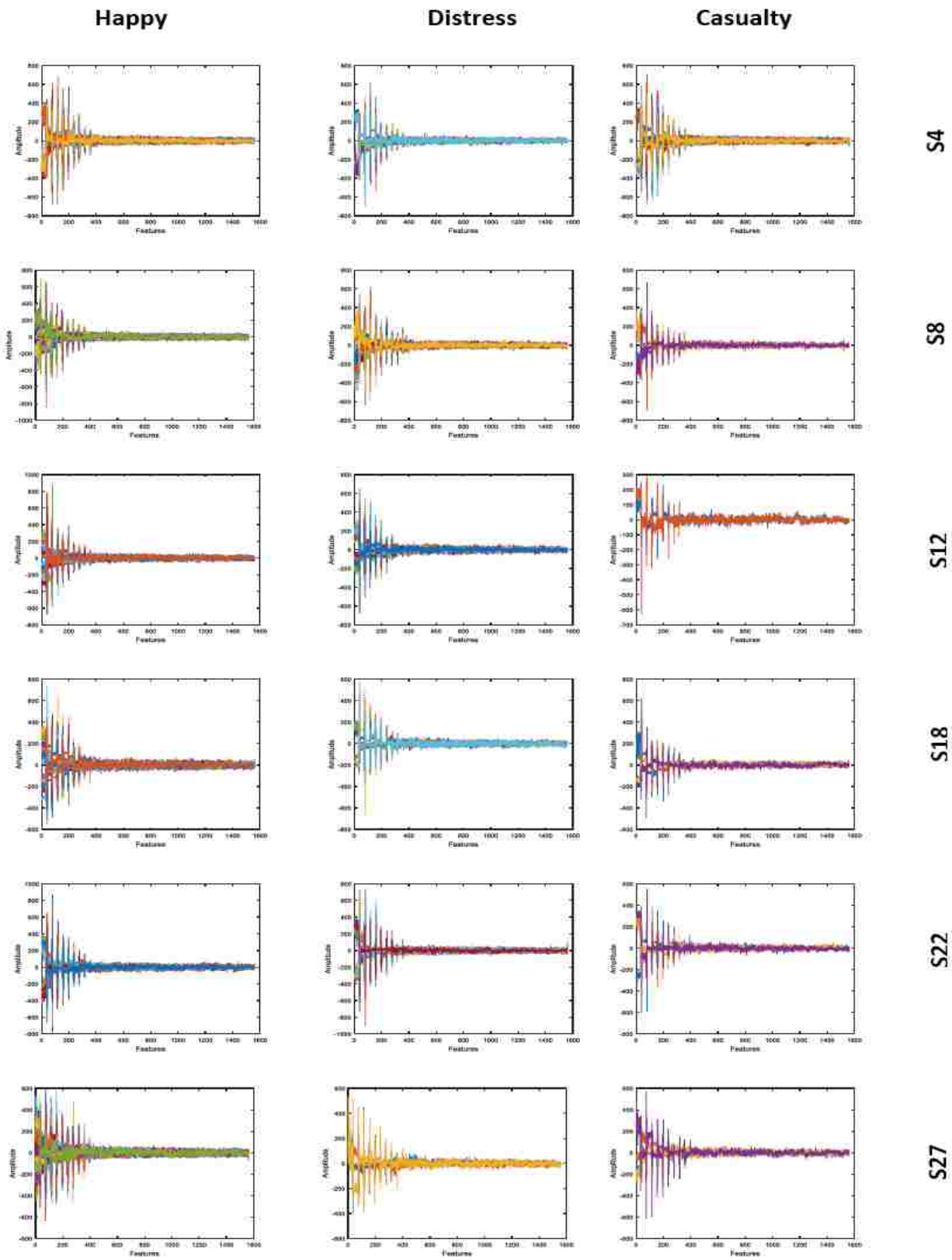


Figure 6.17: Sample histogram features of distress phase dimension using fused modality.

The summary of the best results obtained under each modality and the corresponding features is shown in Table 6.16. In the Table, the arousal class recorded 85.16% (MSE=0.0398) as its best recognition accuracy for the peripheral physiological modality using the HOG features while the EEG and the fused modalities recorded their best recognition results of 93.36% (MSE=0.2974) and 88.36% (MSE=0.3363) respectively using the HIM features. The peripheral physiological data has fewer data channels, hence fewer feature vector elements as compared with the other modalities. Thus, based on the results obtained, the peripheral physiological data are easily detected by texture, local appearance and shape of their mapped images as provided by the LBP and HOG feature descriptors. However, for the arousal class and across all the modalities considered, the HIM features and EEG modality produced the best result of 93.36% (MSE=0.2974) and this combination is hereby recommended.

As shown in Table 6.16 for the valence class and across the three modalities, the best recognition results of 92.89% (MSE=0.0878) were achieved by the EEG modality using the HIM features. The fused modality's result of 90.08% (MSE=0.4309) and 85.94% (MSE=0.4246) of the peripheral modality were achieved with the HOG features thus confirming only their unique shape and local appearances. Also, because of its superior performance, a combination of the EEG modality and the HIM features is recommended for the valence class in human emotion recognition using the DEAP dataset.

The dominance class recorded as its best results, the recognition accuracies of 89.77% (MSE=0.1183) and 87.34% (MSE=0.3775) for the fused and peripheral modalities respectively with the HIM and HOG features. Following the trend noticed in the arousal and valence classes, the best result of 93.36% (MSE=0.2623) was recorded by the EEG modality using the HIM features and the combination is also recommended.

The summary of the best results obtained under each modality and the corresponding features while using the method proposed in this research work is presented in Table 6.16 while the results obtained by other authors who have used the same DEAP dataset that was utilised for experimentations in this study including the various methods applied by them are shown in Table 6.17 to enable an easy comparison of results obtained.

In summary, as contained in Table 6.16 and graphically represented in Figure 6.18, even for the liking class and distress phase emotion model, the best results of 93.13 (MSE=0.8827) and 91.41% (MSE=0.5248) respectively were recorded by the EEG modality using the HIM features. The overall best recognition result is 93.36% which was obtained by arousal and dominance classes, but the dominance class has a lower mean square error of 0.2623 as well as the lower number of neurons utilised in the hidden layer.

Table 6. 16: Summary of experimental results across dimensions, modalities and features.

	Modalities					
	Peripheral		EEG		Fused	
Emotional Dimension	ACC (%)	Feature	ACC (%)	Feature	ACC (%)	Feature
Arousal	85.16	HOG	93.36	HIM	88.36	HIM
Valence	85.94	HOG	92.89	HIM	90.08	HOG
Dominance	87.34	HOG	93.36	HIM	89.77	HIM
Liking	86.33	LBP	93.13	HIM	89.38	HIM
Distress Phase	82.03	HOG	91.41	HIM	86.02	HIM

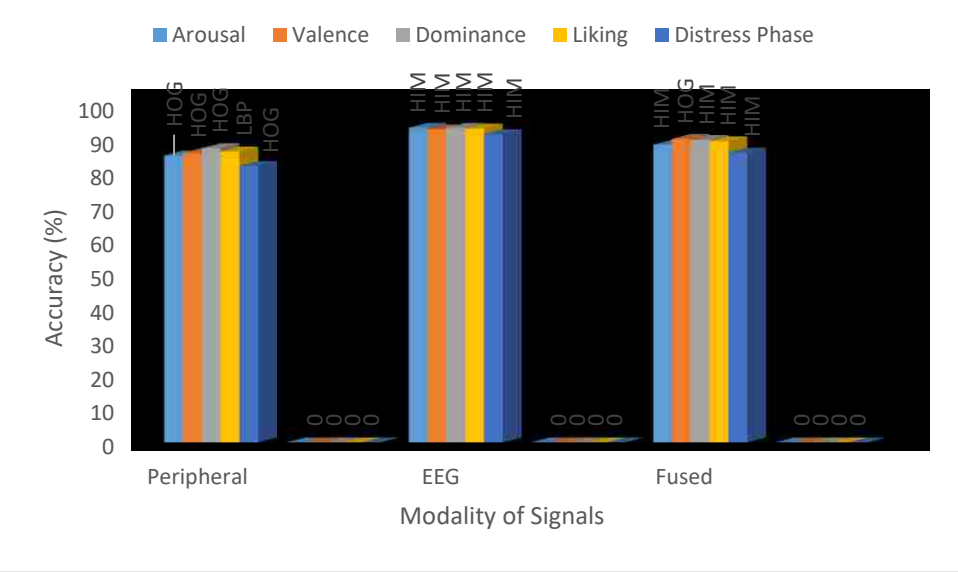


Figure 6.18: Plots of experimental results summary across dimensions, modalities and features.

However, because the aim of this study is to detect and track individuals in a distress phase situation, the distress phase emotion model and HIM feature of the EEG and fused modalities; as well as the HOG features of the peripheral modality are hereby

recommended for incorporation in an emotion recognition module of an affective system. Bearing in mind the cognizance that, for instance with the EEG modality, the best recognition accuracy of 91.41% (MSE=0.5248) was achieved by the distress phase emotion model as shown in Figure 6.18 and Table 6.10. This result competes with the results of other classes within the same and across modalities despite having more number of classes and containing the characteristics of all the other four classes combined.

The various best results obtained by each of the three modalities and features are state of the art and better than the results obtained in recent research studies (Wang and Shang 2013; Li et al. 2015; Zhuang et al. 2017; Li et al. 2017; Yin et al. 2017; Alhagry, Fahmy and El-khoribi 2017; Menezes et al. 2017; Nakisa et al. 2018) including those reported by other authors as shown in Table 6.17 that utilised the DEAP dataset. These results obtained in the current research work conducted and reported in this dissertation are also better than those obtained by (Li et al. 2015; Li et al. 2017; Yin et al. 2017; Alhagry, Fahmy and El-khoribi 2017) despite the trending deep learning approaches applied in those studies, thereby making the developed distress phase model very novel with state of the art results obtained for human emotion recognition tasks.

Table 6. 17: Experimental results of other studies across dimensions, modalities and features.

Method		Arousal (%)	Valence (%)	Dominance (%)	Liking (%)	Features/Classifier/ Modality
Koelstra et al. (2012)	Peripherals	57.0	62.7	-	59.1	Statistical time-frequency domain features, power spectral features, difference between spectral power/Naïve Bayes
	EEG	62.0	57.6	-	55.4	
	MCA	65.1	61.8	-	67.7	
Chung and Yoon (2012)	2 classes	66.4	66.6	-	-	Power spectral density/ Bayes classifier/EEG
	3 classes	53.4	51.0	-	-	
Rozgic et al. (2013)		69.1	76.9	73.9	75.3	Spectral power differences between symmetric channel pairs/RBF-SVM/EEG
Wang and Shang (2013)		60.9	51.2	65.1	68.4	DBN generated features/DBNs/EOG+ EMG
Chen et al. (2015)		76.2±6.8	73.6±7.9	-	-	EEG connectivity features/SVM/EEG
Bahari and Janghorbani (2013)		64.6±10.7	58.1±9.3	-	67.42	Recurrence Quantification/KNN/ EEG
Naser and Saha (2013)		66.2	64.3	68.9	70.2	Sub-band features of real and imaginary trees/SVM/EEG
Zhuang et al. (2014)		68.4±12.1	76.9±6.4	73.9±11.1	75.3±10.6	Spectral power and spectral power differences/SVM/EEG

Li et al. (2015)		64.3	58.4	65.8	66.9	DBN features/Deep Belief Network/EEG
Liu, Zheng and Lu (2016)		80.5	85.2	84.9	82.4	Bimodal Deep Auto Encoder (BDAE) high network RBM features/multimodal deep learning/EEG
Menezes et al. (2017)	2 classes	69.2	88.4	-	-	Bandwaves spectral power density/SVM/EEG
	3 classes	59.5	55.9	-	-	
Chen et al. (2017)		77.57	43.57	-	-	Statistical and power spectral features/ Three-stage decision/EEG+Peripherals
Zhuang et al. (2017)		72.0±7.8	69.1±7.0	-	-	Intrinsic Mode Functions by Empirical Mode Decomposition/ SVM/EEG
Arnau-González, Arevalillo-Herráez and Ramzan (2017)	Naïve Bayes	65.6	68.0	-	-	Relative energy, spectral power and mutual information/Naïve Bayes, SVM/EEG
	SVM-RBF	67.7	69.6	-	-	
Tang et al. (2017)		83.23±2.6	83.82±5.0	-	-	Differential Entropy/Bimodal LSTM/EEG+Peripherals

Yin et al. (2017)	77.19	76.17	-	-	425 salient features/MESAE/EEG +Peripherals
Qiu, Li and Hu (2018)	84.79	86.45	-	-	Gated recurrent units features/ Correlated Attention Networks/ EEG + eye movements
Choi and Kim 2018)	74.65	78.0	-	-	Automatic feature extraction and selection/ LSTM model based Deep Learning /EEG+Peripherals
Liu et al. (2018)	74.3±8.4	77.2±8.6	-	-	14 different features extracted /Random Forest (EEG)
	72.4±8.8	76.1±7.4	-	-	14 different features extracted /SVM (EEG)

CHAPTER SEVEN

Summary, Conclusion and Future Works

The safety of lives and properties are germane for national development and peaceful co-existence. Human beings, who are the major operators of all the aspects of living, including health, sport, economy, education, transportation and agriculture among others, need to feel safe or secured to perform these basic functions. However, human beings are being faced with several threats to lives and properties through abductions/kidnappings, human trafficking, terrorism, road accidents, fire outbreaks, floods and various other emergency situations.

To enhance safety, remotely monitoring an individual through any or a combination of his emotional state, human activity and location information could prove vital and useful as a threat to safety can easily be identified and curtailed. This study specifically focused on human emotion recognition tracking of an individual with a view to detecting negative emotions resulting from potential threat and capable of leading to imminent danger. A succinct architectural description of how human emotion can be combined with human activity and location information to thwart the escalation of the imminent dangers that characterizes a distress phase situation model was also presented.

7.1 Summary

A thorough literature review of human emotion and its detection was done to discover the different emotion types, representation models, dimensions, elicitation in a laboratory environment, sensors for harvesting physiological data, emotion application domains as well as emotion recognition techniques among others. Several experiments utilising the peripheral physiological, EEG and fused (EEG+peripheral physiological) modality data were performed, using the HOG, LBP and HIM features extracted from the DEAP physiological signals dataset. The RBFNN pattern classifier was applied to these features and classification done along the arousal, valence, dominance, liking and distress phase

model dimensions with a view to comparing methods and discovering the combination that gave the best result in recognizing human emotion using the DEAP dataset.

The automatic recognition of human emotion remains very challenging in affective computing, as physiological data are continuously being researched to enhance emotion recognition accuracies. Different datasets and methods are continuously applied by various researchers hence the varied results that are obtained.

In the experiments conducted on the peripheral physiological data of the DEAP data set reported in this thesis, the HOG, LBP and the HIM feature extraction techniques of the digital image processing domain were leveraged on to extract discriminating features of the peripheral physiological data consisting of electrooculogram (EOG), electromyogram (EMG), GSR, respiration amplitude, skin temperature, BVP and body temperature. The Gaussian RBFNN pattern classifier was applied to the extracted features and results were compared as the intrinsic ability of the neural network model to handle the non-linearity in data to achieve promising results was explored.

With the peripheral physiological modality data, the best recognition result of 87.34% was achieved with the dominance dimension using the HOG features. This combination is recommended in an affective system if just a single emotional dimension is desired since it gave the best result among the arousal, valence, liking and distress phase dimensions. However, under this modality, for the emotion tracking of individuals with a view to prompting an emergency services as it has been designed in this study, the 82.03% result achieved by the emotion model named the distress phase dimension is recommended since it was built from the other four dimensions and bears their characteristics. With the results obtained here, there is a strong prospect of its adoption of an automated emotion recognition paradigm in application domains.

Similarly, in the various experiments conducted with the EEG modality, both the arousal and dominance dimensions recorded 93.36% performance accuracy achieved with the HIM features as their best recognition results but the dominance dimension has a lower MSE value. Nevertheless, for the reasons earlier highlighted under the peripheral

physiological modality, the 91.41% accuracy performance with the HIM features achieved by the distress phase dimension for the purpose of tracking an individual in a distress phase situation using EEG modality is hereby recommended.

Furthermore, the fused modality data recorded 90.08% along the valence dimension with the HOG features as its best result across the four emotion dimensions while 86.02% highest result was obtained by the distress phase dimension with the HIM features and therefore recommended for the fused modality physiological data.

However, for the purpose of emotion tracking of an individual in a distress phase situation, as shown in the results of the various experiments conducted, across all modalities and features, the best result of 91.41% accuracy achieved with the HIM features along the distress phase dimension is recommended above the 93.36% of the dominance dimension.

In order to analyse the architecture, experimentations and results obtained with respect to the aims and objectives of this study, the research questions of this study are re-stated below:

i. What physiological properties are suitable for identifying an individual's emotional state?
ii. How can existing methods be used to develop an intelligent system to identify an individual's emotional state?
iii. How can detection of and matching between body signals and an individual's emotional state be enhanced?
iv. How reliable is an automated system of identification of an individual's emotional state using data from an existing database?

The research objectives are also re-stated below, including how they have been achieved and how the research questions were accordingly addressed and explained.

The first research objective, as stated in section 1.3(i) was "to discover a set of physiological properties that are suitable for identifying an individual's emotional state".

This objective was addressed by an extensive literature survey. Links between some physiological signals and emotions were established and highlighted in Chapter 2 of this dissertation. Skin temperature, ECG, EEG, EMG, EOG, Heart Rate, BVP, respiration, GSR were reported in literature to change with different emotional states (Soleymani et al. 2012; Karthik, Varghese and Chaitanya 2013; Schneegaas et al. 2013; Solovey et al. 2014; Salazar-Lopez et al. 2015; De Witte et al. 2016). This motivated this current study to choose and utilise most of these physiological signals, as contained in the DEAP data set for the experimentations. The state-of-the-art results that were obtained using these identified physiological properties, from which features were consequently extracted surpassed the results in recent studies reported in (Wang and Shang 2013; Li et al. 2015; Zhuang et al. 2017; Li et al. 2017; Yin et al. 2017; Alhagry, Fahmy and El-khoribi 2017; Menezes et al. 2017; Nakisa et al. 2018) that utilised the DEAP dataset. These results are also better than what was reported by (Li et al. 2015; Li et al. 2017; Yin et al. 2017; Alhagry, Fahmy and El-khoribi 2017) despite the trending deep learning approaches applied in those studies thus affirming the assertion of the suitability of these physiological properties in identifying an individual's emotional state.

Subsequently, "to explore existing methods that could help to develop an intelligent system to identify an individual's emotional state" was the second objectives stated for this research. This objective was realised by extracting various features including the HOG, LBP and HIM from a standard and publicly available physiological signals emotion datasets, namely the Database for Emotion Analysis using Physiological Signals (DEAP).

In comparison with other methods, unique data preprocessing steps were developed in this study based on the tanh normalizer and a simplified form of inverse Fisher transformation was consequently applied in the pre-processed data. This is toward the extraction of local features with subsequent dimension reduction and feature selection using the PCA algorithm, which yielded discriminatory features that were eventually concatenated to give an improved result. The performances of the RBF-ANN pattern recognition algorithm with varied configurations applied to different extracted features of HOG, LBP and HIM along the five emotional dimensions of arousal, valence, dominance, liking and the developed distress phase emotion model was rigorously compared. The

best results along the various dimensions for each modality were noted while the best results of the distress phase emotion model, which was achieved with the HIMS features, were recommended for human emotion recognition for tracking of individual in a distress phase situation.

The study's third research objective, which was "to enhance detection of and matching between body signals and an individual's emotional state", was addressed by this section. The proposed architecture in chapter 3 postulated the capability of harvesting human physiological signals, using wearable sensors with a view to utilizing these signals for emotion recognition. Emotion detection is enhanced through appropriate and quality data collection, feature extraction procedures and classification algorithms. The incorporation of human activity recognition systems, involving the collection of human activity data using the accelerometer sensors and synching the recognised activity and emotion with the location information provided by the GPS, is appropriate for tracking of an individual in a distress phase situation. A procedural programming procedure can utilise the Bayesian, HMM and other machine learning algorithms for dynamic situation prediction and classify it accordingly to normal/happy, distress and casualty phases. An emergency response is only needed for the distress and casualty situations, but the aim is to thwart the progression from distress phase to casualty phase

However, the fourth research objective, which was "to test the reliability of the automated identification of an individual's emotional state by the intelligent system using data from an existing database", was addressed as this thesis describes evaluation metrics, including the MSE and accuracy (Domingo 2015) to validate and measure the reliability of the methods applied in the study. The classification accuracy is a measure of the ratio of correct predictions of the total number of samples. On the other hand, the MSE measures the mean of the square of the difference between the expected output and the actual output of a pattern classifier. The highest values obtained for the accuracy and lowest values for the MSE metrics are an indication of the excellent reliability of the automated emotion identification system. A testing module was also incorporated in the generic architecture of the human emotion recognition model. This was to test the various

models with ‘seen’ and ‘unseen’ physiological data as impressive test results were obtained indicating the models’ high performance with the experimental data.

7.2 Conclusion

Taking full advantage of the tanh normalizer and the inverse Fisher transform algorithm applied to emotion laden physiological signals while utilizing various digital image processing domain feature descriptors, a data pre-processing method, extracted discriminatory features as well as a human emotion model developed for emergency assistance, distinctive state of the art results for human emotion recognition was obtained. Some of these results surpassed those achievements with the deep learning approaches while also extending the research horizon, as the methods that have been applied including the features generated constitute a swerve from the traditional method.

7.3 Limitations

The experiments conducted in this thesis are only on physiological signals collected from the DEAP emotion corpus. The prospect of combining human emotions, human activity and location information of an individual was discussed in this study and only an architecture for its implementation was presented. Various experiments conducted were limited to emotion physiological signals while a field-testing using the proposed architecture has not been implemented.

7.4 Future Works

In future, the utilization of more other emotion based physiological signals corpuses, such as the MAHNOB-HCI and SJTU Emotion EEG Datasets (SEED) (Zheng and Lu 2017) physiological signals corpuses as well as human activities corpuses in order to perform

intensive experiments with other features and machine learning methods will be considered. This will enhance a robust implementation and field-testing of the proposed architecture of the HER, HAR and location detection system reported in this thesis for the purpose of real life deployment. The incorporation of an audio signal with physiological signals to enable a robust system will also be undertaken in future works while acknowledging the rigorous task to be done in order to fully, further develop and deploy an efficient real time emotion recognition module that can be incorporated in an affective system.

References

Abayomi, A., Olugbara, O.O., Adetiba, E., Heukelman, H. 2016. Training pattern classifiers with physiological cepstral features to recognize human emotion. Springer international publishing, Switzerland. N. Pillay et al. (eds.), *Advances in Nature and Biologically Inspired Computing. Proceedings of the 7th World Congress on Nature and Biologically Inspired Computing (NaBIC2015)*, Pietermaritzburg. South Africa. Advances in Intelligent Systems and Computing. 419: 271-280.

Abdelrahman, S.A. and Abdelwahab, M.M. 2018. Accumulated grey-level image representation for classification of lung cancer genetic mutations employing 2D principle component analysis. Electronics Letters, 54(4): 194–196.

Abdulkadir, S., Yanhui, G. and Yaman, A. 2016. Time–frequency texture descriptors of EEG signals for efficient detection of epileptic seizure. *Brain Informatics*, 3(2): 101–108.

Abe, B.T., Olugbara, O.O. and Marwala, T. 2014. Experimental comparison of support vector machines with random forests for hyperspectral image land cover classification. *Journal of Earth System Science*, 123(4): 779–790.

Abraham, A. 2005. Artificial neural networks. In: Sydenham, P.H. and Thorn, R. eds. Handbook of measuring system design. John Wiley & Sons, Ltd., London. 901-908.

Ada, F. and Kau, R. 2013. Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(3):187-190.

Adetiba, E. and Olugbara, O. O. 2015. Improved Classification of Lung Cancer Using Radial Basis Function Neural Network with Affine Transforms of Voss Representation. *PLoS ONE*, 10(12): e0143542. doi:10.1371/journal.pone.0143542.

Ahmad, F., Roy, K., O'Connor, B., Shelton, J., Dozier, G. and Dworkin, I. 2014. Fly wing biometrics using modified local binary pattern, SVMs and random forest. *International Journal of Machine Learning and Computing*, 4(3): 279–285.

Ahonen, T., Hadid, A. and Pietikäinen, M. 2006. Face description with local binary patterns: application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12): 2037–2041.

Alarcao˜, S. M. and Fonsec, M. J. 2017. Identifying emotions in images from valence and arousal ratings. *Multimedia Tools and Application*, 1-23.

Alhagry, S., Fahmy, A.A. and El-Khoribi, R.A. 2017. Emotion recognition based on EEG using LSTM recurrent neural network. *International Journal of Advanced Computer Science and Applications*, 8(10):355-358.

Ali, K.M. and Pazzani, M.J. 1996. Error reduction through learning multiple descriptions. *Machine Learning*, 24(3): 173– 202.

Al-Shawaf, L., Conroy-Beam, D., Asao, K and Buss, M.D. 2016. Human emotions: An evolutionary psychological perspective. *Emotion Review*, 8(2):1–14.

Andina, D. and Pham, D. C. 2007. Computational Intelligence for Engineering and Mathematics. Andina D and Pham DC eds. Springer Science & Business Media, Netherlands. Chapter 5:120-121.

Andreoni, G., Standoli, C.E. and Perego, P. 2015. Wearable monitoring of elderly in an ecologic setting: the SMARTA project. In: Proceedings 2nd International conference on sensors and applications, Online.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S. 2012. The 'K' in K-fold cross validation. In: Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 441–446.

Anthony, H. and Seth, F. 2018. Multimodal sentiment analysis to explore the structure of emotions. In KDD ’18: The 24th ACM SIGKDD International Conference on knowledge discovery & data Mining, London, United Kingdom.

Arnau-González, P., Arevalillo-Herráez, M. and Ramzan, N. 2017. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals. *Neurocomputing,* 244(2017): 81-89.

Atallah, L., Lo, B., Ali, R., King, R. and Yang, G. 2009. Real-time activity classification using ambient and wearable sensors. *IEEE Transactions on Information Technology in Biomedicine,* 13(6): 1031-1039.

Ateke, G., Ataollah, A. and Atefeh, G. 2016. Combination of empirical mode decomposition components of HRV signals for discriminating emotional states. *Iranian Journal of Medical Physics,* 13(2): 86-99.

Avcı, C., Akbaş, A. and Yüksel, Y. 2014. Evaluation of statistical metrics by using physiological data to identify the stress level of drivers. 3rd International Conference on Environment, Chemistry and Biology IPCBEE, Mauritius: 78(2014).

Averill, J. R. 1980. A constructivist view of emotion. In: Plutchik, R. and Kellerman, H. eds., *Emotion: Theory, research, and experience.* New York: Academic Press, 305–339.

Bao, L. and Intille, S. S. 2004. Activity recognition from user-annotated acceleration data. *Pervasive*: 1–17.

Bahari, F. and Janghorbani, A. 2013. EEG-based emotion recognition using recurrence plot analysis and K nearest neighbor classifier. *In proceedings of 20th Iranian Conference on Biomedical Engineering (ICBME 2013)*, Tehran, Iran.

Baveye, Y., Dellandrea, E., Chamaret, C. and Chen, L. 2015. LIRIS-ACCEDE: A video database for affective content analysis. *IEEE Transactions on Affective Computing*, 6(1): 43–55.

Beauxis-Aussalet, E. and Hardman, L. 2014. Visualization of confusion matrix for non-expert users. *In* P*roceedings of the IEEE symposium on Information Visualization (IEEE InfoVis)*.

Benedek, K., Shayn, L. and Mahzarin, R. B. 2017. Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, 49(2): 457-470.

Benitez-Quiroz, C.F., Srinivasan, R. and Martinez, A.M. 2016. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June –1 July 2016: 5562–5570.*

Berchtold, M., Budde, M., Gordon, D., Schmidtke, H. and Beigl, M. 2010a. Actiserv: Activity recognition service for mobile phones. In: *International Symposium on Wearable Computers, (ISWC)* Seoul, South Korea*:* 1–8.

Berchtold, M., Budde, M., Schmidtke, H. and Beigl, M. 2010b. An extensible modular recognition concept that makes activity recognition practical. In: *Advances in Artificial Intelligence, Lecture Notes in Computer Science.* Berlin/Heidelberg: Springer, 400–409.

Bergström, J., Dahlström, N., Henriqson, E. and Dekker, S. 2010. Team coordination in escalating situations: An empirical study using mid-fidelity simulation. *Journal of Contingencies and Crisis Management,* 18(4): 220-230.

Betella, A. and Verschure, P.F.M.J. 2016. The affective slider: a digital self-assessment scale for the measurement of human emotions. *PLoS ONE*, 11 (2): e0148037. doi:10.1371/journal.pone.0148037.

Beverley, F. and Russell, J.A. 1984. Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology*: *General*, 113(3): 464-486.

Bishop, C.M. 2006. *Pattern recognition and machine learning*. Springer, New York.

Boashash, B., Mesbah, M. and Coldit, P. 2003. Time frequency detection of EEG abnormalities. In: Boashash, B. ed. *Time-frequency signal analysis and processing: A Comprehensive Reference.* Amsterdam: Academic Press, 663-670.

Bonner, M. A. and Wilson, G. F. 2001. Heart rate measures of flight test and evaluation. International Journal Aviation Psychology, 12(1): 63–77.

Borges, M.R.S., Engelbrecht, A. and Vivacqua, A.S. 2011. Digital tabletops for situational awareness in emergency situations. In: Proceedings of the 15th International Conference on computer supported cooperative work in design (CSCWD), Lausanne, Switzerland.

Bors, A.G. 2001. Introduction of the Radial Basis Functions (RBF) Networks. The York Research database. Online Symposium of Electronics Engineers.

Bos, D.O. 2006. EEG-based emotion recognition: the influence of visual and auditory stimuli. *Emotion,* 57(7):1798–806.

Bradley, M.M. and Lang, P.J. 1999a. International Affective Digitized Sounds (IADS): stimuli, instruction manual and affective ratings, *Tech. Rep. No. B-2, Gainesville, FL*: The Centre for Research in Psychophysiology, University of Florida.

Bradley, M. M. and Lang, P. J. 2007. *International Affective Digitized Sounds* (IADS-2): *stimuli, instruction manual and affective ratings*. 2nd ed. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Bradley, M.M. and Lang, P.J. 1999b. *Affective Norms for English Words (ANEW): stimuli, instruction manual and affective ratings*. Technical report C-1, Gainesville, FL.: The Center for Research in Psychophysiology, University of Florida.

Breiman, L. 1998. Arcing classifiers. *The Annals of Statistics,* 26(3): 801–849.

Brezmes, T., Gorricho, J. L. and Cotrina, J. 2009. Activity recognition from accelerometer data on a mobile phone. In: *Distributed Computing, Artificial Intelligence, Bioinformatics,*

Soft Computing, and Ambient Assisted Living, (5518) of Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, 796–799.

Brian, J.R. and Daniel, H.M. 2008. Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin,* 34(5): 907–926.

Broek van den, E.L., Schut, M.H., Westerink, J.H.D.M. and Tuinenbreijer, K. 2009. Unobtrusive Sensing of Emotions (USE). *Journal of Ambient Intelligence and Smart Environments*, 1(3): 287-299.

Brouwer, A.M., van Wouwe, N., Mühl, C., van Erp, J. and Toet, A. 2013. Perceiving blocks of emotional pictures and sounds: effects on physiological variables. *Frontiers in Human Neuroscience*, 7(295).

Brusco, M. and Nazeran, H. 2005. Development of an intelligent PDA-based wearable digital phonocardiograph. In proceedings: International Conference of the IEEE Engineering in Medicine and Biology Society. Shanghai, China. 4:3506-9.

Buciu, I., Kotropoulos, C. and Pitas, I. 2001. Combining support vector machines for accurate face detection. In: p*roceedings of International Conference on Image Processing*, Thessaloniki, Greece, 1(2001): 1054-1057.

Burkhardt, F. 2005. Emofilt: the simulation of emotional speech by prosody-transformation. In proceedings of Interspeech '05, Lisbon, Portugal. 509-512.

Cacioppo, J. T. and Tassinary, L. G. 1990. Inferring physiological significance from physiological signals. *The American Psychologist*, 45(1):16–28.

Carvalho, S., Leite, J., Galdo-Alvarez, S. and Goncalves, O. 2012. The Emotional Movie Database (EMDB): a self-report and psychophysiological study. *Applied Psychophysiology and Biofeedback*, 37 (4): 279–294.

Castrillón-Santana, M., Lorenzo-Navarro, J. and Ramón-Balmaseda, E. 2013. Improving gender classification accuracy in the wild. *Progress in pattern recognition, image analysis, computer vision and applications. Lecture Notes in Computer Science*. José Ruiz-Shulcloper, Gabriella Sanniti di Baja eds., 8259: 270-277.

Cha, I. and Kassam, S. A. 1995. Channel equalization using adaptive complex radial basis function network. *IEEE Journal on Selected Areas in Communications*, 13(1):122-31.

Chadha, R. 2016. The Junto Emotion Wheel: Why and How We Use It. Funto Institute of Entrepreneurial Leadership. Chicago. U.S.A. http://blog.thejuntoinstitute.com/the-junto-emotion-wheel-why-and-how-we-use-it. Accessed on 30 September, 2018.

Chanel, G., Kronegg, J., Grandjean, D. and Pun, T. 2006. Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals. *Multimedia Content Representation, Classification Security,* 4105: 530-537.

Chanel, G., Kierkels, J.J.M., Soleymani, M. and Pun, T. 2009. Short-term emotion assessment in a recall paradigm. *International Journal of Human-computer Studies*, 67(8):607-627.

Chapman, R. A., Norman, D. M., Zahirniak, D. R, Rogers, S. K. and Oxley, M. E. 1991. Classification of correlation signatures of spread spectrum signals using neural networks. *In: Proceeding of IEEE National Aerospace and Electronics Conference (NAECON) Dayton, OH, USA, 1: 485-491*.

Chen, G., Anas, S.I., Aguilar–Pelaez, E. and Rodriguez–Villegas, E. 2015. Algorithm for heart rate extraction in a novel wearable acoustic sensor. *Healthcare Technology Letters*. 2(1): 28–33.

Chen, J., Hu, B., Wang, Y., Moore, P., Dai, Y., Feng, L. and Ding, Z. 2017. Subject-independent emotion recognition based on physiological signals: a three-stage decision method. *BMC Medical Informatics and Decision Making*. 17(3): 167.

Chen, M., Han, J., Guo, L., Wang, J. and Patras, I. 2015. Identifying valence and arousal levels via connectivity between EEG channels. *In proceedings 2015 International Conference on Affective Computing and Intelligent Interaction (ACII 2015)*, Xi'an, China.

Chen, R., Sharman, R., Rao, H. and Upadhyaya, S. 2011. An exploration of coordination in emergency response management. *Communications of the ACM*, 51 (5): 66-73.

Chen, Y.H., Chen, H. H., Chen, T. C., Chen L. G. 2011. Robust heart rate measurement with phonocardiogram by on-line template extraction and matching. IEEE Engineering in Medicine and Biology Society (EMBC), Boston, MA, USA. 2011: 1957-60.

Chen, Y.P., Yang, J.Y., Liou, S.N., Lee, G.Y. and Wang, J.S. 2008. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. *Applied Mathematics and Computation*, 205 (2):849–860.

Choi, E.J. and Kim, D.K. 2018. Arousal and valence classification model based on Long Short-Term Memory and DEAP Data for mental healthcare management. *Healthcare Informatics Research*, 24(4): 309–316.

Chollet, F. and Allaire, J.J. 2017. Deep learning with R. MEAP edition. Manning Publications Co. Shelter Island. New York. 11964. Page 10.

Choujaa, D. and Dulay, N. 2008. Tracme: Temporal activity recognition using mobile phone data. *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, *Shanghai, China*, 1:119–126.

Christin, D., Reinhardt, A., Kanhere, S. S. and Hollick, M. 2011. A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software*, 84 (11):1928–1946.

Christos, D.K., Nikolaos, K., George, G. and Dimitrios, I.F. 2008. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on systems, man and cybernetics – part A: systems and humans,* 38(3): 502-512.

Chubb, H and Simpson, J. M. 2012. The use of Z-scores in paediatric cardiology. *Journal of Annals of Pediatric Cardiology*, 5(2): 179–184.

Chung, S.Y. and Yoon, H.J. 2012. Affective classification using Bayesian classifier and supervised learning, *Proceedings of 12th International Conference on Control, Automation and Systems (ICCAS 2012)*, JeJu Island, South Korea.

Chun-yan, N., Hai-xin, S. and Wang, J. 2013. The relationship between chaotic characteristics of physiological signals and emotion based on approximate entropy. In: *proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, Hangzhou, China: 552-555.

Ciuti, G., Mencassi, A. and Dario, P. 2011. Capsule endoscopy: from current achievements to open challenges. *IEEE Reviews in Biomedical Engineering,* 4: 59–72.

Cohen, D. S. 2012. Kidnapping for ransom: the growing terrorist financing challenge.

Cohn, J.F. and Tronick, E.Z. 1988. Mother-infant interaction: the sequence of dyadic states at three, six, and nine months. *Development Psychology*, 24(3): 386-392.

Colan, S. D. 2013. The Why and How of Z Scores. *Journal of the American Society of Echocardiography,* 26(1): 38–40.

Collet, C., Clarion, A., Morel, M., Chapon, A. and Petit, C. 2009. Physiological and behavioural changes associated to the management of secondary tasks while driving. *Applied Ergonomics*, 40(6): 1041–46.

Cong, Z. and Chetouani, M. 2009. Hilbert-Huang transform based physiological signals for emotion recognition. *International symposium on Signal processing and Information Technology (ISSPIT), Ajman, United Arab Emirates*: 334-339.

Cooking Hacks. 2017. https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical/. Accessed on 15 June, 2018.

Cooper, E.A., Garlick, J., Featherstone, E., Voon, V., Singer, T. and Critchley, H.D. 2014. You turn me cold: evidence for temperature contagion. *PLoS ONE*, 9(12): e116126. https://doi.org/10.1371/journal.pone.0116126.

Cortes, C. and Vapnik V. 1995. Support-vector networks. *Machine Learning,* 20(3): 273–297. Kluwer Academic Publishers, Boston.

Crawford, J. R. and Henry, J. D. 2004. The Positive and Negative Affect Schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 43: 245–265.

Curry**,** C. and Hughes, D. 2014, What You Need to Know Now About the Nigerian Kidnapped Girls. ABC News. http://abcnews.go.com/International/nigerian-kidnapped-girls/story?id=23590323. Accessed on 15 April, 2016.

Cybenko, G. 1989. Approximation by super positions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4): 303–314.

Dalal, N. and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, Calif, USA*: 886–893.

Damasio, A.R. 1999. The feeling of what happens: body and emotion in the making of consciousness. New York, Harcourt Brace.

Dan-Glauser, E.S. and Scherer, K.R. 2011. The Geneva Affective Picture Database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2):468–477.

Davidson, R.J., Jackson, D.C. and Kalin, N.H. 2000. Emotion, plasticity, context and regulation: Perspectives from affective neuroscience. *Psychological Bulletin,* 126: 890-909.

Dawson, M. E., Schell, A. M. and Filion, D. L. 2017. Handbook of psychophysiology. New York: Cambridge University Press.

Dawson, M.R.W., Kelly, D.M., Spetch, M.L. and Dupuis, B. 2010. Using perceptrons to explore the reorientation task. *Cognition.* 114: 207-226.

De Sousa, R. 1987. The rationality of emotion. Cambridge, MA. MIT Press.

De Witte, N.A.J., Sütterlin, S., Braet, C. and Mueller, S.C. 2016. Getting to the heart of emotion regulation in youth: the role of interoceptive sensitivity, heart rate variability and parental psychopathology. *PLoS ONE*, 11(10).

Dimitriev, D.A., Saperova, E.V. and Dimitriev, A.D. 2016. State anxiety and nonlinear dynamics of heart rate variability in students. *PLoS ONE* 11(1).

Dipankar, D. 2014. Activity recognition using histogram of oriented gradient pattern history. *International journal of computer science, engineering and information technology.* 4(4): 23-31.

Domingo, M. 2015. Computer vision for X-ray testing: Imaging, systems, Image databases and algorithms. 229-231.

Dooley, E. E., Golaszewski, N. M. and Bartholomew, J. B. 2017. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR mHealth and uHealth*, 5(3):e34.

Ekman, P. 1982. Emotion in the human face, Second ed. Cambridge University. Press.

Ekman, P. 1984. Expression and the nature of emotion. *In Scherer, K. and Ekman, P. (Eds.), Approaches to Emotion*, Hillsdale, NJ, Lawrence Erlbaum: 319-343.

Ekman, P., Davidson, R.J. and Friesen, W.V. 1990. Emotional expression and brain physiology II: The Duchenne smile. *Journal of Personality and Social Psychology*, 58:342-353.

Ekman, P. and Davidson, R. J (Eds.). 1994. The nature of emotion: moods, emotions and traits. Oxford University Press: 56-58.

Ekman, P., Matsumoto, V. and Friesen, W.V. 1997. Facial expression in affective disorders, what the face reveals. Ekman P and Rosenberg EL. eds. Oxford University press: 32-342.

Ekman, P. 1999. Handbook of cognition and emotion. Edited by Dalgleish T. and Power, M. John Wiley & Sons Limited.

Ekman, P. and Rosenberg, E.L. 2005. What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system, Second ed. Oxford University Press.

European Commission, 2012. Road transport - a change of gear. Luxembourg: Publications Office of the European Union.

Ferdinando, H., Ye, L., Seppänen, T. and Alasaarela, E. 2014. Emotion recognition by heart rate variability. *Australian Journal of basic & applied sciences,* 8(14): 50-55.

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Roning, J., Forlizzi, J.F. and Dey, A.K. 2014. Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. In Computational Intelligence, Cognitive Algorithms, Mind and Brain (CCMB), IEEE Symposium, Orlando, FL: 39-48.

Foerster, F., Smeja M. and Fahrenberg, J. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5): 571-583.

Fontaine, J.R.J., Scherer, K.R., Roesch, E.B. and Ellsworth, P.C. 2007. The world of emotions is not two-dimensional. *Psychological Science*, 18(12): 1050-1057.

Franken, R.E. 1994. Human Motivation, 3rd ed. Belmont, CA: Brooks/Cole Publishing Company.

Fried, E. 1976. The impact of nonverbal communication of facial affect on children's learning. PhD dissertation, Rutgers University.

Fuller, R. and Santos, J.A. (Eds) 2002. Human factors for highway Engineers. *Elsevier*, Langford lane, Oxford.

Gamero L.G. and Watrous R. 2003. Detection of the first and second heart sound using probabilistic models. *IEEE Engineering in Medicine and Biology Society*, *Cancun, Mexico.*

Garcia-Olalla, O., Alegre, E., Fernandez-Robles, L., Garcia- Ordas, M.T. and Garcıa-Ordas, D. 2013. Adaptive local binary pattern with oriented standard deviation (ALBPS) for texture classification. *EURASIP Journal on image and video processing,* 1(31): 1-11.

Garmin. 2016. Activity tracking and fitness metric accuracy. http://www.garmin.com/en-US/legal/atdisclaimer. Accessed on 15 April, 2018.

Gebejes, A. and Huertas, R. 2013. Texture characterization based on grey-level co-occurrence matrix. *In proceedings of the conference of Informatics and Management Sciences, ICTIC*. 375–378.

Glenn, E. 2010. The physics hypertextbook: Intensity vs amplitude. Physics info. https://physics.info/. Accessed on 13 December, 2018.

Goldberger, A. L., Amaral, L.A.N, Glass, L., Hausdorff, J.M., Ivanov, P.C.H., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K. and Stanley, H.E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]. Accessed on 12 June, 2018.

Gorman, J. M., Martinez, J. D., Coplan, J. D., Kent, J. and Kleber, M. 2004. The effect of successful treatment on the emotional and physiological response to carbon dioxide inhalation in patients with panic disorder. *Biological Psychiatry*, 56(11): 862–867.

Goshvarpour, A., Abbasi, A. and Goshvarpour, A. 2017. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. *Biomedical Journal*, 40 (2017): 355-368.

Grimnes, S., Jabbari, A., Martinsen, Ø.G. and Tronstad, C. 2011. Electrodermal activity by DC potential and AC conductance measured simultaneously at the same skin site. *Skin Research and Technology*, 17(1):26-34. Doi: 10.1111/j.1600-0846.2010.00459.x. Accessed on 10 May, 2017.

Gross, J. J. and Levenson, R. W. 1995. Emotion elicitation using films. *Cognition and Emotion*, 9(1): 87-108.

Grossman, P., Watkins, L.L., Wilhelm, F.H., Manolakis, D. and Lown, B. 1996. Cardiac vagal control and dynamic responses to psychological stress among patients with coronary artery disease. *The American journal of cardiology*: 78.

Gu, Y., Tan, S.-L., Wong, K.-J., Ho, M.-H. R. and Qu, L. 2010. A biometric signature based system for improved emotion recognition using physiological responses from multiple subjects *In 8th IEEE International Conference on Industrial Informatics (INDIN), Osaka*.

Guo, W., Brennan, D. and Blythe, P. 2013. Detecting older drivers' stress level during real-world driving task. In International conference on applied psychology and behavioural sciences. Paris, France. World Academy of Science, Engineering and Technology.

Haifeng, S., Jun, G., Gang, L. and Qunxia, L. 2005. Non-stationary environment compensation using sequential EM algorithm for robust speech recognition. In: Jorge, A. et al. (eds.): 264–273. Springer-Verlag, Berlin Heinderg.

Hamdam, J., Sethu, S., Smith, T., Alrevic, A., Alhaidari, M., Atkinson, J., Ayala, M., Box, H., Cross, M., Delaunois, A., Dermody, A., Govindappa, K., Guillon, J.M., Jenkins, R., Kenna, G., Lemmer, B., Meecham, K., Olayanju, A., Pestel, S., Rothfuss, A., Sidaway, J., Sison-Young, R., Smith, E., Stebbings, R., Tingle, Y., Jean-Pierre, V., Williams, A.,

Williams, D., Park, K. and Goldring, C. 2013. Safety pharmacology -Current and emerging concepts. *Toxicology and Applied Pharmacology*, 273(2):229-41.

Hamed, M., Sazzad, H.M., Rafael, A.C. 2014. Using remote heart rate measurement for affect detection. In p*roceedings of the twenty-seventh international Florida artificial intelligence research society conference:* 118-123.

Hammerschmidt, K. and Jurgens, U. 2007. Acoustical correlates of affective prosody. *Journal of Voice*, 21(5):531-540.

Hanai, Y., Nishimura, J. and Kuroda, T. 2009. Haar-like filtering for human activity recognition using 3d accelerometer. *IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*: 675–678.

Hanjalic A. and Xu, L. Q. 2005. Affective video content representation and modeling. *IEEE Transactions on multimedia*, 7(1):143-154.

Happy, S.L. and Aurobinda, R. 2015. Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing,* 6(1): 1-12.

Harrendorf, S., Heiskanen, M. and Malby, S. 2010. International Statistics on Crime and Justice.

Hartmut, R.P. and Christian, K. 2008. Amplitude and amplitude variation of emotional speech. *In proceedings of annual conference of the Interspeech '08*, Brisbane, Australia. 1036-1039.

Hassan, M.M., Uddin, Md. Z., Mohamed, A. and Almogren, A. 2018. A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, 81(2018):307-313.

Hastie, T., Tibshirani, R., Friedman, J. and Franklin, J. 2005. The elements of statistical learning: data mining, inference and prediction. *Mathematical Intelligence,* 27(2): 83–85.

He, Z.Y. and Jin, L.W. 2008. Activity recognition from acceleration data using ar model representation and SVM. *In proceedings*: *International Conference on Machine Learning and Cybernetics*, 4:2245–2250.

He, Z. and Jin, L. 2009. Activity recognition from acceleration data based on discrete consine transform and SVM. *In: IEEE International Conference on Systems, Man and Cybernetics*: 5041–5044.

Healey, J.A. and Picard, R.W. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transaction Intelligent Transportation Systems*, 6(2): 155-166.

Healey, J. 2014. Physiological sensing of emotion. In Calvo R, D'Mello S, Gratch J, Kappas A. (Eds.), *The oxford handbook of affective computing*. New York, NY: Oxford University Press.

Hein, G., Lamm, C., Brodbeck, C. and Singer, T. 2011. Skin conductance response to the pain of others predicts later costly helping. *PLoS ONE*, 6(8): e22759. https://doi.org/10.1371/journal.pone.0022759.

Helander, M. 1978. Applicability of drivers' electrodermal response to the design of the traffic environment. *Journal of Applied Psychology*, 63(4): 481-488.

Heng, Y. P., Lili, N.A., Alfian, A.H. and Puteri, S.S. 2013. A study of physiological signals-based emotion recognition systems. *International journal of computer and technology*, 11(1): 2189-2196.

Hightower, J. and Borriello, G. 2001. Location systems for ubiquitous computing. *Computer*, 34(8). IEEE Computer Society Press.

Hiscox Group, 2014. Report and Loss Accounts – 2013. https://www.hiscoxgroup.com/investors/report-and-accounts/latest-report. Accessed on 8 Decemeber, 2015.

Hollnagel, E. and Woods, D. D. 2005. Joint Cognitive Systems - foundations of cognitive systems engineering. CRC press.

Hölzle E. 2002. Pathophysiology of sweating. In Kreyden, O.P., Böni, R. and Burg, G. (eds): Hyperhidrosis and Botulinum Toxin in Dermatology. *Current Problems in Dermatology*, Basel, Karger, 30:10-22.

Honig, F., Wagner, J., Batliner, A. and Noth, E. 2009. Classification of user states with physiological signals: On-line generic features versus specialized feature sets. In: 17th European Signal Processing Conference 2009, Glasgow, Scotland: 2357–2361.

Hsu, C.W., Chang, C.C. and Lin, C.J. 2000. A Practical Guide to Support Vector Classification. Technical report.

Huang, K. L., Kanhere, S. S. and Hu, W. 2010. Preserving privacy in participatory sensing systems. *Computer Communications*, 33 (11): 266–1280.

Huang, Y. 2009. Advances in artificial neural networks—methodological development and application. *Algorithms (Basel).* 2(3): 973–1007.

Iglesias, J., Cano, J., Bernardos, A. M. and Casar, J. R. 2011. A ubiquitous activity-monitor to prevent sedentariness. *In IEEE Conference on Pervasive Computing and Communications*. 978-1-4244-9529-0/11/: 319-321.

Iman, B., Norihide, M., Hiroshi, M., Hasan, D. and Gholamreza. A. 2017. Histogram-based feature extraction from individual gray matter similarity-matrix for Alzheimer's disease classification. *Journal of Alzheimer's Disease*, 55(4):1571-1582.

Iris, B., Theo van der, V., Peter, V. and Jan de, B. 2014. Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology, 33(3):* 405–421.

Itsara, W. and Peerapon, V. 2014. An evaluation of feature extraction in EEG-based emotion prediction with Support Vector Machines. In *IEEE 11th International Joint Conference on Computer Science and Software Engineering (JCSSE),* 106-110.

Izard, C. E. 1972. Patterns of emotions: a new analysis of anxiety and depression. San Diego, CA: Academic Press.

Izard, C. E. 1977. Human emotions. New York: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-2209-0. Accessed on 19/08/2107.

Jabarouti, M.M., Rahmani, R. and Soltanian-Zadeh, H. 2009. Automatic segmentation of putamen using geometric moment invariants. *The 15th Iranian conference on biomedical engineering, Mashad, Iran*.

Jack, V.T. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology,* 49(11): 1225-1231.

Jain, A., Duin, R. and Mao, J. 2000. Statistical pattern recognition: A review. *IEEE Transaction pattern analysis and machine intelligence*, 22(1): 4-37.

James, L. and Diane, N. 2000. Road rage and aggressive driving. Steering clear of highway warfare. Prometheus Books, Amherst, NY.

Jatoba, L. C., Grossmann, U., Kunze, C., Ottenbacher, J. and Stork, W. 2008. Context-aware mobile health monitoring: evaluation of different pattern recognition methods for classification of physical activity. *30thAnnual International Conference of the IEEE Engineering in Medicine and Biology Society*: 5250–5253.

Jerritta, S., Murugappan, M., Nagarajan, R. and Wan, K. 2011. Physiological signals based human emotion recognition: a review. *Signal Processing and its Applications (CSPA), Proceedings of IEEE 7th International Colloquium*: 410-415.

Jirayucharoensak, S., Pan-Ngum, S. and Israsena, P. 2014. EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation. Hindawi: *The Scientific World Journal*, 2014:10. Article ID 627892. http://dx.doi.org/10.1155/2014/627892.

Johnson, G. 2009. Theories of emotion. (http://www.iep.utm.edu/e/emotion.htm) Accessed on 19 June, 2018 from Internet Encyclopedia of Philosophy.

Jonghwa, K. and Andre, E. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern analysis and machine Intelligence*, 30(12): 2067-2083.

Jovanov, E., Milenkovic, A., Otto, C. and De Groen, P. C. 2005. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of neuro engineering and rehabilitation*, 2(1): 6.

Jukka, K., Suvi, T., Xiaohua, H., Xiaobai, L., Seppo, L., Matti, P. and Tapio, S. 2012. Multimodal emotion recognition by combining physiological signals and facial expressions: A Preliminary Study. *34th Annual International Conference of the IEEE EMBS* San Diego, California USA.

Jung, J., Ha, K., Lee, J., Kim, Y. and Kim, D. 2008. Wireless body area network in a ubiquitous healthcare system for physiological signal monitoring and health consulting. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 1 (1): 47-54.

Junkai, C., Zenghai, C., Zheru, C. and Hong, F. 2014. Facial expression recognition based on facial components detection and HOG Features. *Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields,* Koc University, Istanbul/Turkey. 64-69.

Jun-Wen, T., Adriano, O.A., Hang, L., Steffen, W., David, H., Stefanie, R., Kerstin, L., Holger, H. and Harald, C.T. 2016. Recognition of intensive valence and arousal affective states via facial electromyographic activity in young and senior adults. *PLoS ONE*, 11(1): e0146691. doi:10.1371/journal.pone.0146691.

Kanade, T., Cohn, J. and Tian, Y. 2000. Comprehensive database for facial expression analysis. *In Proceedings of IEEE international Conference on Face and Gesture Recognition (AFGR '00)*: 46-53.

Kao, T.P., Lin, C.W. and Wang, J.S. 2009. Development of a portable activity detector for daily activity recognition. *IEEE International Symposium on Industrial Electronics*: 115–120.

Karthik, S., Varghese, A. and Chaitanya, S. 2013. Analysis of physiological signals in response to stress using ECG and respiratory signals of automobile drivers. Imac4s: 574-579.

Kasteren, T. V, Englebienne, G. and Krse, B. 2010. An activity monitoring system for elderly care using generative and discriminative models. *Personal and Ubiquitous Computing, 14(6):* 489–498.

Katsis, C. D., Ntouvas, N. E., Bafas, C. G. and Fotiadis, D. I. 2004. Assessment of muscle fatigue during driving using surface EMG. In Proceedings, 2nd IASTED International Conference, Biomedical Engineering: 259–262.

Keltner, D., Haidt, J., and Shiota, M. N. 2006. Social functionalism and the evolution of emotions. In M. Schaller, J. A. Simpson, D. T. Kenrick (Eds.), Evolution and social psychology. New York: Psychology Press: 115–142.

Kim, E., Helal, S. and Cook, D. 2010. Human activity recognition and pattern discovery. *Pervasive Computing*, *IEEE*, 9(1): 48-53.

King, R.D., Feng, C. and Shutherland, A. 1995. Statlog: Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence*, 9(3): 259-287.

Koelsch, S. 2014. Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience,* 15(3): 170–180.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A. and Patras, I. 2012. DEAP: a database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1): 18– 31.

Kollias, S. and Karpouzis, K. 2005. Multimodal emotion recognition and expressivity. Multimedia and Expo, *ICME 2005. IEEE International Conference*: 779-783.

Komensky, T., Jurcisin, M., Ruman, K., Kovac, O., Laqua, D. and Husar, P. 2012. Ultra-wearable capacitive coupled and common electrode-free ECG monitoring system. In proceedings of the 2012 annual international conference of the IEEE engineering in medicine and biology society, San Diego, CA, USA. 1594–1597.

Konica Minolta. 2014. Oxygen saturation monitor PULSOX-300i. Available at http://www.konicaminolta.com/. Accessed on 7 June, 2018.

Krug, E., Sharma, G. and Lozano, R. 2000. Commentaries: The global burden of injuries. *American Journal of Public Health*, 90 (4): 523-526.

Kumar, P. and Lahudkar, S.L. 2016. Automatic speaker recognition using LPCC and MFCC. *International Journal of Recent and Innovation Trends in Computing and Communication,* 3(4):2106–2109.

Labatut, V. and Cherifi, H. 2012. Accuracy measures for the comparison of classifiers*. In proceedings of the 5th International Conference on Information Technology, amman : Jordanie, 2011*.

Lachezar, B., Petia, G., Isabel, S., Ana, P. and Carlos, S. 2015. EEG-based subject independent affective computing models. *INNS Conference on Big Data*, 53:375–382.

Lakens, D., Fockenberg, D.A., Lemmens, K.P., Ham, J. and Midden, C.J. 2013. Brightness differences influence the evaluation of affective pictures. *Cognition and Emotion*, *27(7):*1225-46.

Lan L. and Ji-hua, C. 2006. Emotion recognition using physiological signals from multiple subjects. In *Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP '06*: 355-358.

Lanatà et al. 2015. How the autonomic nervous system and driving style change with incremental stressing conditions during simulated driving. *IEEE Transactions on Intelligent Transportation Systems*, 16(3).

Lang, P. J., Bradley, M. M. and Cuthbert, B. N. 2008. International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual (Technical Report A-8). Gainesville: University of Florida, Center for Research in Psychophysiology.

Lara, O. D., Perez, A. J., Labrador, M. A. and Posada, J. D. 2011. Centinela: A human activity recognition system based on acceleration and vital sign data. *Journal on Pervasive and Mobile Computing*, 8 (2012): 717–729.

Larson, J. and Rodriguez, C. 1999. Road rage to road-wise. 1st edition, Tom Doherty Associates Inc. New York, NY.

Lass, R.N., Regli, W.C., Kaplan, A., Mitkus, M. and Sim, J.J. 2008. Facilitating communication for first responders using dynamic distributed constraint optimization. IEEE Conference on technologies for homeland security.

Lech, M., Stolar, M., Bolia, R. and Skinner M. 2018. Amplitude-frequency analysis of emotional speech using transfer learning and classification of spectrogram images. *Advances in science, technology and engineering systems journal*, 3(4):363-371.

Lee, J., Heo, J., Lee, W., Lim, Y., Kim, Y. and Park, K. 2014. Flexible capacitive electrodes for minimizing motion artifacts in ambulatory electrocardiograms. *Sensors,* 14:14732–14743.

Leimin, T., Michal, M., Catherine, L., Johanna, D. M., Theodoros, K., Patrizia, L., Thierry, P. and Guillaume, C. 2017. Recognizing induced emotions of movie audiences: are

induced and perceived emotions the same? 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII).

Leonard, J. A. and Kramer, M. A. 1991. Radial basis function networks for classifying process faults. *IEEE Control Systems Magazine*, 11(3):31-8.

Li, X., Zhang, P., Song, D., Yu, G., Hou, Y. and Hu, B. 2015. EEG based emotion identification using unsupervised deep feature learning. *In proceedings of SIGIR2015 Workshop on Neuro-Physiological Methods in IR Research. NeuroIR'15, August 2015, Santiago, Chile.*

Li, Y., Huang, J., Zhou, H. and Zhong, N. 2017. Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Applied Sciences*, 7(1060).

Liang, Y.L.Y, Reyes, M.L. and Lee, J.D. 2007. Real -time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems,* 8:340–350.

Lindasalwa, M., Mumtaj, B., Elamvazuthi, I. 2010. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *Journal of Computing*, 2(3):138–143.

Liu, J., Meng, H., Li, M., Zhang, F., Qin, R. and Nandi, A.K. 2018. Emotion detection from EEG recordings based on supervised and Unsupervised dimension reduction. *Concurrency and Computation Practice Experience*, (30).

Liu, W., Zheng, W.L. and Lu, B.L. 2016. Emotion recognition using multimodal deep learning. *In Proceedings: International Conference on Neural Information Processing (ICONIP 2016), A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, D. Liu (eds). Lecture Notes in Computer Science, Springer, Cham, 9948*.

Lizette, C. L. and Nasoz, F. 2005. Affective intelligent car interfaces with emotion recognition. *In proceedings of 11th International Conference on Human Computer Interaction, Las Vegas, NV, USA: 1-10*.

Louise, K.C. 2007. Crisis management in hindsight: cognition, communication, coordination and control. *Public Administration Review*, 67(s1): 189-197.

Lu, Z., Szafron, D., Greiner, R., Lu, P., Wishart, D.S., Poulin, B. et al. 2003. Predicting sub-cellular localization of proteins using machine-learned classifiers. Oxford University Press.

Luminita, V. 2010. An introduction to mathematical image processing. IAS, Park City Mathematics Institute, Utah.

Lundberg, J. and Asplund, M. 2011. Communication problems in crisis response. In proceedings of the 8th international ISCRAM conference. Lisbon, Portugal.

Maaoui, C. and Pruski, A. 2010. Emotion recognition through physiological signals for human machine communication. *Cutting Edge Robotics, Vedran Kordic (Ed)*.

Madzarov, G., Gjorgjevikj, D. and Chorbev, I. 2009. A multi-class SVM classifier utilizing binary decision tree. *Informatica,* 33(2): 233–241.

Majumder, S., Mondal, T. and Jamal Deen, M. 2017. Review wearable sensors for remote health monitoring. *Sensors*, 17: 130.

Maldwyn, W.T. 2012, Top ten kidnap hotspots. www.red24.com/uploads/topten_kidnapthreat_areas_040412.pdf. Accessed on 15 February, 2015.

Manar, M.F.D., Aliaa, A.A.Y. and Atallah, H. 2014. Spontaneous Facial expression recognition based on histogram of oriented gradients descriptor. *Computer and Information Science.* Canadian Center of Science and Education. 7(3): 31-37. ISSN 1913-8989 E-ISSN 1913-8997.

Martin, R., Nicholas, M., Sonia, E., Nico, P. and Jean-Marc, D. 2018. Advances in human factors in wearable technologies and game design. Advances in Intelligent Systems and Computing 608. Proceedings of the AHFE 2017 International Conference on Advances in Human Factors and Wearable Technologies, July 17-21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA. T Ahram and C. Falcao (eds.), DOI 10.1007/978-3-319-60639-2_2.

Matsumura, K., Rolfe, P., Lee, J. and Yamakoshi, T. 2014. iPhone 4s Photoplethysmography: Which light color yields the most accurate heart rate and normalized pulse volume using the iPhysioMeter application in the presence of motion artifact? *PLoS ONE*, 9(3): e91205. https://doi.org/10.1371/journal.pone.0091205.

Maurer, U., Smailagic, A., Siewiorek, D. P. and Deisher, M. 2006. Activity recognition and monitoring using multiple sensors on different body positions. *International Workshop on Wearable and Implantable Body Sensor Networks, Washington, DC, USA. IEEE Computer Society.*

McCormick, C. 2013. Computer vision and machine learning projects and tutorials. Radial Basis Function Network (RBFN) Tutorial. http://mccormickml.com/2013/08/15/radial-basis-function-network-rbfn-tutorial/. Accessed on 19 May, 2018.

McCraty, R. and Childre, D. 2004. The appreciative heart. The psychophysiology of positive emotions and optimal functioning. In Emmons, R.A. and McCullough, M.E (eds.), the psychology of gratitude. Oxford University, New York, NY: 230-255.

Media Center, World Health Organization (WHO). 2015. Despite progress, road traffic deaths remain too high. Available at http://www.who.int/mediacentre/news/releases/2015/road-safety-report/en/. Accessed on 22 August, 2018.

Ménard, M., Hamdi, H., Richard, P. Daucé, B. and Takehiko, Y. 2015. Emotion recognition based on heart rate and skin conductance. *In proceedings of the international*

conference on physiological computing system, ESEO, Angers, Loire Valley, France. SciTePress 2015, ISBN 978-989-758-085-7.

Menezes, M. L. R., Samara, A., Galway, L., Sant'Anna, A., Verikas, A., Alonso-Fernandez, F., Wang, H. and Bond, R. 2017. Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset. *Personal and Ubiquitous Computing*, 21(6):1003–1013.

Mimma, N., Gaetano, V., Alberto, G., Antonio, L. and Enzo, P.S. 2015. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, 6(4):1-1. DOI: 10.1109/TAFFC.2015.2432810.

Miyaji, M., Danno, M., Kawanaka, H. and Oguri, K. 2008. Driver's cognitive distraction detection using AdaBoost on pattern recognition basis. IEEE International Conference on Vehicular Electronics and Safety (ICVES), Columbus, OH, USA: 51–56.

Mohamad, F.S., Manaf, A.A., Chuprat, S. 2011. Nearest neighbor for histogram-based feature extraction. *In Proceedings of International Conference on Computational Science, ICCS 2011. Nanyang, Singapore*. Procedia Computer Science, 4 (2011): 1296–1305.

Mohamed, D. and Jean, M. 2011. Emotion recognition using dynamic grid-based HOG features. *Ninth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011)*, CA, USA. 21-25.

Mondher, F. and Ahmed, B.H. 2012. A comparative survey of ANN and Hybrid HMM/ANN architectures for robust speech recognition. *American journal of intelligent systems*. 2(1): 1-8.

Montefinese, M., Ambrosini, E., Fairfield, B. and Mammarella, N. 2014. The adaptation of the Affective Norms for English Words (ANEW) for Italian. Behavior Research Methods, 46(3):887-903. Doi: 10.3758/s13428-013-0405-3.

Moor, M. 2008. Kidnapping is Booming Business. A Lucrative Political Instrument for Armed Groups Operating in Conflict Zones. IKV Pax Christ: 64. ISBN 9070443414, 9789070443412.

Morsbach, H., & Tyler, W. J. 1986. A Japanese emotion: Amae. In R. Harré (Ed.). The social construction of emotions. Oxford, UK: Blackwell: 289–307.

Myrtek, M., Weber, D., Brugner, G. and Muller, W. 1996. Occupational stress and strain of female students: results of physiological, behavioral, and psychological monitoring. *Biological Psychology*, 42(3): 379-391. Psychophysiology of Workload.

Najstrom, M. and Jansson, B. 2007. Skin conductance responses as predictor of emotional responses to stressful life events. *Behaviour research and therapy*, 45(10):2456-63.

Nakisa, B., Rastgoo, N.M., Tjondronegoro, D. and Chandra, V. 2018. Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Systems with Applications*, 93(2018): 143-155.

Narayanan, M. R., Redmond, S. J., Scalzi, M. E., Lord, S. R., Celler, B. G. and Lovell, N. H. 2010. Longitudinal falls-risk estimation using triaxial accelerometry. *Biomedical Engineering, IEEE Transactions*, 57 (3): 534-541.

Naser, D.S. and Saha, G. 2013. Recognition of emotions induced by music videos using DT-CWPT. *Proceedings of 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT 2013)*, Kharagpur, India.

Nasoz, F., Alvarez, K., Lisetti, C. and Finkelstein, N. 2004. Emotion recognition from physiological signals using wireless sensors for presence technologies. Cognition, Technology & Work, 6(1): 4–14.

Nemati, E., Deen, M. and Mondal, T. 2012. A wireless wearable ECG sensor for long-term applications. *IEEE Communications Magazine*, 50:36–43.

Nguyen, T.T., Nguyen, L.M and Shimazu, A. 2007. Improving the accuracy of questions classification with machine learning. *In Proceedings of IEEE International Conference on Research, Innovation and Vision for the Future*. 234–241.

Niemic, C.P. 2002. Studies of Emotion: A theoretical and empirical review of psychophysiological studies of emotion. *Journal of Undergraduate Research,* 1:15-18.

Noppadon, J., Setha, P. and Pasin, I. 2013. Real-time EEG based happiness detection system. Hindawi Publishing Corporation. *The scientific world journal*, 2013 (article id 618649).

Noppadon, J., Setha, P. and Pasin, I. 2013. Emotion classification using minimal EEG channels and frequency bands. *In Proceedings of the 10th International Joint Conference on Computer Science and Software Engineering (JCSSE '13)*: 21–24.

Nweke, H.F., Teh, Y.W., Al-garadi, M.A and Alo, U.R. 2018 Deep learning algorithms for human activity recognition using Mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, 105 (2018):233–261.

NYA International Crisis Prevention and Response. 2016. Global Kidnap Review. https://presswire.com/sites/presswire.com/files/pr/nya/160203-NYA-January-Kidnap-Review.pdf. Accessed on 5 May, 2018.

Ogunduyile, O., Olugbara, O. O. and Lall, M. 2013. Development of wearable systems for ubiquitous Healthcare Service Provisioning. *APCBEE Procedia.* 7: 163-168.

Oh, K., Park, H.S and Cho S.B. 2010. A mobile context sharing system using activity and emotion recognition with Bayesian networks. *International Conference on Ubiquitous Intelligence Computing*: 244–249.

Ojala, T., Pietikainen, M. and Harwood, D. 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1): 51–59.

Ojala, T. and Pietikäinen, M. 1999. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(3): 477-486.

Ojala, T., Pietikäinen, M. and Mäenpää, T. 2002. Multi resolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7): 971-987.

Ojala, T., Valkealahti, K., Oja, E. and Pietikäinen, M. 2001. Texture discrimination with multidimensional distributions of signed gray-level differences. *Pattern Recognition,* 34(3): 727–739.

Olugbara, O., Ojo, S. and Adigun, M. 2011. A grid enabled framework for ubiquitous healthcare service provisioning, Advances in Grid Computing. Zoran Constantinescu. *InTech, Croatia, ISBN*: 978-953.

Oscar, D.L. and Labrador A.M. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications surveys & tutorials*, 15(3):1192-1209.

Özdemir, A. T. and Barshan, B. 2014. Detecting falls with wearable sensors using machine learning techniques. *Sensors*, 14 (6): 10691-10708.

Pachori, R.B. and Sircar, P. 2008. EEG signal analysis using FB expansion and second-order linear TVAR process. *Signal Processing,* 88: 415-420.

Palo, H.K., Mohanty, M.N. and Chandra, M. 2015. Use of different features for emotion recognition using MLP network. *I.K. Sethi (ed.), Computational vision and robotics, Advances in intelligent systems and computing 332*, Springer: 7-15.

Panagiotis, C.P. and Leontios, J.H. 2010. Emotion recognition from EEG using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):186-197.

Papantonopoulos, G., Takahashi, K., Bountis, T. and Loos, B.G. 2014. Artificial neural networks for the diagnosis of aggressive periodontitis trained by immunologic parameters. *PLoS ONE,* 9(3). e89757. doi:10.1371/journal.pone.0089757.

Park, J.H., Jang, D. G., Park, J. and Youm, S. K. 2015. Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric sensor. *Sensors*, 15: 23402–23417.

Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J. and Korhonen, I. 2006. Activity classification using realistic data from wearable sensors. *IEEE Transaction on Information Technology in Biomedicine*, 10(1): 119–128.

Patterson, J.A.C., Mcllwraith, D.G. and Yang, G.Z. 2009. A flexible low noise reflective PPG sensor platform for ear-worn heart rate monitoring. *Sixth international workshop on wearable and implantable body sensor networks. Berkeley*. 286-291.

Pazhanirajan, S. and Dhanalakshmi, P. 2013. EEG signal classification using linear predictive cepstral coefficient features. *International Journal of Computer Applications,* (0975–8887), 73(1):28-31

Peipei, S., Zhou, C. and Xiong, C. 2011. Automatic speech emotion recognition using support vector machine. *In proceedings of International Conference on Electronic and Mechanical Engineering and Information Technology*, 621–625.

Peira, N., Pourtois, G. and Fredrikson, M. 2013. Learned cardiac control with heart rate biofeedback transfers to emotional reactions. *PLoS ONE,* 8(7): e70004. https://doi.org/10.1371/journal.pone.0070004.

Perez, A.J., Labrador, M.A. and Barbeau, S.J. 2010. G-sense: a scalable architecture for global sensing and monitoring. *IEEE Network*, 24(4): 57-64.

Petrantonakis, P. C. and Hadjileontiadis, L. J. 2010. Emotion recognition from EEG using Higher Order Crossings. *IEEE Transactions on Information Technology in Biomedicine,* 14(2):186–197.

Picard, R.W., Vyzas, E. and Healey, J. 2001. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 23(10): 1175-1191.

Popescu, M. C., Balas, V. E., Perescu-Popescu, L. and Mastorakis, N. 2009. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7): 579–588.

Pung, H. K., Gu, T., Xue, W., Palmes, P. P., Zhu, J., Ng, W. L., Tang, C. W. and Chung, N. H. 2009. Context-aware middleware for pervasive elderly homecare. *IEEE Journal on Selected Areas in Communications*, 27 (4): 510-524.

Qi, P.L. 2011. Speaker Authentication. Springer Science and Business Media. 174–175.

Qiu, J., Li , X. and Hu, K. 2018. Correlated attention networks for multimodal emotion recognition. *In proceedings IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2018), Madrid, Spain*.

Rahim, M. A., Hossain, M. N., Wahid, T. and Azam, M. S. 2013. Face recognition using Local Binary Patterns (LBP). *Global Journal of Computer Science and Technology Graphics & Vision*, 13(4).

Ramakrishnan, S. 2012. Recognition of emotion from speech: A review. In: Ramakrishnan, S. (ed.) Speech Enhancement, Modeling and Recognition-Algorithms and Applications. Intec, (2012). ISBN: 978-953-51-0291-5.

Ramanan, D. and Forsyth, D. A. 2003. Finding and tracking people from the bottom up. *In: Proceedings 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'03). IEEE, II-467-II-474: 462.

Ramchand, H., Narendra, C. and Sanjay, T. 2013. Recognition of facial expressions using Local Binary Patterns of important facial parts. *International Journal of Image Processing (IJIP),* 7(2): 163-170.

Rani, P., Liu, C., Sarkar, N. and Vanman, E. 2006. An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern analysis & applications,* 9: 58-69.

Raschka, S. 2014. About Feature Scaling and Normalization – and the effect of standardization for machine learning algorithms. http://sebastianraschka.com/Articles/2014_about_feature_scaling.html. Accessed on 20 March, 2018.

Ravelo-Garcial, A.G., Navarro-Mesal, J.L., Hemadez-Perezl, E., Martin-Gonzalezl, S., Quintana-Moralesl, P., Guerra-Morenol, I., et al. 2103. Cepstrum feature selection for the classification of sleep Apnea-Hypopnea syndrome based on heart rate variability. *Computing in Cardiology*, 40: 959–962.

Redfern, W.S., Tse, K., Grant, C., Keerie, A., Simpson, D.J., Pedersen, J.C., Rimmer, V., Lauren, L., Klein, S. K., Karp, N. A., Sillito, R., Chartsias, A., Lukins, T., Heward, J., Vickers, C., Chapman, K. and Armstrong, J. D. 2017. Automated recording of home cage activity and temperature of individual rats housed in social groups: The Rodent Big Brother project. PLoS ONE, 12(9): e0181068. https://doi.org/10.1371/journal.pone.0181068.

Reeves, J. 1993: The face of interest. *Motivation and Emotion*, 17: 353-376.

Renukadevi, N.T. and Thangaraj P. 2013. Performance evaluation of SVM–RBF Kernel for medical image Classification. Global Journal of Computer Science and Technology Graphics and Vision: 13(4).

Repovš, G. 2010. Dealing with noise in EEG recording and data analysis. *Informatica Medica Slovenica,* 15(1): 18-25.

Riboni, D. and Bettini, C. 2011. Cosar: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing*, 15: 271–289.

Richardson, R. C. 1996. The prospects for an evolutionary psychology: human language and human reasoning. *Minds and Machines*, 6:541–557.

Rigas, G., Katsis, C.D., Ganiatsas, G. and Fotiadis, D.I. 2007. A user independent, biosignal based, emotion recognition method. *User Modeling, Lecture notes in Computer Science,* 4511: 314-318.

Rigas, G., Goletsis, Y., Bougia, P. and Fotiadis, D.I. 2011. Towards driver's state recognition on real driving conditions. *International Journal of Vehicular Technology*, 1-14.

Roche, D.T.D, Harney, M.T.D. and McDowell M.T.D. 2004. A framework for major emergency management.

Roggen D., Magnenat S., Waibel M., Tröster G. 2011. Wearable computing. *IEEE Robotics & Automation Magazine,* 2:83–95.

Roisman, G.I., Tsai, J.L. and Chiang, K.S. 2004. The emotional integration of childhood experience: physiological, facial expressive and self-reported emotional response during the adult attachment interview. *Developmental Psychology*, 40(5): 776-789.

Rottenberg, J., Ray, R.D. and Gross, J.J. 2007. Emotion elicitation using films. Handbook of emotion elicitation and assessment. *Series in affective science,* Oxford University. Press. 9-28.

Rozgic, V., Vitaladevuni, S.N. and Prasad, R. 2013. Robust EEG emotion classification using segment level decision fusion. *In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing: 1286–1290*.

Russell, J.A. 1980. A circumplex model of Affect. *Journal of Personality and Social Psychology,* 39(6): 1161-1178.

Russell, J.A. 1991. Culture and categorization of emotions. *Psychological Bulletin*, 110(3): 426-450.

Russell, J. A. and Mehrabian, A. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3): 273-294.

Saastamoinen, A., Pietilä, T., Värri, A., Lehtokangas and M., Saarinen, J. 1998. Waveform detection with RBF network—application to automated EEG analysis. *Neurocomputing*, 20: 1–13.

Sabyasachee, B., Rahul, G. and Shrikanth, N. 2017. A knowledge transfer and boosting approach to the prediction of affect in movies. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA: 2876 - 2880

Salasznyk, P. P. and Lee, E. E. 2006. A systems view of data integration for emergency response. *International Journal of Emergency Management*, 3(4): 313-331.

Salazar-López , E., Domínguez, E. , Juárez Ramos, V., de la Fuente, J., Meins, A., Iborra, O., Gálvez, G., Rodríguez-Artacho, M.A. and Gómez-Milán, E. 2015. The mental and subjective skin: Emotion, empathy, feelings and thermography. *Consciousness and Cognition*, 34(2015): 149-162.

Salehi, Y. and Darvishi, M. T. 2016. An Investigation of Fractional Riccati Differential Equations. *Optik - International Journal for Light and Electron Optics*, 127(23).

Sanderson, C. and Paliwal, K.K. 2004. Identity verification using speech and face information. *Digital signal processing,* 14(5): 449–480.

SAPS - South Africa Police Service. 2017. Crime situation in Republic of South Africa 2016-2017.

Sarah, N.A., Ayman, A. and Mostafa-Sami, M.M. 2015. Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal.* 16(213). ISSN 11108665.

Sarkar J., Vinh, L., Lee, Y.K. and Lee, S. 2010. Gpars: a general-purpose activity recognition system. *Applied Intelligence*: 1–18.

Savithiri, G. and Murugan, A. 2011. Performance analysis on half iris feature extraction using GW, LBP and HOG. *International journal of computer applications (0975-8887).* 22(2): 27-32.

Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. 2010. Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cognition and Emotion*, 24(7): 1153-1172.

Scherer, K.R. 2005. What are emotions? And how can they be measured?" *Social Sci. Information*, 44(4): 695–729.

Schlmilch, O.X., Witzschel, B., Cantor, M., Kahl, E., Mehmke, R. and Runge, C. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5):571-583.

Schneegaas, S., Pfleging, B., Broy, N., Schmidt, A. and Heinrich, F. 2013. A data set of real world driving to assess driver workload. *In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive UI '13). ACM, New York, NY, USA: 150-157*.

Shimizu, H. 2013. Shimizu's textbook of dermatology, Second edition. Willey Blackwell. ISBN-13: 978-1119099055.

Shiqing, Z., Xiaoming, Z. and Bicheng, L. 2012. Facial expression recognition based on Local Binary Patterns and Local Fisher Discriminant Analysis. *WSEAS Transactions on signal processing,* 1(8):21-31.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. 2011. Real-time human pose recognition in parts from single depth images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA*, 56(1):1297-1304.

Siemer, M., Mauss, I. and Gross, J.J. 2007. Same Situation—Different Emotions: How Appraisals Shape Our Emotions. *Emotion*, 7(3): 592–600.

Silvia, M.F. and Marius, D.Z. 2013. Emotion recognition in Romanian language using LPC features. In: The 4th IEEE International Conference on E-Health and Bioengineering—EHB 2013, Grigore T. Popa University of Medicine and Pharmacy, Iaşi, Romania, Nov 21–23, 2013.

Singh, Y.N. and Gupta, P. 2007. Quantitative evaluation of normalization techniques of matching scores in multimodal biometric systems. S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642: 574–583.

Smith, D. 2007. Smart clothes and wearable technology. *AI & Society 2007*, 22: 1–3.

Soares, A.P., Comesaña, M., Pinheiro, A.P., Simões, A. and Frade, C.S. 2012. The adaptation of the Affective Norms for English Words (ANEW) for European Portuguese. Behavior Research Methods, 44(1):256-69.

Soares, A.P., Pinheiro, A.P., Costa, A., Frade, C.S., Comesaña, M. and Pureza, R. 2013. Affective auditory stimuli: adaptation of the International Affective Digitized Sounds (IADS-2) for European Portuguese. Behavior Research Methods, 45(4):1168-81.

Soleymani, M., Lichtenauer, J., Pun, T. and Pantic, M. 2012. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1): 42–55.

Soleymani, M., Larson, M., Pun, T. and Hanjalic, A. 2014. Corpus development for affective video indexing. *IEEE Transactions on Multimedia*, 16(4): 1075.

Solovey, E., Zec, M., Perez, E. A. G., Reimer, B. and Mehler, B. 2014. Classifying driver workload using physiological and driving performance data: Two field studies. Presented at the CHI conference on human factors in computing systems, Toronto, Canada: 1–10.

Soma, B., Shanthi, T. and Madhuri, G. 2015. Emotion recognition using combination of MFCC and LPCC with supply vector machine. IOSR Journal of Computer Engineering. (IOSR-JCE), 17(4): 01–08.

SOMNOmedics GmbH. 2014. SOMNOscreen. Available at http://www.somnomedics.eu/. Accessed on 7 September, 2018.

Staffan, B.R.F.K.S. 2011. Emergency response systems: concepts, features, evaluation and design. Center for Advanced Research in Emergency Response (CARER). URL: http://www.liu.se/forskning/carer. Accessed on 20 October, 2018.

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S. and Samson, A.C. 2017. Facial expression analysis with AFFDEX and FACET: A validation study. Behavior Research Methods, 50(4): 1446-1460.

Taabish, G., Anand, S. and Sandeep, S. 2014. Comparative analysis of LPCC, MFCC and BFCC for the recognition of Hindi words using artificial neural networks*. International Journal of Computer Application (0975–8887)*, 101(12):22–27.

Tajitsu, Y. 2015. Piezoelectret sensor made from an electro-spun fluoropolymer and its use in a wristband for detecting heart-beat signals. *IEEE Transactions on Dielectrics and Electrical Insulation*, 22: 1355–1359.

Takeda, Y., Sato, T., Kimura, K., Komine, H., Akamatsu, M. and Sato, J. 2016. Electrophysiological evaluation of attention in drivers and passengers: toward an understanding of drivers' attentional state in autonomous vehicles. *Transportation Research Part F: Transport Psychology and Behaviour*, 42(1): 140-150.

Tamura, S. and Tateishi, M. 1997. Capabilities of a four-layered feedforward neural network: four layer versus three. *IEEE Transaction Neural Networks,* 8: 251-255.

Tang, H., Liu, W., Zheng, W. and Lu, B. 2017. Multimodal emotion recognition using deep neural networks. In proceedings *International Conference on Neural Information Processing (ICONIP 2017)*, D. Liu et al. (Eds.), Part IV, LNCS 10637 2017, 811–819.

Tapia, E. M., Intille, S. S., Haskell, W., Larson, K., Wright, J., King, A. and Friedman, R. 2007. Real-time recognition of physical activities and their intensities using wireless

accelerometers and a heart monitor. *International Symposium on Wearable Computers*, Boston, MA, USA.

Thagard, P. 2001. How to make decisions: coherence, emotion and practical inference. In Milligram E. (ed.), Varieties of practical inference. Cambridge, MA: MIT Press, 2001.

Thamizhvani, T. R., Hemalatha, R. J., Babu, B., Dhivya, A.J.A, Joseph, J.E. and Chandrasekaran, R. 2018. Identification of skin tumours using statistical and histogram based features. *Journal of Clinical and Diagnostic Research*, 12(9): 11-15.

Thammasan N, Moriyama K, Fukui K, Numao M. 2017. Familiarity effects in EEG-based emotion recognition. *Brain Informatics,* 4(1):39-50.

Tolstikov A., Hong, X., Biswas, J., Nugent, C., Chen, L. and Parente, G. 2011. Comparison of fusion methods based on dst and dbn in human activity recognition. *Journal*, *Control Theory and Applications*, 9: 18–27.

Tomkins, S. S. 1980. Affect as amplification: Some modifications in theory. *In R. Plutchik & H. Kellerman (Eds.), Emotion: theory research and experience.* New York: Academic Press, 1: 141-187.

Tseng, K.C., Lin, B.S., Liao, L. D., Wang, Y.T. and Wang, Y. L. 2014. Development of a wearable mobile electrocardiogram monitoring system by using novel dry foam electrodes. *IEEE System Journal*, 8(3):900–906.

Tu, J., Inthavong, K. and Loong Wong, K.L. 2015. Computational Hemodynamics—Theory, modelling and applications. Biological and medical physics, biomedical engineering, Springer. ISBN: 978-94-017-9594-4 (eBook).

Ugur, H. 2004. Artificial Neural Networks. Chapter 9, EE543 Lecture Notes.

Uhrig M.K., Trautmann N., Baumgärtner U., Treede R.D., Henrich F., Hiller W. and Marschall S. 2016. Emotion elicitation: a Comparison of pictures and films. *Frontiers in Psychology,* 7:180.

Uroš, M. and Božidar, P. 2015. Automated facial expression recognition based on histograms of oriented gradient feature vector differences. *Journal of Signal Image and Video Processing (SIViP)*, 9(1): 245–253.

Valentini, G., Muselli, M. and Ruffino F. 2003. Bagged ensembles of support vector machines for gene expression data analysis. In proceedings of the International joint conference on neural networks. *IEEE Computer Society,* 1844-1849.

Vannini, M., Detotto, C. and Mccannon, B. C. 2015. Ransom Kidnapping. Encyclopedia of Law and Economics. Springer. Editors: Jurgen Georg Backhaus.

Venkatesan, P. and Anitha, S. 2006. Application of a radial basis function neural network for diagnosis of diabetes mellitus. *Current Science*, 91(9): 1195-1199.

Vergara-Laurens, I. J. and Labrador, M. A. 2011. Preserving privacy while reducing power consumption and information loss in lbs and participatory sensing applications. *In: IEEE Workshop on Ubiquitous Computing and Networks*, 1247-1252.

Vyzas, E. and Picard, R.W. 1999. Offline and online recognition of emotion expression from physiological data. I*n proceedings: Workshop emotion-based agent architectures, third international conference on autonomous agents, Seattle, WA*: 135-142.

Walla, P., Brenner, G. and Koller M. 2011. Objective measures of emotion related to brand attitude: a new way to quantify emotion-related aspects relevant to marketing. *PLoS ONE*, 6(11): e26782. https://doi.org/10.1371/journal.pone.0026782

Wang, X., Jin, J. and Li, S. 2008. Measurement and analysis of heart signal based on the pressure sensor. Proceedings of 6th IEEE International conference on industrial informatics, Daejeon, 619-622.

Wang, X., Jin, C., Liu, W., Hu, M., Xu, L. and Ren, F. 2013. Feature fusion of HOG and WLD for facial expression recognition. In 2013 IEEE/SICE International Symposium on System Integration (SII), 227–232.

Wang, D. and Shang, Y. 2013. Modeling physiological data with deep belief networks. *International Journal of Information and Education Technology*, 3 (5): 505–511.

Wang, J., Chen, Y., Hao, S., Peng, X. and Hu, L. 2018. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 0000(2018):1-9.

Wan-Hui, W., Yu-Hui, Q., Guang-Yuan, L. 2009. Electrocardiography recording, feature extraction and classification for emotion recognition. In: *WRI World Congress on Computer Science and Information Engineering Los Angeles, CA*.

Watson, D. and Clark, L. A. 1994. The PANAS-X: Manual for the positive and negative affect schedule–expanded form.

Watson, D., Clark, L.A. and Tellegen A. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063-70.

Watson, J, B. 1919. Psychology from the standpoint of a behaviorist. Philadelphia: Lippincott.

Wenger, M. A. 1950. Emotion as visceral action: An extension of Lange's theory. *In M, L. Reymert (Ed.); The second international symposium on feelings and emotions*. New York: McGraw-Hill: 3-10.

White, J., Thompson, C., Turner, H., Dougherty, B. and Schmidt, D.C. 2011. WreckWatch: Automatic traffic accident detection and notification with smartphones. *Mobile networks and applications*, 16: 285-303.

Wiberg, H., Nilsson, E., Lindén, P., Svanberg, B. and Poom, L. 2015. Physiological responses related to moderate mental load during car driving in field conditions. *Biological Psychology*, 108: 115–125.

Wilson, G. F. 2001. An analysis of mental workload in pilots during flight using multiple psychophysiologic measures. *International Journal of Aviation Psychology*, 12 (1): 3–18.

Within reach, 2018. Most popular forms of transport. Available at http://www.within-reach.org.uk/most-popular-forms-of-transport.html. Accessed on 24/06/2018.

World Health Organization (WHO). 2015. Global status report on road safety.

World Health Organization (WHO). 2011. UN decade of action for road safety 2011-2020: saving millions of lives.

World Health Organization (WHO). 2014. Factsheet: Burns.

www.who.int/mediacentre/factsheets/fs365/en/. Accessed on 3 May, 2018.

Wundt, W. 1924. An introduction to psychology. Pintner, Trans., London: Allen & Unwin (Original work published 1912).

Xiang, L., Dawei, S., Peng, Z., Yazhou, Z., Yuexian, H. and Bin, H. 2018. Exploring EEG Features in Cross-Subject Emotion Recognition. *Frontiers in Neuroscience*, 12:162.

Xianhai, G. 2011. Study of emotion recognition based on electrocardiogram and RBF neural network. *Procedia Engineering*, 15: 2408–2412.

Xiao, Y., Wu, J., Lin, Z. and Zhao, X. 2018. A deep learning-based multi-model ensemble method for cancer prediction. *Computer methods and programs in Biomedicine*, 153 (2018): 1-9.

Xu, H. and Plataniotis, K.N. 2012. Affect recognition using EEG signal. *In proceedings of the 14th IEEE International workshop on Multimedia Signal Processing (MMSP 2012)*, Banff, Canada: 299–304.

Yamacli, M., Dokur, Z. and Ölmez, T. 2008. Segmentation of S1-S2 sounds in phonocardiogram records using wavelet energies. IEEE ISCIS, Istanbul, Turkey.

Yamakoshi, T., Yamakoshi, K., Tanaka, S., Nogawa, M. Park, S.B., Shibata, M., Sawada, Y., Rolfe, P. and Hirose, Y. 2008. Feasibility study on driver's stress detection from

differential skin temperature measurement. *IEEE 30th International conference in medicine and biology society*, 1076-1079.

Yan, O. and Nong, S. 2013. Robust automatic facial expression detection method", *Journal of Software*, 8(70): 1759-1764.

Yang J., Lee, J. and Choi, J. 2011. Activity recognition based on RFID object usage for smart mobile devices. *Journal Computer Science and Technology*, 26: 239–246.

Yin, J., Yang, Q. and Pan, J. 2008. Sensor-based abnormal human-activity detection. *IEEE Transaction of Knowledge and Data Engineering*, 20(8):1082–1090.

Yin, Z., Zhao, M., Wang, Y., Yang, J. and Zhang, J. 2017. Recognition of emotions using multimodal physiological signals and an ensemble deep learning. *Computer Methods and Programs in Biomedicine*, 140 (2017): 93–110.

Yisi, L., Olga, S. and Minh, K.N. 2011. Real-time EEG-based emotion recognition and its applications. *Gavrilova ML et al. (Eds.): Transactions on computational science XII, LNCS 6670*, Springer-Verlag Berlin Heidelberg. 256-277.

Yisi, L. and Olga, S. 2013. EEG Databases for Emotion Recognition. 2013 International Conference on Cyberworlds. IEEE. 978-1-4799-2245-1/13.

Yixiong, P., Peipei, S. and Liping, S. 2012. Speech emotion recognition using support vector machine. International Journal on Smart Home, 6(2): 101–108.

Yojna, A., Abhishek, S. and Abhay, B. 2014. A study of applications of RBF network. *International Journal of Computer Applications (0975–8887),* 94(2).

Yoo, K.S. and Lee, W.H. 2011. Mental stress assessment based on pulse photoplethysmography. IEEE 15th international symposium on consumer electronics, Singapore: 323-326.

Yoon, S. and Cho, Y. H. 2014. A skin-attachable flexible piezoelectric pulse wave energy harvester. *Journal of physics, conference series*: 557, 012026.

Yuen, P.W.T and Richardson, M.A. 2010. An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. *The Imaging Science Journal*, 58:241-253.

Yunan, Z., Yali, L. and Shengjin, W. 2015. Facial expression recognition using coarse-to-fine classifiers. *Computer Science and Applications – Hu (Ed.). Taylor & Francis Group, London*, ISBN: 978-1-138-02811-1: 127-135.

Zayed, N and Elnemr, H. A. 2015. Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities. International Journal of Biomedical Imaging. 2015 (Article ID 267807) : 7.

Zenn, J. 2018. The Terrorist Calculus in Kidnapping Girls in Nigeria: Cases from Chibok and Dapchi. Combating Terrorism Center at West Point CTCSENTINEL, 2(3): 1-8.

Zhang, T.T., Ser, W., Daniel G.Y.T., Zhang, J., Yu, J. Chua, C. and Louis, I.M. 2010. Sound based heart rate monitoring for wearable systems. International conference on body sensor networks, Singapore. IEEE Computer Society Washington, DC, USA. 139-143.

Zhao, S., Yao, H., Yang, Y. and Zhang, Y. 2014. Affective image retrieval via multi-graph learning. In ACM, MM'14 Proceedings of the 22nd ACM international conference on Multimedia Orlando, Florida, USA, 1025-1028.

Zheng, W.L. and Lu, B.L. 2017. A multimodal approach to estimating vigilance using EEG and forehead EOG. *Journal of Neural Engineering*, 14(2).

Zhihong, M., Pantic, Glenn, I.R. and Thomas, S.H. 2009. A survey of affect recognition methods: audio, visual and spontaneous expressions. *IEEE Transactions on pattern analysis and machine intelligence*, 31(1): 39-58.

Zhouyu, F., Guojun, L., Kai, M.T. and Dengsheng, Z. 2013. Optimizing cepstral features for audio classification. *In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China*: 1330–1336.

Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H. and Yan, B. 2017. Emotion recognition from EEG signals using multidimensional information in EMD domain. *BioMed Research International*, 2017 (Article ID 8317357): 9 pages.

Zhuang, X., Rozgic, V. and Crystal, M. 2014. Compact unsupervised EEG response representation for emotion recognition. *In proceedings of IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2014), Valencia, Spain.*

Zou, K. H., O'Malley, A.J. and Mauri, L. 2007. Receiver operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5):654–657.