DEVELOPMENT OF STRESS INDUCTION AND DETECTION SYSTEM TO STUDY IT'S EFFECT ON BRAIN

A Thesis

submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

NISHTHA PHUTELA

(1720446C602)

Under the supervision of

Dr. Devanjali Relan (BML Munjal Univeristy, Gurugram),
Prof. Goldie Gabrani (VIPS, New Delhi) and
Prof. Ponnurangam Kumaraguru (IIIT, Hyderabad)

BML MUNJAL UNIVERSITY™
FROM HERE TO THE WORLD

Bml Munjal University

Gurugram, Haryana, India-122413,

June 2022

Certificate

This is to certify that the thesis titled "**DEVELOPMENT OF STRESS INDUCTION AND DETECTION SYSTEM TO STUDY IT'S EFFECT ON BRAIN**" submitted for the award of the Doctor of Philosophy is original to the best of our knowledge. The work was carried out by **Ms. Nishtha Phutela** under our guidance and has not been submitted in parts or full to this or any other university for award of any degree or diploma. All the assistance and help received during the course of study has been duly acknowledged.

Nishtha Phutela

Supervisors

Dr. Devanjali Relan
Department of Computer Science & Engineering
BML Munjal University, Gurugram

Prof. Goldie Gabrani
VIPS Technical Campus
College of Engineering, New Delhi

Prof. Ponnurangam Kumaraguru
Department of Computer Science, IIIT Gachibowli Hyderabad

Abstract

Stress has become a significant mental health problem of the 21st century. The number of people suffering from stress is increasing rapidly. Thus, easy-to-use, inexpensive, and accurate biomarkers are needed to detect stress during its inception. Early detection of stress-related diseases allows people to access healthcare services. This thesis focuses on the development of stress stimuli and the detection of stress induced by these stimuli. Identifying brain regions affected while exposing the subject to these stressful stimuli has also been done. Three different stimuli, viz. videos, gamified application, and a game, are investigated to study their effect as stress induction stimuli. To this end, in this thesis, a system is proposed to classify participants into stressed and non-stressed categories using machine learning, deep learning, and statistical techniques. The statistical significance between stressed and non-stressed was found using Higuchi Fractal Dimensions (HFD) feature extracted from EEG. This feature also helped identify the brain's most affected region due to stress. Another outcome of this thesis is the extra annotation of the ground truth which further helps to validate the participant's experience under the influence of stressful stimuli. This annotation was performed by evaluating participant performance under time pressure. In addition, a technique based on in-game analytics is presented to complement the betterment of self-reported data. Further, another dimension utilizing signatures from WiFi Media Access Control (MAC) layer traffic is presented to detect stress indicators in a device-agnostic way.

Dedication

To,

The Mystic Force and the Divine Powers,

Shri Guru Maharaj Ji and Shri Krishna Ji

My parents who always believe in me,

Mrs. Lakshmi Phutela and Mr. Shyam Lal Phutela

My parents-in-law without whom this would not have been possible,

Mrs. Ranjna Bahl and Mr. Ashok Kumar Ganda

My spouse for his care, advice, patience and support,

Mr. Akshay Arora

And my baby, for taking my stressors away with his touch,

Nirvan.

Acknowledgement

First of all, I am thankful to the Almighty for giving me immense strength and determination to complete this thesis work. I take this opportunity to thank all the fantastic people who supported me in this challenging yet beautiful Ph.D. journey.

I am deeply grateful to Dr. Devanjali Relan (Associate Professor, BML Munjal University) for her guidance and valuable suggestions in my research work. She always motivated me to look at finer details and helped me improve my technical writing and presentation skills by giving consistent feedback. The brainstorming sessions with her were very productive and full of enthusiasm. I had a wonderful experience learning from her insightful comments.

I also like to express my sincere gratitude to Prof. Goldie Gabrani (Professor, VIPS Technical Campus, New Delhi, and former Professor at Delhi College of Engineering and BML Munjal University). I am indebted to her for her enlightening support and motivation throughout my Ph.D. tenure. I got life lessons from her, and the power of her wisdom always amazes me. Somehow she always knew what was right for me. She always believed in me and gave me the strength to overcome all the challenging periods during my Ph.D. She taught me to understand and analyze things holistically. In view of this, I am highly grateful and appreciative of the mystic force that helped me to tap my inner potential and overcome all the hardships faced during my Ph.D. work joyfully. I thank Prof. Goldie for always being with me and helping me cope with work pressures.

I would like to extend my sincere thanks to Prof. Ponnurangam Kumaraguru (IIIT Hyderabad, former Professor IIIT Delhi) for his selfless help. Despite his hectic schedule, Prof. Ponnurangam was always there to mentor me. All his suggestions were thoughtful. I especially thank him for connecting me to the students from 'Precog', his research lab. After some of the brainstorming sessions with his Precog students, I understood the importance of peer evaluation. Prof. Ponnurangam is the person who introduced the Human-Computer Interaction (HCI) subject to me at IIIT Delhi, which has now become one of my favorite research areas.

My Ph.D. research also benefited from the insightful comments from my research committee members from BML Munjal University, consisting of Prof. Jaskiran Arora, Dr. Maheshwar Dwivedy, and Dr. Kiran Sharma.

I am immensely thankful and indebted to Dr. Dheryta Jaisinghani (University of Northern Iowa, Cedar Falls, Iowa) for her invaluable advice. She co-guided me during the time of my masters degree at IIIT Delhi and mentored me at various stages of my Ph.D. journey as well. She helped me to connect the *dangling pointers* in my work on game based approaches and guided me in the work on device-agnostic techniques for detecting stress indicators.

I am profoundly grateful to the President of BML Munjal University, Ms. Swati Munjal, Vice-Chancellor Prof. Manoj K Arora, Dean of School of Engineering and Technology and Dean Research, Prof. Anirban Chakraborti, Registrar, Mr. Abhay Sharma, Controller of Examinations, Ms. Suneet Soni, Assistant Dean Doctoral & Post-Doctoral Programmes, Dr. Sambhunath Bera and the entire management team of BML Munjal University for their unconditional support. I also take this opportunity to thank Dr. Vishal Baloria (Assistant Professor, BMU), Dr. A.K Prasada Rao (Professor, BMU), Dr. Tabish Rasheed (Associate Professor, BMU), Dr. Suchitra Rajput Chauhan (Assistant Professor, BMU), Mr. Akash Agnihotri (Senior Lab Technician) for providing timely and speedy administrative help. I also want to acknowledge Dr. Akhlaq Husain (Associate Professor, BMU) and Dr. Arpit Bhardwaj (Associate Professor, BMU) for providing guidance and support whenever needed. I also thank all management and administrative offices at BML Munjal University for always being kind and available to address all my queries. Sincere thanks to BML Munjal University authorities for providing me necessary facilities for the smooth completion of my work.

I thank the examiners of my thesis and the anonymous reviewers of the published papers, whose comments helped me to significantly improve this work.

On a personal front, I significantly extend appreciation to my beloved son *Nirvan* for being very supportive. The precious moments I would have spent cherishing his childhood were spent doing Ph.D. work. I also thank my parents-in-law, my husband and my house help *Rajni* for taking care of my son while I was out for work. I could not imagine completing this work without their support.

Nishtha Phutela

Contents

List of Figures

List of Tables

List of Abbreviations

CSI Channel State Information

CSV Comma-Separated-Value

CWMT Color Word and Memory Test

DNN Deep Neural Network

DSM Diagnostic and Statistical Manual of Mental Disorders

EDA Electrodermal Activity

EEG Electroencephalography

GSR Galvanic Skin Response

HCI Human Computer Interaction

HFD Higuchi Fractal Dimension

IoT Internet of Things

LSTM Long Short Term Memory

ML Machine Learning

MEEGA+ Method for the Evaluation of Educational Games for Computing Education

MLP Multi Layer Perceptron

MRI Magnetic Resonance Imaging

PHRs Patient Health Records

PHQ Patient Health Questionnaire

POC Proof of Concept

PSS Perceived Stress Scale

QIDS Quick Inventory of Depressive Symptomatology

RSSI Received Signal Strength Indicator

SAM Self Assessment Manikin

SCWT Stroop Color Word Test

STAI State Trait Anxiety Inventory

WESAD Wearable Stress and Affect Detection

WHO World Health Organisation

Chapter 1

Introduction

1.1 Motivation

Mental health is a state of well-being where every individual can realize his potential, cope with the everyday stresses of life, work productively, and contribute to his family and community. If any of these conditions are not being met, the individual may have a mental illness. The World Health Organisation (WHO) predicts that by the year 2030, mental illnesses will be the leading disease and will burden society globally. Approximately 15.5% of the global population is affected by mental illnesses, and those numbers are rising. Stress is also a type of mental illness that can badly affect an individual's health. It is one of the most common problems in the western world and is increasing in the middle-class population in India due to the adoption of the western lifestyle. In today's world, work and occupation-related stress are increasing day by day. Moreover, the job that demands to multitask is another major cause of stress [1].

According to the American Institute of Stress, around 73-77% population experience stress that affects not only the physical health but also the mental well being. Further, around 48% of people are suffering from sleep disorder due to stress [2]. Recent survey on LinkedIn showed that 40% of working Indian professionals experience increased stress or anxiety [3]. The survey also showed that 36% of them feel that stress is adversely impacting their work-life balance .

Traditionally, stress was analysed only by medical personnel without the use of any technology. Medical staff trained in mental health used to perform psychotherapies which involved face to face interaction with the people facing mental health concerns. Advances in computing technology have created opportunities for close collaboration between computer engineers and medical practitioners studying mental health. With the plethora of sensing devices now being available, emerging technologies like Human Computer Interaction (HCI), Internet of Things (IoT), and Machine Learning (ML) have started to emerge as technologies with capabilities to develop applications that help people with their stress-related mental health problems [4, 5, 6]. These technologies have become an umbrella to offer new opportunities for screening and predicting stress-related mental health problems. Coupled with the power of data science, these can transform the way technology can be used to identify and treat people who have stress-related mental illnesses.

Stress can be triggered by the change in the body's emotional response to various situations such as depression, anxiety, anger, grief, guilt, low self-esteem, or performance pressure. Stress is the root cause for a variety of mental health problems like depression, dementia and has an adverse effect on a person's performance [7]. Issues related to stress are rising exponentially worldwide therefore, the detection and quantification of stress are of utmost importance [8].

There are different ways to detect stress. Traditionally, the stress can be detected only through self-reports [9] in the form of questionnaires or interviews taken by psychologists. Some standard questionnaires are available, and the answers filled by subjects to those questionnaires are mapped to some predefined scales. Each question is assigned some score based on the answer given, and the total score is calculated from all the questions answered [10]. Different standard scales are used in clinical settings through which stress can be quantified [11, 12, 13] but they are subjective indicators. Moreover, in developing countries, people do not prefer to go to psychologists and mental health clinics, due to the social stigmas associated with it. Thus, there is a need for a system that can automatically classify the subjects into stress and non-stress. It may also help to develop preventive measures, for instance, to make the people aware about their mental health [14].

It has been observed that stress has a significant effect on specific physiological parameters like skin conductance, blood pressure and brain signals [15, 16]. Various physiological signals acquired from different sources such as Electroencephalography (EEG), Functional Magnetic Resonance Imaging (fMRI), and Positron Emission Tomography (PET) were used for the detection of stress [17, 18, 19]. Among these, EEG has gained acceptance in monitoring stress levels as it is non-invasive [20], non-expensive and gives very high temporal resolution [21]. With recent improvements, wearable EEG devices can quantify human stress by directly collecting central nervous system activities during stress. Potentially, the EEG can play an important role in detecting stress but detection systems apart from EEG should also be explored.

Numerous researchers have quantified stress in two different settings: controlled laboratory settings and natural settings (real-time). For instance, in the laboratory setting, stress is induced intentionally, [22, 23]. Whereas, a natural setting is more realistic than a laboratory setting as it is non-intrusive. It does not interfere with the everyday activities of the person. Consequently, stress induction and its detection are different in the above settings.

To this background, it is essential to devise a system to induce the appropriate amount of stress through proper stress elicitation stimuli. This system should consider different techniques to annotate the ground truth to effectively capture the influence of the stimulus at an individual level. Additionally, the system must detect stress in a non-invasive manner and accurately classify an individual as stressed or non-stressed.

1.2 Aims and Objectives

The aim of this thesis is to explore and develop empirical methods to understand the manifestation of stress. In this regard, the major objectives are:

1. To design and use stress elicitation stimuli.

2. To create experimental setup and collect stress related data.

3. To explore and propose techniques to identify stress biomarkers.

4. To classify whether a person is stressed or not stressed, using machine learning and other techniques.

1.3 Significant Contributions

In view of the above objectives, the following are the significant contributions of this thesis:

1. Acquired EEG signals from participants when exposed to two different stimuli: Video and Color Word and Memory Test (CWMT).
2. Developed an innovative stress stimulus called Color Word and Memory Test (CWMT) in the form of gamified mobile application which was inspired by Stroop Color Word Test (SCWT) stress inducer.
3. Examined the effect of different stress stimuli on human mental state and classified participants into stressed and non stressed.
4. Proposed a system to capture the complexity using Higuchi Fractal Dimension (HFD) from EEG signal.
5. Designed and developed instrumented version of the educational game *Unlock Me* to:
   (a) record human-game interactions and validate player experience.
   (b) analyze the difficulty in game progression and improve learning outcomes
6. Created an experimental setup and collected the data set to distinctly identify different stress related human activities: Walking, Sitting, Sleeping, Awake, Using Phone and Not Using Phone. The Media Access Control (MAC) layer signatures were used as features to classify respective activities.

1.4 Outline of Thesis

This thesis is organized as follows: Chapter 2 presents the anatomy of the brain and the technique that was utilized for brain signals measurement i.e EEG. Chapter 3 reviews the state-of-the-art techniques for detecting stress. Chapter

4 presents stress detection using video stimuli. State-of-the-art machine learning and deep learning classifiers were used to differentiate between the stressed and non-stressed emotions elicited upon watching the videos. Chapter 5 presents the statistical analysis performed using fractals to investigate the existence of significant difference between stressed and non-stressed groups. Chapter 6 presents an evaluation study conducted to understand the impact of game based learning to enhance learning experience of the players and to evaluate the playing behavior in response to in-game events. Chapter 7 delves into smartphone based approaches for stress related activity monitoring. This chapter also presents a proof of concept of using device agnostic approach for classifying various human activities. The thesis is concluded in Chapter 8. This chapter also summarizes the significant contributions and the different ways this thesis can be extended in future.

Chapter 2

Background

In this thesis, various experiments were performed using EEG signals which show the brain activity. The subsequent section presents anatomy of the brain, EEG signals and various frequencies associated with it.

2.1 Human Brain

Human brain is the largest and the most complex organ which is made up of 100 billion neurons [24, 25]. It receives and processes various signals from the sense organs. The various brain regions are:

1. Forebrain: In this area, the cerebrum is encountered. The cerebrum is the most significant region in brain. It initiates movement, language, memory and reasoning. The electrodes to measure brain activity record changes in electrical activity from the cortex which is in turn a part of the cerebrum located on its outer side [22, 26, 27, 28, 29].

2. Midbrain: A significant portion of the brainstem is the midbrain. Its the brainstem that connects the forebrain with the spinal cord. Strokes are found to happen in the midbrain [30].

3. Hindbrain: The cerebellum is located in this part. The movements in the cerebellum are coordinated by the hindbrain. The pons are also present in this region. The pons function to relay signals from the cortex to assist in controlling movement. It is also involved with the control of sleep and arousal.

The cerebral cortex is now described in detail since it is the source of recording of brain activity done in this thesis using non-invasive sensors. There are several lobes in this cortex. As shown in Figure 2.1, the cerebral cortex consists of frontal, temporal, occipital, and parietal lobes, with each having its particular function. There is some ground established in the neuroscience community to understand the common tasks associated with each lobe of the brain.

The frontal lobe monitors our personality. Its functions include motor function, problem solving, initiation, judgment, and spontaneity. The frontal lobes contain important asymmetrical differences. For instance, language-related movement is controlled by left frontal lobe, and non-verbal abilities are controlled by right frontal lobe. Some researchers emphasize that this rule is not absolute and that both lobes are involved in nearly all behavior in many people.

There are two functional regions in which the parietal lobes are divided. The first takes cares of sensation, cognition and perception and the second is primarily associated with the visual system and creates a representation of the picture around us.

A damage in the temporal lobe causes language problems. The recognition of words get affected if the damage is on left side of temporal lobe and talking gets affected if their is a damage on right side of temporal lobe. There is also a significant association between temporal lobes and memory skills.

The occipital lobes are the center of our visual perception system. Disorder in the occipital lobe can cause hallucinations.
There are various techniques to decode mental states by measuring brain activity like fMRI, EEG, ECoG [31]. The work presented in this thesis used EEG signals therefore the next section explains the EEG signal and its associated frequency bands.

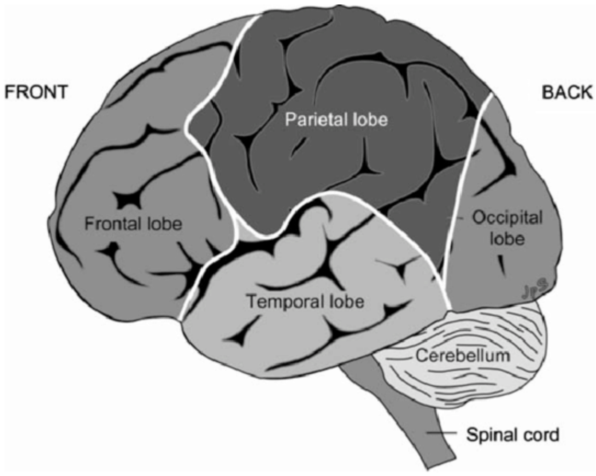


Figure 2.1: Cerebral cortex of brain divided into lobes (Source: [32])

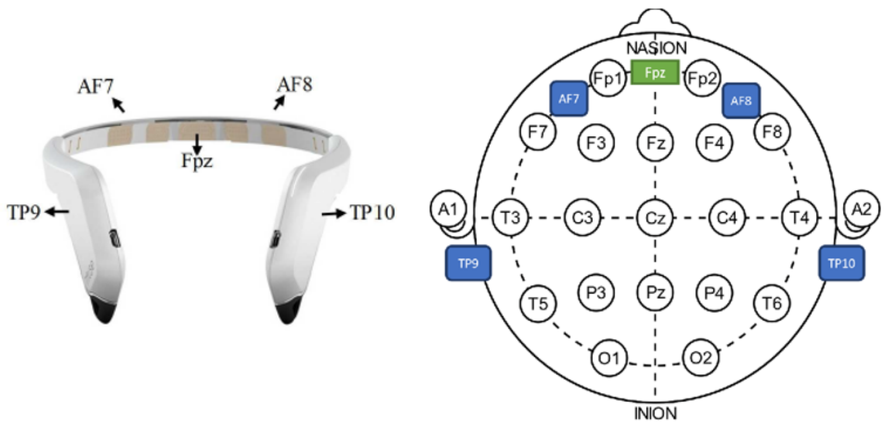


Figure 2.2: (a) MUSE headband to record EEG signal (b) Electrode positioning on head scalp (Source: [23])

2.2 Electroencephalography

The electrical activity happening within ensembles of neurons generate some electric potential. EEG is the technique of recording this potential from the surface of the scalp.

EEG electrodes embedded within EEG devices are used to take measurements from specific parts of the brain. While conducting experiments with EEG, one should choose the position for placing EEG electrodes on the cortex depending on the activity to be studied. This is because each region is associated with a different activity. In order to place these electrodes one needs some standard rule. One such standard is the 10-20 system of electrode placement (refer Figure

2.2). It is internationally accepted and is known as 10-20 system, because the distance between electrodes are between 20% of each other from the whole area of the scalp and it is placed 10% above the nasion and inion which are referred as reference points.

Before the 10-20 system can be used, some reference points should be determined for appropriate placement of electrodes with reference to the skull. These points consist of the nasion, inion, and ears. Nasion is located above the nose on the skull, below the forehead. While inion can be inferred by bony end and marks the transition between skull and neck.

As seen in Figure 2.2, the naming convention in the 10-20 system is determined as per the lobes that the electrodes should be placed on. For instance, the Fp electrode represents pre-frontal lobe which is placed above the eye, F electrodes which represent the frontal lobe, C to denote central part of the cortex, T for temporal lobe, P for parietal lobe, and O for occipital lobe. Additionally, there are three sections in which the scalp is divided. The first is left hemisphere, second is right hemisphere and the third is central area. This division is done using a straight line starting from the front and going towards the back. As a convention, the left section is labelled with odd numbers and the right section with even numbers. A 'z' notation is used for the central area. For instance, the electrodes located in the left side of prefrontal region are named $Fp1$ and the ones present in the right side are named as $Fp2$. The electrodes at the centre of prefrontal region are called Fz.

2.3 Stress Detection From EEG

Stress detection of an individual has been done by many researchers by analysing the waves generated by the brain. These waves generate neural activity in the brain that produces some electrical signals. These electrical signals generate certain waveforms when the electrodes are placed on a person's scalp. The features of these waveforms are shape, amplitude and frequency. There are five classes into which the frequency of the EEG signals is divided. These are delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-32 Hz), and gamma (>32 Hz).

There is a separate state of mind associated with each of these classes [33].

Initially the state of mind is assumed to be neutral. This is also called resting state EEG [34]. In controlled experiments, the state of mind has to be activated by using some stimulus. For instance, this stimulus can be in the form of audio and video signals [23]. These tapes arouse a certain category of emotion in the brain of the participant. It can be a positive emotion or a negative emotion. It has been observed that during negative emotions, the activities in right hemisphere are dominant than those in the left hemisphere [35, 36]. This indicated that the area for stress detection is the right hemisphere. Among the frequencies generated by the EEG waves, alpha and beta bands indicate consciousness, while delta and theta bands characterise unconscious state of mind. Particularly, beta wave frequencies are considered as the characteristic indicator of stress [33].

Various features and classification techniques have been used by researchers for evaluating stress from EEG signals. These techniques use signal processing tools like Fourier transform and wavelet transform. There are different domains in which the EEG signal can be analysed viz. frequency, time and space. The features produced by these signals are frequency band power, peak frequency, cross-correlation among the powers of each band etc. Since the EEG data is a time series data, these features vary over time. On a similar note, the proportion of power spectral densities of alpha and beta bands has also been calculated for the analysis of stress. Stress brings changes in the absolute power of the EEG signal. It also affects some connectivity parameters like coherence [37, 38].

Chapter 3

Related Work

Stress is one of the mental health concerns. It is challenging to quantify and hence difficult to measure. As per Lazarus and Folkman, "Stress is the relationship between the person and a situation, which adversely impacts the happiness and health of the person suffering" [39]. Another definition of stress is "a physiological reaction that aims to protect the individual from possible threats emanating from the environment" [40]. One can infer from these definitions that stress arises from a threatening situation.

Our body activates its defense mechanism to adapt to or overcome stressful circumstances. Upon the disappearance of the stressor, our body returns to regular operation. However, it takes a specific time because stress causes chemical changes in the body [41]. If the body is continuously exposed to these chemical changes, stress can prevent the body from coming to regular operations and thus have long-term health problems. These health ailments range from psychological to cardiovascular problems and are dependent on the duration of stressor [33]. Based on the duration, stress can be transient or prolonged [42]. Transient stress might not affect health but might influence a person's decision-making ability [43] and his memory [44], whereas prolonged stress can have adverse effects on health. Thus, it is important to have a system of stress detection as it helps to detect the stress level timely, which reduces the risk of adverse health consequences [1].

[1]The content in this chapter is revised from my publications: [28, 45, 46, 47, 48]

3.1 Machine Learning Approaches

Different stress stimuli were proposed to induce stress. Also, different stress induction protocols were followed. Using these procedures, different datasets were generated. Various authors used multiple methods using these datasets to classify stress. The accuracy of these methods relies on multiple factors such as sensors that can measure physiological signals, quality of signal and the machine learning model.

In one of the studies, the authors related stress with the circumplex model of affect. This model characterizes several emotions in the domain of arousal and valence [36]. The authors used the dataset for Emotional Analysis using Physiological signals (DEAP) [49] for the detection of stress. They extracted time-based, spectral features from complex non-linear EEG signals. They found that stressed state is associated with reduced asymmetry as compared to non-stressed state. Using coherence analysis, they also found that during the stressed state the activity in the right side of the brain is more than the left.

In another study, the authors used Stroop Color Word Test (SCWT) stressor to induce the stress and used a combination of power based features, fractal dimension(FD) and statistical features for the inter-subject classification [27]. The features were extracted from fourteen electrode EEG headband. The stress level was classified using k-nearest neighbours (k-NN) and support vector machine (SVM). Finally, the 5-fold cross-validation was performed to validate the model. It was found that SVM outperformed k-NN when a combination of statistical and FD features was used. Three levels stress classification achieved an average accuracy of 75.22 % and two levels stress classification resulted in an accuracy of 85.17% using SVM.

In another work, a system was proposed to classify different mental states - relaxing, neutral and concentrating [50]. The authors tested various feature selection algorithms and classifiers. They compared the performance of the proposed system in terms of accuracy and number of features used. A 10-fold cross validation was performed to validate the accuracy of the designed model. The authors summarized optimal 44 features, from a set of 2100 features, required for the stress classification. The designed system resulted in overall accuracy of

87% with random forest classifier.

Many authors have used frequency based features to classify stress. For instance, in [29], the frequency domain features were extracted from EEG recordings. They found SVM as the best classifier, among all the tested classifiers, to classify human stress when used with alpha asymmetry as one of the feature. They also found that alpha asymmetry can be regarded as one of the potential bio-marker for stress classification, when labels are assigned using expert evaluation.

The authors in [51] utilized frequency based features to classify four types of negative emotions using four channel EEG signals. They used movie clips as emotion elicitation material. They tried multiple machine learning algorithms and found that Long Short Term Memory (LSTM) is able to achieve the best accuracy of 92.84% by using 10-fold cross validation.

Recently deep learning has been widely used in the domain of stress recognition through EEG [41, 52]. A detailed review of deep learning techniques for classification tasks using EEG signals is reported in [53]. The advantage of using deep learning is that it can directly take input from raw data and identify the most prominent features automatically without any feature engineering and pre-processing [54, 55]. As a result, the difficulty of selecting the best appropriate pre-processing algorithm and feature selection methods has been overcome, making this more applicable. But according to [56], although deep neural networks are capable of learning features, it is better to do feature extraction beforehand. This is because EEG signals contain noise and interference. Furthermore a lot of data is needed for training purpose in building a deep learning model.

Stress can be detected in natural setting or controlled lab setting. Under controlled lab setting, various stress elicitation material was used such as evoked emotional stress through multitasking [1], Paced Auditory Serial Addition Test (PASAT) [57] and Stroop Color Word Test (SCWT) [27].

The authors of [41] attempted to record the pattern of workers' brain waves at a construction site (natural setting) when they were under stress. Their aim was the early detection and mitigation of stress for the construction workers. To obtain the ground truth, the saliva of the workers was collected which contains a hormone called cortisol responsible to regulate stress. The authors conducted this study on 9 construction workers using fourteen electrodes mobile EEG device. The intrinsic signal artifacts were removed by using Independent Component Analysis (ICA) and the extrinsic signals artifacts were removed by using low pass filter, high pass filter and notch filter with appropriate frequencies. They proposed a system which utilizes 14 electrodes and build a model using convolutional deep learning network and a fully connected deep neural network architecture for binary stress classification of stress. The accuracy reported by using fully connected DNN was 86.62 %. Their DNN architecture consisted of two hidden layers with eighty three neurons in the first layer and twenty three neurons in the second hidden layer. While their approach reported an average increase in the stress classification accuracy as compared to their previous approach of using SVM [58], the major limitation of their study was the collection of saliva. Collecting saliva might not be favourable to the subjects under study. They have also not specified if a medical practitioner was in their team to label the stress level based on collected saliva sample.

The application of LSTM network for classification of brain signals has been reported by [59, 60, 61]. LSTM with attention mechanism has been used by [62] to develop cross-subject generalised solution for classifying limb movements using EEG. They have used frequency as well as time based features as input to LSTM network and obtained an accuracy of 83.2 %. In [63] the authors reported the highest accuracy of 92.8 % in context of driver stress classification tasks at different weather conditions and other ambient factors, using single physiological signal i.e. electrocardiogram (ECG) signal. They used LSTM and CNN to detect driver's stress.

In [23], the authors record the EEG signals of twenty-seven participants using a four-channel MUSE headband. Responses obtained from the State-Trait Anxiety Inventory (STAI) questionnaire were used as the ground truth. First, frequency domain features (namely absolute power, relative power, coherence, phase lag, and amplitude asymmetry) were extracted. After that, two and three classes classification was performed using four different classifiers (sequential minimal optimization, stochastic gradient descent, multi layer perceptron, and logistic regression). Using the logistic regression technique, these authors obtained a stress classification accuracy of 98.76% and 95.06% for two- and three-level classification.

In another study, authors used a multi-tasking framework to achieve near-real-time induction of stress [1]. These authors used Stroop and Memory test as stress induction stimuli. They used a fourteen-channel EEG device for data collection. A three-item questionnaire was used to elicit the ground truth of stress-induced after each task. Subsequently, band power features were used to classify stress from non-stressed conditions using Support Vector Machines (SVM). The authors reported a stress classification accuracy of 77.53%.

Authors in [64] used a single-channel EEG device for real-time monitoring of stress. They analyzed frontal asymmetry, which is a significant feature of stress assessment. They concluded that machine learning classifiers, when trained on 50% of the dataset, achieved 90% accuracy in classifying EEG of stressed subjects. The stress detection system proposed by the authors in [27] used EEG signals from nine subjects. They used SCWT as a stressor, and EEG data of participants were recorded during the experiment. They performed a four-level, three-level, and two-level classification of stress and achieved maximum accuracy of 85.71% using SVM when a combination of fractal dimensions and statistical features were used for two-level stress classification.

In order to analyze the activity in different brain regions under various mental stress conditions, the authors in [65] curate a series of cognitive tasks. These tasks include solving puzzles, making decisions, mathematical computations, pattern matching, and memorization. The authors have used Multifractal Detrended Fluctuation Analysis (MFDFA) to observe these changes. The authors found that these cognitive tasks affect the frontal and parietal lobes of the brain. Additionally, the authors have used a topological scalp map to validate the results from MFDFA.

The fractal dimension has been proved to be an effective tool for finding the significant difference between the pathological and normal groups. For example, authors in [66] hypothesize that HFD can capture region-specific neural changes in patients with Alzheimer's disease (AD). The experiment was conducted with 52 subjects (20 subjects with Alzheimer's disease and 32 from the normal group). The EEG data of the subjects were recorded, and the mean HFD value was calculated for every channel of EEG. P-values were calculated across both groups to identify EEG channels significantly associated with AD. The authors found that HFD values for subjects with AD were significantly lower in the parietal areas. Thus, HFD values captured from these areas can differentiate AD and normal subjects.

Furthermore, in [67] the authors compared two EEG-based analysis methods (linear spectral asymmetry index (SASI) and HFD) for detecting depression . The results indicated that the SASI and HFD methods both demonstrated a good sensitivity for detecting characteristic features of depression in a single-channel EEG.

Authors in [33] performed statistical analysis and inferred that there exists a significant correlation between stress and high beta activity. They found a statistically significant Pearson's correlation coefficient between salivary cortisol levels and SDNN - a metric derived from EEG and Heart Rate Variability (HRV) data for their analysis. These authors used two physiological signals to infer stress, but this thesis work proves that only one physiological signal, i.e., EEG, can also be used to identify a biomarker for stress.

A stressed emotion dataset called Multimodal Dataset of Stressed Emotion (MuSE) - has been presented in [52] to study the correlation between occurrence of stress and the presence of affect. They have considered stress as one of the confounding factors in influencing the psychological state of a person. The authors collected data from 28 students during the final exams and after the exams period to create datasets for stress and non-stress respectively. The experiment comprised of a series of the following events: emotional stimuli presentation, video watching and emotionally evocative monologues. Perceived Stress Scale (PSS) was used to get self reported scores of stress and Self Assessment Manikins (SAM) was used for emotional assessment. They used a paired t-test to infer that the average PSS scores obtained from the two groups were significantly different. A considerably large recording of forty-five minutes was used. The novelty in their experiment design was the use of different emotional elicitation materials across all sessions, even though the emotional dimension being captured was same. They have used various unimodal deep neural networks and also a multimodal ensemble for modelling valence and activation [68]. They have segmented the video dataset with a window size of 1 second and an overlap of 0.5 seconds. Out of the diverse set of features that have been used in their work, the visual and physiological modalities perform the best for stress elicitation while being influenced emotionally. The reported accuracy from their work is 70%.

Other datasets available for stress detection include [69], it is a real world biometric dataset collected from nurses working in a hospital at the time of COVID-19. The physiological variables measured in this dataset include EDA, heart rate, GSR and accelerometer reading.
Wearable Stress and Affect Detection (WESAD) is another publicly available dataset collected in a controlled lab set up. It contains physiological and motion related data of 15 participants [70]. Various machine learning algorithms have been used to differentiate stressed, neutral and amused emotion. Authors in [71] have introduced another large scale dataset of stress using physiological signals. [72] is a multi domain social media dataset for identifying stressed state of an individual.

As mentioned above, various authors used different dataset, features and techniques to classify participant's data into stress and non-stress groups. Every technique has its own advantage and disadvantage. The work presented in [22, 73] gives the classification accuracy greater than 90% but they lack extra annotation of ground truth whereas in the system proposed in this thesis this advantage has been overcome.

3.2 Technological Approaches

Since stress has an effect on mental wellbeing of a person, a background on different mental health issues and their solutions proposed by various researchers is important.

Worldwide today, problems related to mental health are on an exponential rise and are revealed in a comprehensive manner that includes cognitive capabilities, biological processes, emotional factors, behavioral patterns, and social behavior [6]. As of today, no reliable tools are available to measure these parameters for the analysis of mental disorders. Usually, the assessment of the degree of mental health problems is greatly based on patients' self-disclosed information compared to the diagnostic measures cited in the Diagnostic and Statistical Guidebook of mental disorders. In addition, The Patient Health Records (PHRs) are essential sources for analyzing mental health status. Apart from PHRs, another two significant data sources are in the form of (i) structured records of demographics and prescription notes and (ii) unstructured records consisting of medical transcriptions, DNA sequences & images of the brain produced by MRI technologies. Both of these are present with health care providers. However, with a wide variety of digital tools and techniques in the healthcare sector, the traditional ways of handling mental disorders are undergoing a massive revolution. It has become possible to quantify the individual's characteristics by analyzing the data from the digital objects that one is carrying or wearing. It includes devices like smartphones [74] and wearable sensors [75]. By mining the data gathered through these devices, the mental well-being of a particular individual can be studied. But, the mapping of the signals acquired from these devices to mental illness symptoms is very difficult.

Traditionally, mental health research relied on self-reporting tools like PHQ9, QIDS, DSM-5, etc., which came under active data collection methods. These tools had scalability issues and also faced recall bias. Hence there has been a paradigm shift to focus on passive [76, 77, 78, 79] and heterogeneous [80, 81, 82, 83] methods of data collection. Some of these methods rely on monitoring data from users' smartphones and self-report data. Many researchers have applied the same for understanding the behavior of individuals.

There is a social stigma associated with mental health matters. It makes some people fear judgment if they admit to friends that they need help or if they take the step of scheduling an appointment with a counselor [84]. This challenge makes the screening of mental illness difficult at an individual level. Passive methods cited above can help in this regard.

Harari et al. [85] have also used smartphone-based sensing methods to track the everyday behaviors of the students. They have used an accelerometer and microphone to infer an individual's activity level and sociability across a ten-week duration. They have claimed that their results can be used to predict individuals' activity and sociability trajectories and can be used to design interventions.

A widely discussed study of mental health is by Wang et al. [74] who have assessed an individual's mental health and behavior through smartphones.
Another study by Wang et al. [86] proposed using a person's context in predicting their mental health. A user's context includes activity, body posture, and location data. They also rely on vision-based methods to get information to recognize the user context and compare it with sensor-based methods. Their results show that depression is correlated with the body posture of lying down, so methods should be devised to understand an individual's posture. To avoid blurred images, they combined the accelerometer sensor to detect motion and determine if the picture should be discarded according to the exposure parameters.

A subsequent work by Wang et al. [74] captured the depression dynamics of individuals using DSM-5 [87] for pre and post-ground truth collection. They identified a set of symptoms from the daily mobile and wearable usage patterns of mobile phones and wearable devices users, which could be used as indicators for depression. Researchers have referred to these as digital biomarkers.

As pointed out by Leong et al., [88], a significant component of getting high-quality data is to ensure that the participants feel authoritative about their data. The same has been the focus of Rooksby et al. [84] who have tried to understand user perceptions about collecting their data. Rooksby et al. have analyzed students' perception in sharing their smartphones and wearable data to mental health researchers. As pointed out by Rooksby et al. , work to date in this area has focused on several issues, including monitoring people already diagnosed with mental health issues such as schizophrenia and depression [89], and monitoring general populations for signs of depression [90, 91] or examining mood [92].
Other work has looked at specific populations and contexts, such as looking at workplace stress [93], construction worker's stress [41], firefighters stress [94] and

student's stress [74].

Much of the current work has been of limited scale and primarily for research. However, wide-scale monitoring is envisioned in this area, such as the systematic, population-scale data collections. Suhara et al., through their study DeepMood [92] have accounted for the time of the day and week of the day in the forecasting of depressed behaviors. They have used deep learning models to forecast this behavior. Another body of study has focused on using social media networks for mental health assessment [95] and forecasting [96]. Researchers from the World Well-Being Project (WWBP) analyzed social media with an AI algorithm to pick out linguistic cues that might predict depression [97]. They found that people suffering from depression express themselves on social media differently than those who are not suffering from depression. They were also able to forecast the onset of depression in people through these linguistic markers on social media.

Liang et al. [6] have proposed a framework using big data tools and techniques to address facets like emotion, cognition, behaviour, sociality, genetic, and biological patterns aspects. This framework has discovered mental health through four layers namely: conceptual, sensing, computation layer and application layer. One of the focuses of the work done by Liang et al. is the analysis of human behaviour. Behaviour expression analysis has been an important part of various tests to infer mental health as indicated by [98] who explain that collecting behavioural data in mental health allows for a more comprehensive view of the mental illness which is beyond one-time mental assessment through clinical methods. Traditionally mental health research used to rely on self-reporting tools like PHQ9, QIDS, DSM-5 etc. which came under the ambit of active methods of data collection. These tools had scalability issues and also faced recall bias. Hence there has been paradigm shift to focus on passive [76, 77, 78, 79] and heterogeneous [80, 81, 82, 83] methods of data collection. Some of these methods rely on continuously monitored data from user's smartphones along with the self-report data. The same has been applied by many researchers for understanding the behaviour of individuals.

Harari et al. [85] have also used smartphone based sensing methods to track everyday behaviours of the students. But the feasibility of these methods in correctly monitoring and predicting user behaviours remains a challenge. They have used accelerometer and microphone to infer an individual's activity level and his sociability respectively across ten week duration. They have claimed that their results can be used to predict the activity and sociability trajectories of individuals and can be used to design interventions.

As shown in Table 3.1, authors in [98] measured behavior through smartphone sensors, namely microphone, and accelerometer. While these authors could monitor behavior continuously, the only drawback was that users had to install their application on their mobiles, which most users are reluctant to do for possible security reasons. Authors in [74] also focused on continuous and passive sensing to segregate users into depressed and non-depressed groups, but users had to install their application. Work done in [99] used the completely non-intrusive technique of using mac layer traffic. Authors in [100] measured classroom behavior using wifi which is a non-invasive and cost-effective approach but lacked having ground truth to annotate the data. Another study used wifi infrastructure and levels of analysis but lacked modality [79]. The authors of [83] used a background application based on GPS to infer the severity of depression of the users. However, the drawback for such method is that the energy consumption of GPS is high, thus draining most of the users' batteries. Additionally, this system will not work if GPS is disabled. Thus, authors of [101] tried a better method to create the user's location trajectory to check the places visited by the user. They used GPS data, and when this data was unavailable, they used wifi data to fill it.

Reference	Sources of data collection	Measured factor	Indicators	Technique
[98]	Accelerometer and Microphone	Behavior	1. Physical activity 2.Loneliness	Statistical (SEM), Longitudinal analysis
[74]	PHQ-8, PHQ-4, Microsoft band, light, GPS, accelerometer, microphone, phone lock/unlock	Behavior, Depression	1. Sleep changes 2. Concentration 3. Diminished interest in activities 4. Heart rate variability	ANNOVA, regression analysis, longitudinal analysis
[99]	WiFi	Behavior	1. Social interaction 2. Proximity relationship 3. Occupancy pattern	Probe request, Null data frames
[100]	WiFi	Educational behavior	1. Attendance 2. Late arrival 3. Early departure	RSSI fingerprinting
[79]	PHQ-9, QIDS, WiFi	Depression	Location	Multi feature regression, WiFi infrastructure
[102]	WiFi	Behavior	Location, SMS and call usage, app usage	Heuristics
[83]	GPS, accelerometer, Fitbit and QIDS	Depression, Behavior	Location, activity, heart rate and sleep quality	Multi-task learning, classification, regression, merge android and iphone features

Table 3.1: Overview of different smartphone based approaches to measure user's mental health related behavior.

The work in stress detection is now focused on measuring real-life stress indicators in contrast to lab experiments. Smartphones, the most pervasive devices, have significantly contributed to real-life stress detection [103]. Numerous works have used smartphones to obtain stress-related digital markers [104]. Research efforts in this direction utilize smartphone sensors like microphones, accelerometers, and GPS. Many studies analyzed users' stress-related behavior [105, 106] through smartphone usage patterns, location trajectories, and camera-based monitoring. Another parameter that plays a vital role in differentiating the type of stress a person is experiencing is the context [107, 108, 109]. In addition to the above solutions, a parallel thread of research uses WiFi based solutions. The applicability of these solutions in detecting stress and related issues is explained in this thesis.

Authors in [110] presented a stress and depression detection system (StressMon) by using location information from unobtrusively sensed WiFi data. They used this location data to extract movement and patterns of physical group interaction as features. They did not require any action from the user or install any software on the user's device. These features were input to machine learning classifiers. The evaluation of these classifiers was done on a set of 108 student participants through a longitudinal study and performed stress detection and depression detection. They achieved a 96.01% True Positive Rate (TPR), an 80.76% True Negative Rate (TNR), and a 0.97 area under the ROC curve (AUC) score using a 6-day prediction window for stress detection. For depression detection, they achieved 91.21% TPR, 66.71% TNR, and 0.88 AUC using a 15-day window.

Authors in [111] proposed WiFiTrace, a network-centric method for contact tracing to prevent the spread of infectious diseases. This method uses passive WiFi sensing (instead of the traditional Bluetooth-based method for contact tracing) and has no involvement from the client side. These authors use WiFi network logs gathered by enterprise networks to reconstruct the trajectories of connected devices for contact tracing. They also propose a graph algorithm to scale up their system to trace tens of thousands of users. The graph-based approach outperforms the standard PostgresSQL in memory by 4.5 times.

Authors in [112] tried to identify individuals who behave differently than their groups using passively collected WiFi data. The study participants were 62 college students, and the authors collected the whole semester's WiFi data to identify mobility patterns. Ground truth of group information was obtained using self-reports. Using classification algorithms, these authors achieved: 80% True Positive Rate (TPR), 73% True negative rate (TNR), and 78% Accuracy (ACC). They proposed a system that can help distinguish students who are more likely to struggle with negative workgroup appraisals and enable interventions to improve their overall team experience.

Authors in [113] developed a system *Serene* to capture respiration rate and body motions. It is based on WiFi Channel State Information (CSI). CSI based methods are dependent on location of wifi transmitter and receiver. The authors analyze the impact of respiration on the multipath components of WiFi signals to quantify the effect of small breathing movements on the CSI signals. They take sleep monitoring as a case study, where they collected more than 550 hours (80 nights) of data from 5 users at their respective apartments in real-world full-night sleep monitoring settings. They try to overcome drawback of CSI based methods by using line-of-sight (LOS) as well as non LOS scenarios in their experimental setup.

Authors in [114] develop a system called *Sleep Hunter* to detect sleep stages: Wake, Light Sleep and REM. They obtain a stress classification accuracy of 64% to differentiate between these three stages. Sleep Hunter is a system that uses the smartphone's microphone, accelerometer, light sensor, etc. to collect data related to environmental disturbances (light, noise, etc.) and events during sleep (like movements, cough, snore, etc.) to detect three sleep stages (e.g., REM, deep, and light sleep). These authors use specific feature-extraction methods for each of sleep-related events identified above based on their physical characteristics. These authors also compare their approach against actigraphy and found comparable results. Sleep Hunter was made for Android platform and the authors conducted evaluation experiments on 15 participants from various age groups.

Thus, to overcome drawbacks of previous studies, this thesis presents a proof of concept for monitoring stress-related behavior of the person through non-intrusive methods and utilizing device-agnostic approach.

3.3 Game Based Approaches

Worldwide, stress-related problems are on an exponential rise and affect a person's cognitive abilities. One such cognitive ability is decision making under stressful circumstances [115, 116, 94]. In order to accommodate stress detection during such circumstances, one has to devise stressors that have high external validity [57]. Cognitive effort and time pressure are two such stressors as one frequently encounters these in stressful situations. In order to mimic a stressful scenario and understand the person's responses during such scenario, previous studies had utilized serious games as stress elicitation material [117]. In these studies, stress was indirectly measured by the appropriateness of decisions a person makes during gameplay.

Authors of [117] were pioneers in introducing game-based digital biomarkers for modeling a person's mental health. Further, in one of their works [118], social anxiety was measured using game-based digital biomarkers. They hypothesized that a person's behavior in real life might also manifest in the virtual world. However, their study happens in a lab setting and requires a Virtual Reality headset. Though this headset improves player immersion in the game but drifts away from the goal of assessing stress in a scalable and cost-effective way.

In order to overcome this problem, the technique of in-game analytics [119] was utilized in this thesis. This technique was used to understand in-game player interactions and improve the learning outcomes of an educational game *Unlock Me*. This technique was used to understand in-game player interactions and develop an effective educational game *Unlock Me*. A game is considered effective if it imparts the desired learning outcomes in an easy-to-understand manner. In-game analytics was used for balancing [120] *Unlock Me* to improve learning outcomes and validate players' self-reports about gameplay experience. A series of stressors were provided in the game, and how the players responded to these stressors was recorded through in-game analytics to measure each player's experience and the overall game's effectiveness.

Since a part of this thesis is dedicated to improving learning outcomes through

game-based learning, the background and the previous research in this area is presented next.
The education and health sector has been using game-based learning since the early 2000s. Using games in learning has also proven to be engaging and enable intrinsic motivation among users to learn effectively [121]. Games have elements like fun, goals, rewards, and problem solving that aid learning. The study by [122] states that the users could obtain long-lasting knowledge about healthy lifestyle management. According to a literature review of 512 games, 83% of the 47 knowledge acquisition games were found to be impactful and provided the desired learning outcomes among the users [123]. A comparative study between game-based and non-game-based learners showed that the ones who played competition driven circuit games performed better than the ones playing the non-competitive circuit game [124]. Competitive features like scoreboard, bonus and incentives have proven to motivate and foster deep learning. Learning by doing methodology in educational games helps develop knowledge and positive behaviour in users [125]. Serious games have played a significant role in learning, lifestyle management, and disease awareness. A meta-analysis of the use of serious games for lifestyle management has indicated that serious games positively impact the users [122].

Following the widespread use of serious games in learning, many serious games have been developed. Authors in [126] developed a modified version of the Mario brothers' game to help children learn about Type-1 diabetes and control their blood sugar levels. Another game has been proposed in [127] to help children with cystic fibrosis to learn self management therapies. The game is controlled by a spirometer that measures the breath flow. The intensity of the challenges allows players to use the device and hold their breath efficiently. For helping autistic patients, a game called KickAss was developed [128]. The game mainly aims at helping autistic adolescents in improving their social interaction by confronting them with different social situations and helping them learn from the choices they select for those situations. Besides the games mentioned above, many other serious games have been developed for diseases like cancer, diabetes, alzheimer's, cognitive impairment, autism, and obesity [128].

There are also some games designed to spread awareness among the public. Bruno Santos developed the eVision game to educate people around the world about environmental threats [129]. The game uses augmented reality to capture the surroundings and helps users analyze the ecological threats in that area. The use of augmented reality and a constant interactive assistant called snowkin have proved to increase the persuasive ability of the game. Another game was developed to create hurricane safety awareness among the public of Malaysia [130]. The game provides a hurricane safety checklist to the users and asks them to perform the tasks in the game. Confronting the users with real-life situations and tasks keeps them connected, and as a result, they can apply the tasks in their real-life. The game uses a Game Development Life Cycle, which helps in the design and development of the game in an organized manner. But, with just three levels to progress, the game does not provide a sense of challenge and stimuli to the user. When it comes to promoting awareness for a community disease, a game called Dr. Ludens' LSG was developed to educate local populations about two tropical diseases, Visceral Leishmaniasis (VL) and American Cutaneous Leishmaniasis (ACL) [131]. The game uses a rural farm set up to help users relate to their present environment. The players had to perform specific in-game actions like remove infected dogs, clean leaves, and dry water puddles based on instructions. A quiz was also added to gain extra points along with the points collected from the in-game actions. An in-game quiz encourages deeper learning into the disease and improves the player's retain ability over a more extended period. Although a few experts from the domain evaluate the game, a fixed set of parameters and factors is not considered for measuring the game's impact.

In all the games mentioned above, a proper evaluation model has not been used to evaluate the game. Without evaluating the quality of game one cannot be sure about the effectiveness of the game in imparting the desired learning outcomes. In this regard, Garris presented an input-process-output model of instructional or learning game [132]. A game cycle that included the input-game characteristics and rules, user judgments, behavior, and learning outcomes was followed. Garris considered different dimensions like fantasy, challenge, mystery, rules and goals, and sensory stimuli to describe the game characteristics. User judgments include the level of immersion, interactivity, interest, and confidence, while user behavior has concentration and sustained involvement in the game. User feedback was also collected in the process to support performance. The learning outcomes of the cycle mentioned above are skill-based, procedural, declarative, or affective, depending on the game's goal.

Later, a few other frameworks have also come into existence. Authors in [133] developed a four-dimensional framework to help design an efficient game so that the players could learn without any hindrances. This four-dimensional framework consists of context, representation, learner specification, and pedagogy used. Context mainly consists of place, availability, and tech access to the users and considers the game environment collectively. Learner Specification focuses on the user: their age, demographic, technical expertise, skills, and preferences. Representation includes fidelity, level of immersion, interactivity, user profile, and learning outcome. Pedagogy represents the teaching approach used, be it situative, associative, or cognitive. Although this framework was developed to help design a serious game, it can evaluate a game with the same parameters.

Another game-based learning evaluation model was created in 2015 [134], which is a modified form of input-process-output model by [132]. The input consists of the personal attributes of the player like age, personality, experience, cognitive abilities, and demography. The process consists of design indicators, learning indicators, and environmental influences. The design indicators further consist of usability, challenge, control, and story of the game. Learning indicators include self-efficacy, motivation, engagement, self-directedness, learning activity, and mental effort. The output phase consists of the learning outcomes of the user in terms of knowledge gained, scores obtained, performance, and behavioral changes. The latest model, called the MEEGA+ evaluation model, aims at finding the quality of educational games [135]. It consists of a questionnaire to evaluate factors like usability, player experience, and learning outcomes. These factors further had different dimensions and questions related to them. For example, the usability factor has dimensions like aesthetics, learnability, operability, accessibility, and user error protection. The MEEGA+ model is used in this thesis for base evaluation since it is latest and reliable.

Chapter 4

Video Stimulus Based Stress Classification Using Machine Learning Approaches

4.1 Introduction

Stress, either physical or mental, is experienced by almost every person at some point in his lifetime. Stress is one of the leading cause of various diseases and burdens society globally. It badly affects an individual's well-being and can be a cause of mental disorders. Thus, stress-related study is an emerging field. The estimation of stress in the individual helps in stress management before it invades the human mind and body. In the past decade, a lot of attention has been given to the detection and classification of stress. Electroencephalography (EEG) signals are frequently employed in stress detection research as these are inexpensive and noninvasive modality. Thus, this chapter presents stress classification studies by utilizing EEG signals. These studies are based on stress elicitation through videos. For the first study we utilized covid videos and for the second study we utilized movie clips as stress elicitation material [1].

The outbreak of COVID-19 caused severe harm to people's physical and mental health. Due to the grave situations during the pandemic, much negative news surfaced over the Internet, For instance, an exponential increase of reported

[1]This chapter is a revised version of my publications: [28] and [47]

COVID cases, limited medical resources, lack of hospital beds, and numerous rumors. Many researchers have shown an increase in stress and related mental health problems during COVID-19 [136]. Thus, this motivates to consider COVID news as a stimulus for stress.

Some of the video content providers over the internet release negative information during video watching in order to contribute to creating addiction among youngsters. This addictive behavior leads to a lack of desire to study and degraded performance. The desire to pursue life goals diminishes and a feeling of self-doubt and inferiority complex gains momentum. This further causes stress, anxiety and depression [137]. Additionally, the children's brain receives significantly less amount of dopamine when stress and fear-inducing videos are watched. This hormone dopamine is essential for children because it causes the body to promote reinforcement - the desire to do the task in hand and perform well because of the associated reward [138]. Such stressful videos have been used in various studies in past [139] and this motivates to use such stimulus in present study.

The proposed methodology and the results for first study with COVID news as stress stimulus are given in Sections 4.2.1 and 4.3.1 respectively whereas for second study using movie clips as stress stimulus are given in Sections 4.2.2 and 4.3.2 respectively. For both the studies four groups of features (five PSD features for each of the four electrode positions) were extracted from EEG signals acquired from participants. These features were then used to classify participants into stress and non stress using different machine learning and deep learning techniques.

4.2 Methodology

Various steps involved in the proposed system for stress classification using EEG signals consists of: inducing stress, EEG data acquisition, pre-processing, feature extraction, and classification (see Fig. 4.1).

The EEG signals of subjects were acquired using a four channel MUSE EEG headband in response to COVID-19 news stress stimulus. The MUSE headband is an off-the shelf non-clinical device for capturing the brain signals (see Fig 2.2).

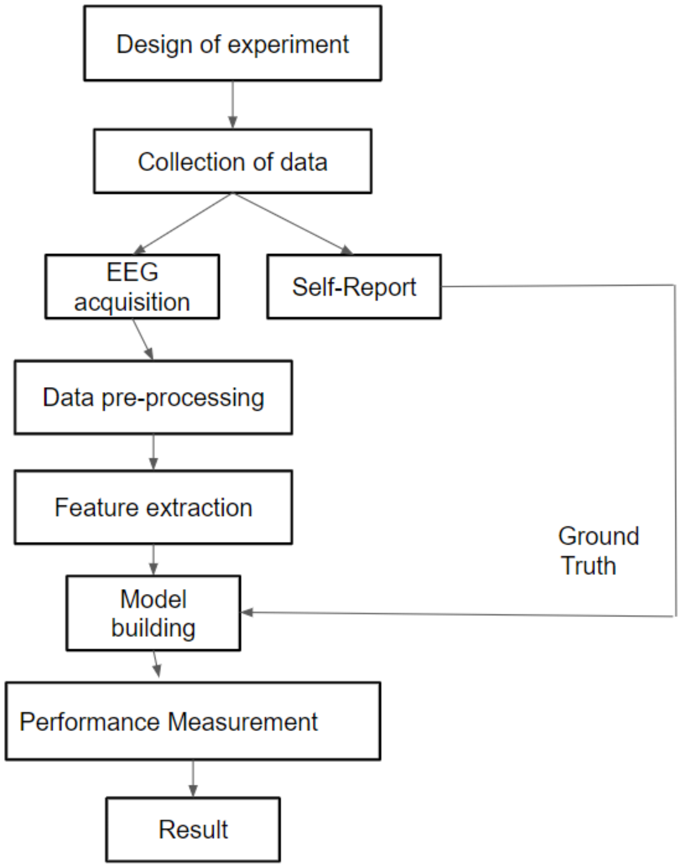


Figure 4.1: Workflow for stress detection using video stimulus

This device contains 4 sensors: TP9, AF7, AF8 and TP10. These sensors are in turn dry electrodes that have been placed according to the 10-20 system of electrode placement (Fig 2.2). The device produces raw as well as pre-processed FFT signals. These signals can be transferred over bluetooth from the device to an android application called MUSE monitor. This application can store the signals in csv format and transfer to the laptop for further processing.

4.2.1 Video Stimulus I

In order to elicitate the emotion of stress, two kinds of video content were taken during this experiment:a) Stressful video content and b) Relaxing video content. The stressful videos were those containing covid related news showing the number of increasing covid cases and deaths and the severity of the disease. The relaxing videos had comedy scenes that would relax the subject. Each video was of a duration three minutes. A gap of two minutes was given between each video clip to avoid the interference of stressed feeling on non stressed feeling and vice versa.

This stimuli is chosen to target two classes of stress: stressed and non-stressed.

A total of 20 healthy subjects, in the age group of 18-30 years, (8 males and 12 females) voluntarily participated in this study. The data from 2 subjects were dropped because two sensors got disconnected in the middle of the experiment. Therefore we performed analysis on the EEG data of 18 subjects. The procedure and protocol was explained thoroughly and consent form was taken from each of the participants. The experiment was conducted according to the principle of Helsinki. After watching each clip a self assessment form was filled by the subjects to rate their experience about the video shown on a five point scale (0-Non stressful at all, 5-lot stressful). A rating of greater than 3 was regarded as stressful. The labels provided by them was considered as ground truth. Since stress also depends on the perception of an individual, these self-reported labels were compared with the labels that were set for the videos. If the labels for a specific subject did not match, its data was removed from our analysis. But the self-reported labels of all the subjects matched with the pre-rated labels. This served as validation for ground truth.

4.2.2 Video Stimulus II

In order to elicitate the emotion of stress, four movie clips were used (refer Table 4.1). Two of these clips were stressful and the other two were not. This stimulus was taken to target two classes of stress: stressed and non-stressed.

Table 4.1: Summary of the excerpts from films shown for stress classification.

Category	**Name of the film**	**Duration (sec)**	**Clip Content**
Non-Stressed	3 Idiots	80	The kick of a stillborn child creates amusement among the surrounding people
	Taare Zameen Par	124	A music teacher delights his students with a motivational song
Stressed	Rang De Basanti	117	The nation mourns during the cremation of a warrior
	Kal Ho Na Ho	96	Friends converse with another friend who is about to die

The data used in this experiment has been collected from participants wearing the Muse headband as shown in Figure 2.2. The headband was adjusted to the comfort of the participant. Subsequently, they were shown the movie clips. EEG signals from 40 subjects were acquired and out of these the data from 5 subjects were corrupted because the connection between electrodes and scalp was loose. This was identified during the manual inspection as the signal has NAN value for these subjects. Finally, 35 subjects were selected (18 males and 17 females) from the age group between 23-55 years. All the subjects were healthy and did not have any kind of neurological disorder. The subjects were instructed not to consume any caffeine product at least 12 hours prior to the start of the experimental process because caffeine is found to interfere with brain activity [140].

While watching the video clip, the participant's EEG signals were recorded. After watching each movie clip, the participant filled an assessment form to infer the level of stress induced by each of these clips. The subjects were asked to complete the State Anxiety questionnaire [13]. State Anxiety is a multi scale questionnaire which we have used to test if the subject has experienced stress after watching the video clip. It has total 20 items. Each of these items is used to infer the feeling of the subject at the current moment. The responses of these questions were taken on a 4-point Likert Scale (1- Not at all stressed, 2-Some what stressed 3-Moderately stressed 4- Very much stressed). The answers from all the respondents were evaluated according to the standard scoring key of State Anxiety scale. The scores of this questionnaire were generally higher after watching the stress inducing videos and lower after watching the non-stressed videos. This data was used as a ground truth to label each instance of EEG recording from the respective person. The procedure followed was in accordance with Helsinki declaration. The records in the data set were labelled as stressed if the score obtained in the State Anxiety scale is greater than or equal to 50 and non-stressed if the score is less than 50.

Authors in [141] showed that stress lies in the top left quadrant of the circumplex model of affect. This quadrant is characterized by high arousal and low valence. The meaning of arousal and valence was explained to the subjects and they were asked to rate the video in terms of arousal and valence. The range of values of arousal and valence for identifying the stressed and non-stressed states were similar to those used by [36]. The data of the participants which did not represent stressed and non-stressed behaviours (in terms of arousal, valence and State anxiety) corresponding to the stressed and non-stressed stimuli was discarded. This was done to ensure that the ground truth of the stress classification model is correct. There were 2 such subjects so their data was discarded. Therefore finally the data of 33 participants was analyzed.

As EEG is a fast and dynamic signal which changes within a short duration of time therefore it is important to process the data in small chunks [142]. Thus a small window size was taken. A smaller window size leads to the model getting trained faster and makes the model more robust and able to capture more information from individual slices of a single sequence of data. LSTM network was used for the binary classification of stress due to its associated advantages. LSTM is used for sequence classification problems and has ability to extract significant temporal information from physiological signals [143], [144]. Moreover, the LSTM network makes predictions based on the individual time steps of the input sequence data.

The dataset was divided into similar length sequences, which is an important step while the training process as input sequences should be of the same length. Thus, the length of the EEG recording for each of the trials was 80 seconds. Each second of the recording has 50 data points. So a total of 4000 data points exist for each of the trials.

During initial analysis only one LSTM layer (LSTM1) was taken containing 8 neurons, gradually a second LSTM layer (LSTM2) containing 16 neurons was included and subsequently another LSTM layer (LSTM3) containing 24 neurons. There was not much difference between the accuracy obtained between the LSTM2 and LSTM3 model. Thus it can be concluded LSTM2 model was sufficient to classify EEG signals as the accuracy does not improve further, which is in accordance with the observation made in [56]. The proposed model is depicted in Figure 4.2.

4.3 Results

This section presents the results of two different experiments with different video stimulus.

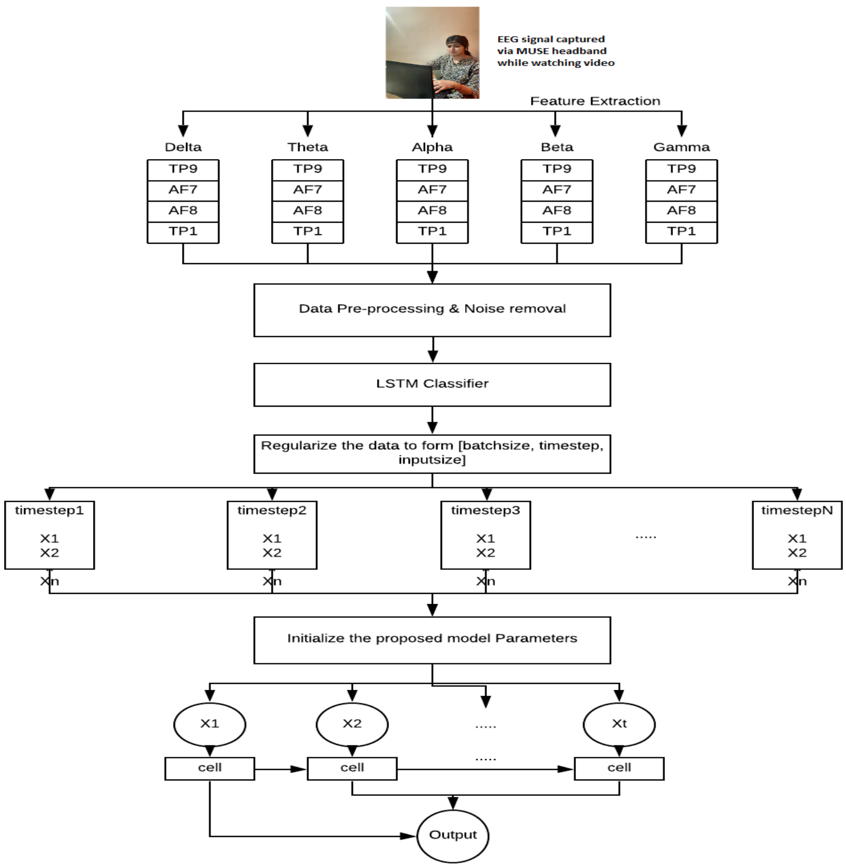


Figure 4.2: Proposed model for stress classification using EEG signals.

4.3.1 Video Stimulus I

In this work, the task of stress classification while watching COVID news has been accomplished. Table 4.2 shows the results obtained from stress classification using

Algorithm		Accuracy	Precision	AUC
Logistic Regression	Min.	77.43	85.76	83.4
	Max.	83.4	96.82	91.56
	Avg.	82.26	87.33	85.75
SVM	Min.	78.88	87.2	91.5
	Max.	88.52	95.17	95.6
	Avg.	85.46	92.26	93.89
Random Forest	Min.	92.85	93.78	92.67
	Max.	96.86	97.67	97.29
	Avg .	94.68	95.55	96.77
Naive Bayes	Min.	76.23	86.26	89.26
	Max.	82.45	91.89	92.57
	Avg.	81.99	84.78	87.64
K-NN	Min.	82.22	86.59	87.20
	Max.	88.43	89.21	90.5
	Avg.	85.91	88.51	89.28
Gradient Boosting	Min	94.43	94.6	93.65
	Max.	97.78	98.29	98.89
	Avg.	95.65	96.54	96.72

Table 4.2: Study 1:Table summarize the results obtained for stress classification using different classifiers

various algorithms. The performance metrics - Accuracy, Precision and Area under the ROC curve (AUC) have been used to compare the results. This table also shows the statistical measure such as: minimum (Min.), maximum (Max.) and average (Avg.) of performance metrics.

For example, the metrics- Min. and Max- denote the minimum and maximum accuracy obtained at a particular fold in 10-fold cross-validation; Avg. denotes the average accuracy obtained from all folds. Results show that the Gradient Boosting algorithm outperformed all other algorithms with an average accuracy of 95.65% as shown in Table 4.2. Moreover, the precision obtained with Gradient boosting is high, and it shows the robustness of the proposed system. High precision shows the percentage of correctly classified instances among the ones classified as stress groups. Furthermore, to reduce the bias of having an already stressed subject in this study, the participants who did not have any causality in their family or immediate family were selected. From the result obtained, it can be concluded that the impact of the stress stimulus was so much that it became possible to differentiate between stressed and non-stressed states with high accuracy.

Method	Validation Method	Specificity	Recall	F1-Score	Precision
MLP	50-50	50.17 ±4.97	79.06 ±2.35	73.17 ±3.35	67.66 ±4.66
LSTM 1	50-50	61.86 ±3.23	84.67 ±4.16	79.17 ±4.73	74.33 ±4.79
LSTM 2	50-50	88.04 ±4.21	86.81 ±4.84	90.04 ±2.11	93.53 ±3.23
MLP	60-40	51.19 ±5.63	82.90 ±3.43	73.83 ±4.43	65.14 ±4.34
LSTM 1	60-40	63.51 ± 3.24	85.41 ± 5.23	80.34 ± 4.39	75.04 ± 5.66
LSTM 2	60-40	86.69 ± 4.44	90.66 ± 4.86	91.68 ± 3.91	82.72 ± 4.27
MLP	70-30	60.60 ± 4.65	79.01 ± 2.28	64.51 ± 3.25	77.26 ± 6.32
LSTM 1	70-30	70.43 ± 4.44	82.16 ± 4.53	83.01 ± 3.93	81.90 ± 6.36
LSTM 2	70-30	84.85 ± 3.72	93.51 ± 3.18	92.45 ± 2.68	91.42 ± 5.23
MLP	10-Fold Cross Validation	61.41 ± 2.23	79.12 ± 3.01	76.81 ± 3.11	76.33 ± 4.69
LSTM 1	10-Fold Cross Validation	71.79 ± 3.65	84.96 ± 4.67	84.62 ± 4.37	82.08 ± 5.53
LSTM 2	10-Fold Cross Validation	88.47 ± 3.42	95.45 ± 2.32	94.94 ± 3.76	94.44 ± 4.43

Table 4.3: Study 2- Performance metrics for stress classification using various classification techniques and training-testing set partitions.

4.3.2 Video Stimulus II

The performance of the stress classification model is measured with the help of the following training-testing data partitions: 50-50, 60-40, 70-30 and 10-fold cross-validation(Refer Table 4.3). This table illustrates the performance metrics used which show the robustness of our proposed approach. Higher recall and precision of our model showed that it gives less false negatives and false positive, respectively. High specificity denotes the true negative rate; i.e., a person classified as nonstressed was actually nonstressed.

The data was split into various proportions for testing and training in all these techniques. 10-fold cross-validation was used to introduce randomisation into the training and test set choice.

Method	**Validation Method**	**Accuracy**		
		Max	Avg	Min
MLP	50-50	70.11	69.33	68.56
LSTM 1	50-50	83.33	81.73	80.14
LSTM 2	50-50	87.22	85.79	84.36
MLP	60-40	73.48	71.80	70.12
LSTM 1	60-40	84.69	82.75	80.76
LSTM 2	60-40	89.28	88.22	87.16
MLP	70-30	75.71	73.77	71.84
LSTM 1	70-30	85.82	83.14	81.87
LSTM 2	70-30	90.33	91.11	91.89
MLP	10-Fold Cross Validation	76.27	72.01	70.59
LSTM 1	10-Fold Cross Validation	87.61	86.55	85.97
LSTM 2	10-Fold Cross Validation	93.17	91.96	90.76

Table 4.4: Classification accuracy comparison for stress classification. The table shows maximum (Max), average (avg.) and minimum (Min) stress classification accuracy obtained with different methods.

As depicted in Table 4.4, LSTM2 gives the average accuracy of 91.96% and maximum accuracy of 93.17% using 10-fold cross-validation, which is much higher than the accuracy obtained through MLP. This result demonstrates the capabilities of LSTM to remember long term information from the sequential data.

4.4 Conclusion

In this chapter, we have presented EEG signals based stress classification using various machine learning and deep learning models. In our first study, the

Gradient Boosting classifier obtains the highest accuracy and in the second study LSTM performed good. Furthermore, various studies used the 4 channels Interaxon Muse headset for stress classification [10, 23, 22, 50, 29] which is the same as used in this study. We achieved either comparable or better accuracy compared to their studies. Moreover, our system outperforms as compared to the approach proposed in [1, 27] for stress detection. But at the same time, it is essential to note that direct comparison is not possible because of the difference in the type of stimulus used, the number of participants, the feature selection techniques and classifiers used in all the studies. In the subsequent chapter, the analysis of EEG signals using Higuchi Fractal Dimensions is presented.

Chapter 5

Higuchi Fractal Dimension Approach to Differentiate Subgroups of Stressed and Non-Stressed Individuals

5.1 Introduction

Everyone experiences stress in their life in one way or the other. As stress can cause harmful effects on health thus, early diagnosis of stress is highly prized. Given the rapid increase in the number of people suffering from stress, there is a need to extract accurate biomarkers that researchers and medical practitioners can use to detect stress during its early stages. Moreover, it will also help patients gain access to appropriate health care services and facilitate the development of new therapies.

The body's physiology changes as a result of the stress response. These physiological alterations are caused by variations in the activity of the brain's central autonomous system [33]. Therefore, a biomarker that provides a measure of changes in the brain in the period of stress would be helpful for early diagnosis of stress. Various techniques for subjective measuring of one's perceived stress have been tried, but these methods rely on imperfect recollection and reconstruction of prior feelings.

Electroencephalography (EEG) is a non-invasive technology utilized in clinical and research settings [145]. With recent improvements, wearable EEG devices can quantify human stress by directly collecting central nervous system activities during stress. Potentially, the EEG can play an important role in detecting stress, but no reliable EEG biomarker exists for the early diagnosis of stress.

Fractals have been widely used to capture complexity in the biomedical signal. By complexity we mean to day that fractal utility has been proved to quantify changes in structure of various signals upon exposure to a stimulus [146]. For instance, fractal were used to capture the complexity from Phonocardiogram (PCG) [147], EEG [148, 149], ECG [150] and Electromyography (EMG) signals [148] for the detection and classification of various diseases such as cardiac abnormalities, obsessive-compulsive disorder (OCD), Multiple Sclerosis (MS) etc. The Higuchi Fractal Dimension (HFD) [151] of the EEG has proven to be very appropriate for electrophysiological data. HFD has characteristics to capture the complexity in the EEG signal [152]. It is non-linear method hence suitable for non-linear signals (non stationary and irregular) like EEG [67]. Authors in [153] hypothesized that the change of fractal dimension (FD) values can reflect the change of emotions and incorporating fractal dimensions causes an improvement in emotion classification accuracy.

Fractals have been used extensively in various biomedical application such as stress classification [27], emotion classification [154], mental state classification from EEG [155], arithmetic task recognition [156], classification of EEG signals of participants with and without schizophrenia [157], etc. Authors in [158] tested the chaotic behaviour of the brain during chronic stress using Higuchi's Fractal Dimension feature from EEG data. Authors in [27] showed an increased stress classification accuracy upon using HFD as one of the features.

Various attempts were made in which multiple features for stress classification were used [22, 23, 27]. In the proposed work, only the Higuchi fractal dimension is used to find the statistical significance difference between stressed and non-stressed sub-groups.

Moreover, it is found in previous research that the brain's right hemisphere is associated with negative emotions and left with positive emotions [159].

Consequently, determining the brain regions that are affected by stress may be helpful in the identification of potential biomarkers and eventually helps in the early diagnosis of stress-related diseases.

Furthermore, numerous researchers have quantified stress in two different settings: controlled laboratory settings and natural settings (real-time). For instance, in the laboratory setting, stress is induced intentionally, and there are few interference from the physiological measurements [22, 23]. A natural setting is more realistic than a laboratory setting, but physical activity can interfere with physiological measurements. Consequently, stress detection and its evaluation are different in the above settings.

A variety of stress-inducing stimuli have been used in the above settings. Existing stimuli had a lack of fun and engagement factors. For example, existing research utilizing SCWT had boring stimulus [160], boredom can also cause stress [161], so the experiment was kept engaging to avoid experiment related stress.

The CWMT was designed considering the input-process output model [132]. This model takes into account multiple features for instructional game design like mystery, stimuli, control etc [48]. As physiological signals are heavily affected by subjects' movements such as running, sitting, and standing, controlling these movements can result in more reliable data [148]. The researcher can restrict the movements in laboratory settings but not in a natural setting. Therefore in this paper, a laboratory setting is used to collect data. Brain signals of the subjects were simultaneously collected using the MUSE headband EEG device while they were being exposed to these stress stimuli.

Thus, in this chapter a system to find the statistical significance difference between stressed and non-stressed groups is proposed and is attempted for the first time to the best of our knowledge. Moreover, we designed gamified Color Word and Memory Test (CWMT) as stimulus.

5.2 Material Used

In this section, the hardware and software setup required for performing stress detection experiment is described.

5.2.1 Hardware- Interaxon MUSE EEG headband

The Interaxon MUSE headband is a four-electrode non-invasive device which is used to record brain activity from four different locations in the brain. It collects data at a notch frequency of 50 Hz. It is easy to adjust and comfortable to wear by the subjects. It is made up of dry electrodes, hence does not require the usage of gels to get the electrodes working. Thus, the MUSE device is highly portable, flexible, and non-intrusive. The 10-20 electrode configuration is used to position the electrodes at $AF7$, $AF8$, $TP9$ and $TP10$ locations (Fig 2.2). The reference electrode for this device is at position F_{pz} (see Fig 2.2). With the help of Bluetooth technology, it can be paired with a mobile. An application called Mind Monitor was used on the mobile phone to record the electrical activity of the brain in Comma-Separated-Value (CSV) format. It was later transferred to the laptop for offline processing.

5.2.2 Software- Color Word and Memory Test(CWMT)

The CWMT was designed based on one of the most popular stress induction technique i.e SCWT [162]. It is a highly reliable and reproducible test for inducing stress as compared to other stress induction methods [28, 163, 164].

In the standard SCWT, there are two parts: a congruent and an incongruent part. The congruent part comprises reading the coloured words, and the incongruent part comprises reading the colour of the ink in which the word is printed. It has been shown in numerous studies that participants require more time in the incongruent section as compared to the congruent section [165, 27]. In the literature, different versions of the Stroop Color-Word test exist. The number of stages, duration of each task in a particular stage, the scoring key used at every stage, and how users provide their input for evaluation (touch/speech) are changed to bring out the variations [27, 33].

Proposed Design of Stress Stimuli

As proposed by [45, 94], stress induction requires the engagement of the user with the stress stimuli. To this end, a stress stimulus containing gamification features was proposed as gamification promotes engagement [166, 48]. So, the existing SCWT stressor was improved by adding different features such as a level-based

S.No	Parameter	Usage in the application
1	Mystery	Score achieved at the end of each level causes curiosity among participants
2	Goals	Clear set of instructions provided before each level to make the goal clear
3	Stimuli	Inclusion of engaging sound and color effects
4	Challenge	Level of difficulty keeps increasing with increase in levels
5	Control	Charge given to participants to choose their actions like selecting particular level to play

Table 5.1: Game design parameters and their usage in the gamified application

approach, score per question(instant feedback), performance accuracy and timer.

Existing stimuli had a lack of fun and engagement factors. Existing research utilizing SCWT had boring stimulus [160], boredom can also cause stress [161], so the experiment was kept engaging to avoid experiment related stress.

Table 5.1 indicates the usage of game parameters from input-process output model in the proposed gamified application [132]. Even though the parameter control was implemented, but in order to be consistent with all the participants they were let to play sequentially.

The protocol for the experiment of stress induction is adapted from [27]. The experiment was designed to induce five different stress levels (from level L1 to L5) accompanied by data acquisition using an EEG headband (See Figure 5.1).

To cause more stress in the participant, the timer is displayed with a clock-like sound (refer Fig. 5.2), as time pressure is significantly correlated with stress [57]. In the beginning, the participants were made familiar with the experiment environment. The experiment procedure was explained to each participant thoroughly. During this phase, they also filled out a consent form to participate in the study. Instruction about the help page of CWMT was also given, and they were made familiar with the different colours used in the experiments before starting the procedure. The CWMT application consists of the following levels (refer Fig. 5.1):

1. Level 1 (L1): This level induces a low-stress state. The participants were

Level Number	L1	L2	L3	L4	L5
Duration (seconds)	10	10	5	3	12

Cognitive warmup — Low — High

Figure 5.1: Experimental design for data collection- Here Level 1(L1) is congruent level, Level 2, 3, and 4 (L2,L3,L4) are in-congruent levels with varying difficulty and Level 5 (L5) is in-congruent level along with memory test. Duration represents the time in seconds given to each question in the particular level. L2 and L3 are categorized as "Low" difficulty and L4 and L5 are categorized as "high" difficulty.

asked to identify the colored word name and chose the answer (refer Fig. 5.3). The participants were supposed to answer each question in ten seconds. Our application automatically calculated and displayed the points scored and the total time taken. Additionally, the interface displayed a continue button to let the participant start the next level at his own pace.

2. Level 2 (L2): This level induces mild stress. The participants need to select the words' font color i.e the color of the ink in which word is written (refer Fig. 5.4 The participants will be more stressed than in Level 1 because of increase in complexity of task. The time given for each question is 10 seconds at this level.

3. Level 3 (L3): This level induces further time pressure. The basic step is the same as Level 2, but only five seconds are given to answer each question.

4. Level 4 (L4): The basic step is the same with levels 2 and 3, but the participants need to respond to the words' font colour within further shortened span of time. Three seconds are given to answer each question.

5. Level 5 (L5): Authors in [1] found that multi-tasking increases stress, so inspired by this approach, at this last level of CWMT, subjects performed two subtasks. Firstly, the subjects memorized the colour coding. For

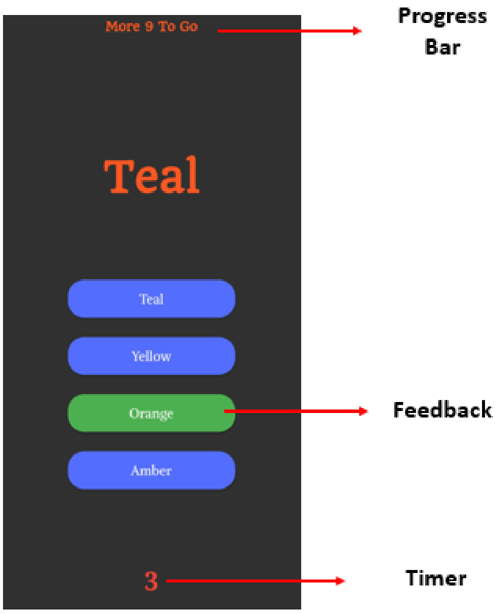


Figure 5.2: Figure shows the various features in the CWMT gamification stress inducer application such as Progress Bar - Shows the remaining questions at particular level, Timer - gives the remaining time with clock like sound, Feedback- Highlighted upon selecting the answer, the blue background of the options which changes to red or green depending on the answer is incorrect or correct respectively.

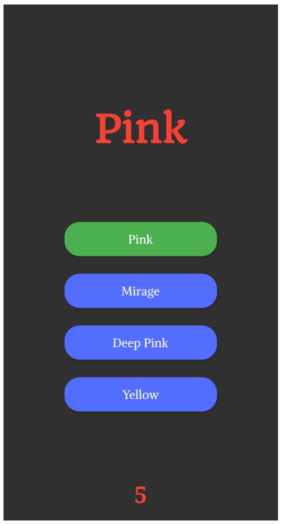
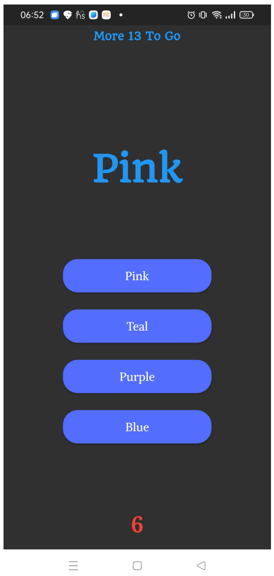
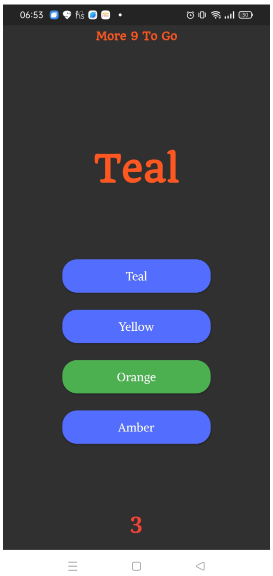


Figure 5.3: Congruent level- The participant is supposed to read the word and choose the answer accordingly. If the answer selected is correct - the colour of the clicked option turns to green otherwise tuned to red if the response is incorrect. The time remaining to complete the task is shown at the bottom of the screen .



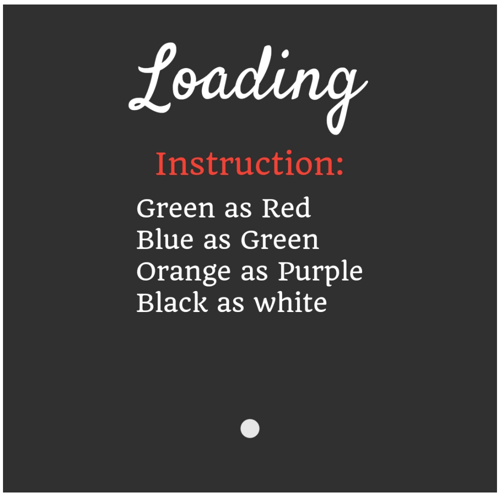
(a) Initial screen with no input selected

(b) Incorrect input selected by user - feedback shown in red color

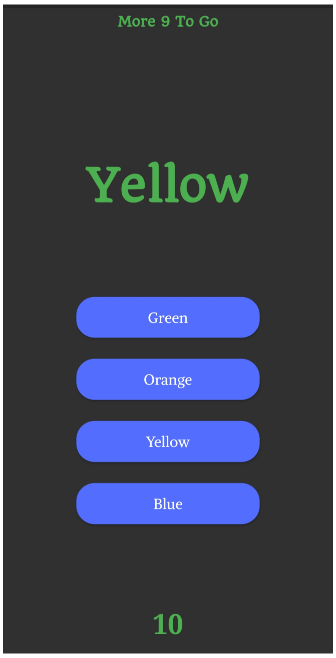


(c) Correct input selected by user - feedback shown in green color

Figure 5.4: Incongruent Stage - User has to select the color in which the displayed word was written. If one selects the incorrect answer then a visual feedback of red color along with sound effect was given whereas if one selects the correct color a visual feedback of green color along with a sound effect is given.



(a) Color coding



(b) Memory Test

Figure 5.5: (a) Color coding shown to the participants only once before the memory test (b) The memory test in the form of in-congruent section

instance, they have to memorize the colours Green as Red; Blue as Green. Secondly, the incongruent section was displayed wherein if the identified colour ink was Green; the participant should select the answer as Red. If the answer is Blue, then the participant should select Green as the correct response (see Fig. 5.5).

Thus, the basic test is the same with level 2,3 and 4, but the subjects face an additional challenge of memory test. Instructions were displayed for a longer time to familiarize the participants with this level. The time given for each question was 12 seconds.

5.3 Methodology

5.3.1 Participants

This study involved 40 participants(18 males and 22 females) aged between 15-17 years, studying in senior secondary school. All the participants were from a similar demographic background.

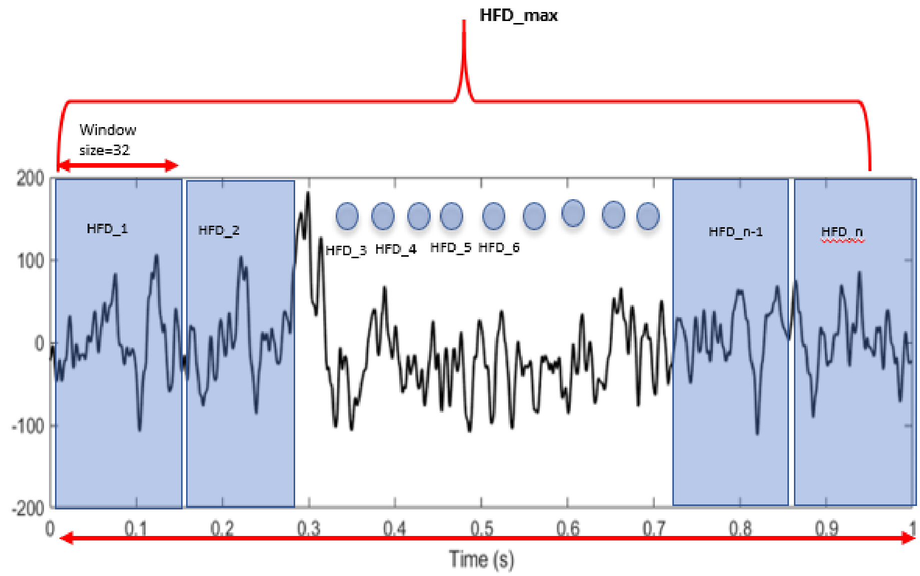


Figure 5.6: Windowing Process to extract HFD feature from EEG recording.

5.3.2 Data Cleaning

The motive of this step was to remove erroneous data readings. It was noticed that these readings generally incurred when a participant tried to wear or remove the device. It also occurred when the device did not correctly fit on the participant's head. Thus, the samples from 8 participants were removed from the dataset during the data cleaning stage because of missing EEG data which occurred because of loose contact between the device's electrodes and the participant's skin.

5.3.3 Feature Extraction

EEG signal was acquired while the participant performed low (Level L2 and L3) and high (Level L4 and L5) difficulty levels. There was one EEG signal per participant for the entire experiment. The acquired EEG signal of each participant was segregated into low and high difficulty tasks. To do this , each participant's time in the low (L2 and L3) and high (L4 and L5) difficulty levels was noted. The EEG signal was divided into low (SL) and high (SH) difficulty task signals by referencing the noted time.

Through the windowing approach, SL and SH signals is segmented. The initial few seconds of recording was removed considering it as a time for cognitive warm up [64]. Next, the entire recording was divided into segments each of size 32 (as shown in Fig. 5.6)[167, 168, 169, 170, 171]. Thus, if the whole signal is of duration x seconds, $(x/32)$ number of windows were obtained. Finally the evaluation of HFD of signals acquired from each window was done and HFD_{max} was extraced among all the acquired HFDs (refer Fig. 5.6), where HFD_{max} is the maximum value of HFD over the window.

5.3.4 Statistical Analysis

Shapiro Wilk's test was used to test the normality of the data. To test for a significant difference in mean fractal dimensions between sub-groups (stressed and non stressed), the p-values were calculated using the two-sided non-parametric Mann Whitney U test with *alpha* set to 0.05. This test was used because it is independent of underlying sample distribution. The test was done at the significance level (alpha) of *0.05*. The null hypothesis ($H0$) that the mean of the two groups is equal was tested. The test was done to ascertain if HFD will be able to differentiate the sample distributions and if the distinction is statistically significant. Figure 5.7 indicates the overall methodology for this work. We performed three types of statistical analysis namely Statistical *Analysis I, II and III*. They are explained in subsequent section.

5.4 Results and Discussions

Three experimental analysis were performed namely *Analysis I*, *Analysis II* and *Analysis III*. The first and second analyses aimed to identify the brain's most impacted region and associated frequency band during stress. Hemispheric differences in the frontal region in the brain during stress were identified through the third analysis. In rest of the chapter, for all analysis, we designate a specify frequency acquired from certain electrode position (place on scalp) for N number of participants as *Freq_EPos_N*, where, *Freq_EPos* is frequency from certain electrode position from N Participant. For example beta frequency band from the $AF8$ region of the brain is designated as BA8 and from the $AF7$ region as BA7. Similarly, alpha frequency band from the $AF8$ region is designated as AA8 and from the $AF7$ region as AA7.

These analyses are described in detail in the below subsections:

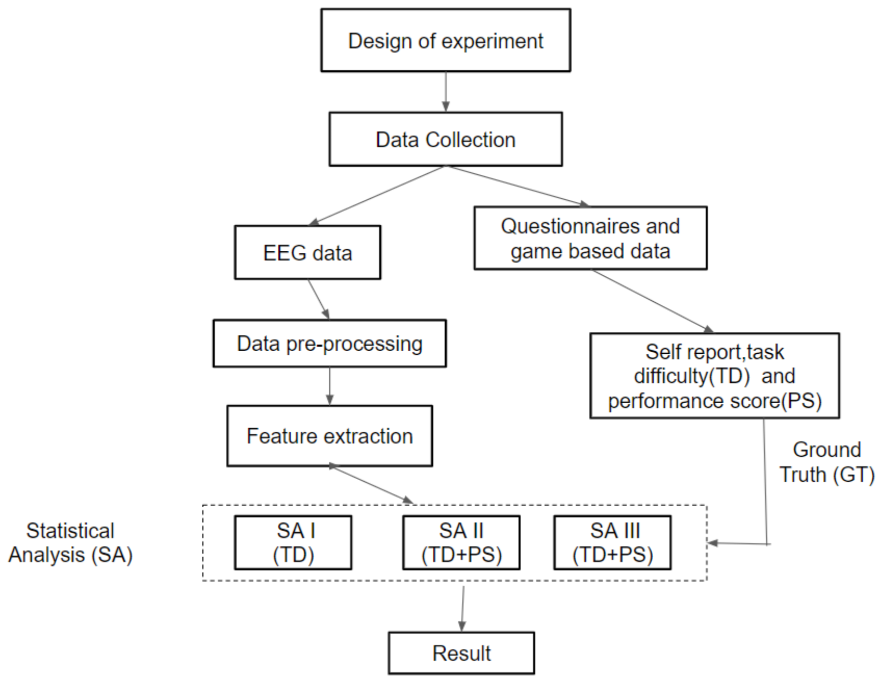


Figure 5.7: Methodology for statistical analysis

1. **Analysis I:** In this experiment, the analysis on all 32 participants was performed. As shown in Table 5.2, the average complexity of HFD increases upon increasing task difficulty i.e mean HFD is higher when participants perform high difficulty task. Table 5.2 shows the statistical significant difference ($p<0.05$) between participant performing low and high difficulty task for both beta and alpha waves originating from $AF8$ region of the brain (i.e for BA8_32 and AA8_32). The table also shows that mean HFD of High difficulty task is also greater then low difficulty tasks in BA7_32 and AA7_32 but this result is not statistically significant (as p-value 0.05). Thus, beta and alpha waves originating from $AF8$ (BA8_32) and $AF7$ (BA7_32) respectively, are able to successfully differentiate between mean HFD of participants while performing low and high difficulty task.

2. **Analysis II**: To further verify the result obtained through analysis I, another analysis was performed in which our initial set of 32 participants was divided into low score (LS) and high score (HS) groups based on their performance in CWMT. This division gave 17 participants with high performance (HS) (considered as low stressed people) and 15 participants

Freq_EPos_N	**Task Difficulty**	**Mean HFD**	**Std**	**p-value**	**Observation**
BA8_32	Low	0.910	0.356	0.001	Sig. Diff.
	High	1.583	0.701		
AA8_32	Low	0.513	0.123	0.011	Sig. Diff.
	High	0.878	0.582		
BA7_32	Low	0.850	0.850	0.097	No Sig. Diff.
	High	1.055	1.055		
AA7_32	Low	0.512	0.512	0.403	No Sig. Diff.
	High	0.517	0.517		

Table 5.2: Analysis I - Aimed to identify significant EEG regions and frequencies while 32 participants perform tasks of varying difficulty. Here, Sig. Diff stands for Significant Difference and Std stands for standard deviation.

Freq_EPos_N	**Ground Truth**	**Task Difficulty**	**Mean HFD**	**Std**	**p-value**	**Observation**
BA8_15	LS	Low	0.883	0.249	0.003	Sig. Diff.
		High	1.443	0.663		
BA8_15	HS	Low	0.783	0.193	0.078	No Sig. Diff.
		High	1.169	0.518		
AA8_15	LS	Low	0.860	0.710	0.460	No Sig. Diff.
		High	0.730	0.550		
BA7_15	LS	Low	0.764	0.166	0.110	No Sig. Diff.
		High	1.238	0.765		
BA7_15	HS	Low	0.984	0.352	0.063	No Sig. Diff.
		High	1.214	0.178		

Table 5.3: Analysis II - Aimed at validating the results of analysis I by dividing group of 32 participants into sub-groups (Low score (LS) and High score (HS)) based on their performance in CWMT. Here, Std stands for standard deviation and Sig. Diff stands for Significant Difference.

with low performance (LS) (considered as high stressed people). To perform statistical analysis, 15 participants were considered in each group, data from 2 participants was dropped randomly from the LS group.

It can be inferred from Table 5.3 that in LS group, beta frequency band from $AF8$ region is significantly affected due to stress. The results obtained are in line with [33] who indicated that stress is correlated with Beta EEG power in pre-frontal area of the brain. The region AF8 that was found to be highly significant in detecting stress is also located in the brain's frontal region.

Table 5.3 also illustrates that there is significant difference between mean HFD extracted from Beta waves from AF8 region while performing low and high difficulty task for LS sub-groups. Thus, Beta $AF8$ region which is located in the frontal part of the brain (see Fig. 2.2) is affected more during high difficulty i.e stressed scenarios.

Freq_EPos_N	**Ground Truth**	**Task Difficulty**	**Mean HFD**	**Std**	**p-value**	**Observation**
BA7_15	LS	Low	0.937	0.630	0.002	Significant Difference
BA8_15		High	1.508	0.719		

Table 5.4: Analysis III: Significant difference found between left ($AF7$) and right ($AF8$) part of brain during stress(low score - indicated by LS). Here, Std stands for standard deviation and Sig. Diff stands for Significant Difference.

3. **Analysis III**: Building upon the above findings, another analysis was performed with the group scoring low (i.e., having high stress group). The objective was to infer if fractals would be able to capture the asymmetry in the left and right parts of the brain during stress. To infer this, the mean HFD of SL and SH (see section 5.3.3) of beta signal acquired from AF7 region i.e $BA7$ (situated in the left frontal part of the brain), and AF8 region i.e $BA8$ (situated in the right frontal part of the brain) was taken. It can be inferred from Table 5.4 that there is a significant difference in HFD acquired from the left and right parts of the frontal region of the brain during stress.

Thus, it can be concluded from the above results that beta frequency from the AF8 region in the frontal part of the brain (BA8_32 and BA8_15) is a characteristic indicator of stress. HFD can capture the intra-individual (Analysis I) and inter individual (Analysis II and III) differences happening in the brain during stress. Also, there is asymmetry of stress occurrence, i.e., the right part of the brain (BA8_15) is more affected than the left part(BA7_15) during stress (Analysis III).

5.5 Conclusion

This work aims to employ fractals to record the irregularities in EEG signal data to attempt to distinguish high-stress conditions from lower-stress conditions and determine if a distinction can be made that is statistically significant. In order to accomplish this aim, an innovative Color-Word and Memory Test CWMT was developed as a gamified mobile application. This test was inspired by standard stressor SCWT. After that, a system was proposed to capture the complexity of EEG signals extracted from different brain regions to understand the region-specific neural changes due to stress. Different experiments were conducted to find a significant difference between sub-groups (stressed vs. not stressed) using fractal dimensions extracted from EEG signals from different brain regions. The analysis results show that the feature computed from data collected by an EEG headband shows statistically significant differences between non-stressed and stressed periods. Specifically, it was found that the $AF8$ region of the brain is affected by stress. At region $AF8$, beta waves are characteristic indicators because of their ability to differentiate stressed (high difficulty) and non-stressed (low difficulty task). This differentiation was performed using HFD from only one physiological signal (EEG).

Thus, in this chapter, a CWMT stress stimulus application was designed and presented for stress elicitation. Even though virtual reality-based tools are more desirable as a stress stimulus as they cause better engagement compared to audio and video-based stress stimuli, such systems are challenging to construct [172]. The approach followed in this chapter, i.e., gamifying benchmarked stimuli, would help get better engagement with users, capturing their feelings more effectively

[45], [118]. Because various people respond to the same stimuli differently, human psychology plays an essential role in labeling the level of stress experienced. To minimize more inaccuracies, the participant should accurately characterize and label the stressful experience. This psychological aspect depends entirely on the individual's comprehension of the event and capacity to label it accurately. Thus, the additional annotation required to authenticate the participant's response is crucial to correctly classify stress from non-stress individuals. The present study can better identify the acquired EEG data by labeling a person as stressed or not stressed based on their performance in the CWMT.

Chapter 6

Game Based Platform to Improve the Experience of Learning

6.1 Introduction

The previous chapter used a gamified app to understand the participant's interaction with the stimulus. While this approach was better than audio-visual stimuli, there was only one extra annotation: using the player's scores to validate the ground truth. In this chapter, the methodology for designing effective and engaging stimuli is explained. Additionally, the technique of in-game analytics is used for the educational game called *Unlock Me* to collect the player interaction metrics in depth [1].

For this study, COVID-19 is considered as a use case. The methodology used and the lessons learned do not just apply to COVID-19 games but to any game which seeks to impart learning and change human behaviour. For instance, spreading awareness about a disease is one such application area.

For spreading awareness about different situations or diseases, many game-based awareness applications have been developed [131] [129] [173]. But to spread COVID-19 awareness, very few games have been developed. To educate users about the precautionary measures related to COVID-19, a mobile game *SurviveCovid-19*, has been developed [174]. To find a game's impact on users, it is essential to thoroughly evaluate the game using the existing game evaluation

[1] This chapter is a revised version of my publication: [48]

models. Although the game *SurviveCovid-19* allows users to learn by doing, it does not test the user's knowledge throughout the game. The game also does not adapt user-centric methodologies in the designing process and does not use a standard evaluation model or in-game analysis to give a behavioral analysis of the player and evaluate game performance. The inception of developing *Unlock Me* came when the first wave of COVID-19 hit India. People's stress levels increased as they were confined to their homes during the lockdown. People broke the government's lockdown rules, not understanding the sensitivity of the situation. When the government made strict rules, people felt locked and wanted to move out of their homes. Hence the name *Unlock Me*.

So the objective was to develop a game that would entertain the people while they were inside their homes. *Unlock Me* simulates real-life situations and experiences to help players understand and feel for discriminating between correct and wrong practices, thereby reducing their risk of contracting the virus. Within *Unlock Me*, certain features have been introduced to help people get accustomed to the game and ease their learning process. A product is effective if it positively influences the user's actions regarding productivity, learning outcomes, or decision-making. To know the level to which the game, *Unlock Me* was influential in creating awareness about COVID-19, extensive game evaluation was performed [175]. There have been various methods to measure the effectiveness of user interface over time, some of which are the input-output model [132], the game-based evaluation model (GEM) [134] and the MEEGA+ evaluation model [135]. These models are replicable. To provide a comprehensive design and evaluation of the proposed game, the parameters and factors used in the MEEGA+ model are considered. It is the latest game evaluation model available to date [135].

The players were presented with multiple challenges throughout the game, and how they respond to these challenges has been recorded to understand the game's effectiveness. An attempt has been made to empirically deduce the factors used in the MEEGA+ model using in-game analysis of the player's progression during gameplay. This method of in-game analysis can prove helpful in data-driven validation of the player experience. It does not require direct interaction with the players [176]. For instance, authors in [177] discuss the importance of validation using factor analysis of the two standard game experience scales - PENS and GEQ. The approach of validation using in-game analysis can aid in the validation of many game scales, provided the mapping of game experience to in-game events is correct.

Our significant contributions, in this chapter:

1. Adopted a user-centric methodology where iterative design activities were conducted to conceptualize *Unlock Me* based on insights from the real world. The inclusion of these activities into the game development process is also explained. The involvement of the users in the game development and the type of usability issues that got uncovered because of involving users and using in-game analysis is described in detail [Section 6.2].

2. *Unlock Me* is designed and developed to have deeper knowledge penetration and easier understanding for users from diverse age groups. The pitfalls of the existing game interfaces have been overcome by using real-world driven game features. These features have been bench-marked by associating them with the design principles used in Human-Computer Interaction (HCI). In addition to this, the significance of utilizing each game feature in instilling awareness about various COVID-19 norms among the players is also presented. *Unlock Me* is freely available on the Google Play Store and runs on all devices with Android version 4.4 and above [Section 6.3] .

3. A novel procedure is used to evaluate the game as a learning tool and as a user-centered game across age groups by combining the standard MEEGA+ game evaluation model and the in-game analysis. The procedure and metrics used for evaluation as a learning tool and as a user-centered game are described. An improvement from a minimum of 22% to a maximum of 90% in the lessons learned by users is seen in the age group 17-21 years; this shows that our game effectively delivered the desired COVID -19 learning. Our findings from other age groups are also presented and it was inferred that even though people from these groups could learn and operate the game, the elderly faced the maximum challenge in completing

it. Evaluation of Unlock Me as a user-centered game has been done using the MEEGA+ model. The validation of this model has also been done by devising metrics from in-game analytics. To the best of our knowledge, we are the first to validate the MEEGA+ model by studying real-time playing behavior. 52.40% of the players in the age group 17-21 years found the game to be usable with a good player experience and learning by the end of the game. School children in the age group 10-16 years also found the game learnable [Section 6.4].
The limitations and future work are presented in Section 6.5 and conclusion of the work is given in Section 6.6.

6.2 Conceptualizing *Unlock Me* with Real-world Driven Building Blocks

The development of an awareness-based educational game requires a deep analysis of the problem area and requirements of the users to create good user experiences and engage the users in playing the game. Hence, it is essential to put users at the center of game design and development. This user-centered approach consists of early focus on end-users, tasks and measuring user reactions at every stage [178]. This approach has been incorporated by doing multiple iterations and frequent user feedback. The users in this study are comprised of people from three different age groups: The first group contains university students aged 17-21 years. The second group contains school children aged 10-16 years. The third group contain few working professionals in the age group of 30-40 years and few elderly in the age group of 41-50 years. The development of *Unlock Me* happened through four phases, namely: discover, define, develop and deliver (Fig. 6.1).

These four phases have been explained below.

6.2.1 Discover

The first phase is the discover phase, where user requirements are collected. An initial study was conducted among 20 people of varying age groups.This step was done to understand the problems faced from the user's perspective. The better we

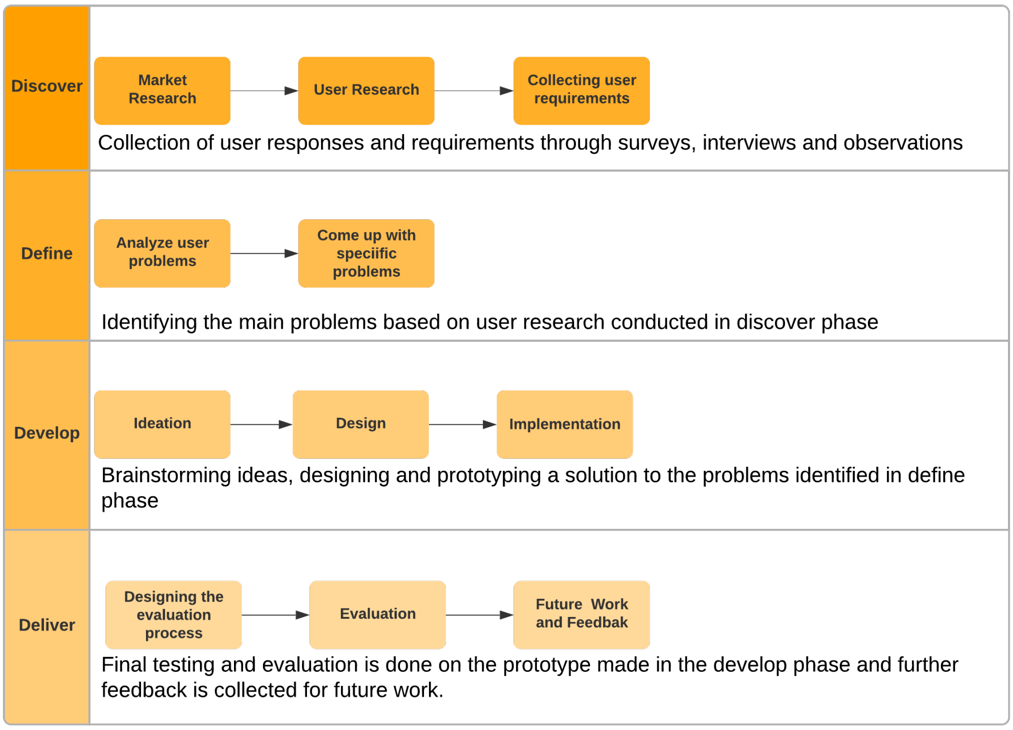


Figure 6.1: User-centred methodology used to design and develop *Unlock Me*

understand our user needs and requirements, the better would be our solutions.

The study comprised face-to-face interviews and group discussions for seven days.

The main objective during the investigation was to know:

1. People's knowledge about the virus.

2. People's opinion about the existing coronavirus based apps.

3. People's expectations from our awareness game.

A set of 9 questions was asked related to people's current awareness about COVID-19 and their opinions about the existing awareness platforms. An overview of responses from the people are given in Fig. 6.2.

It is evident from Fig. 6.2 that most people did not follow the basic precautions like washing their hands after removing masks and using a sanitized mask every day. When detailed questions like "How long does the virus stay on plastic surfaces?" were asked about coronavirus, most people were unable to answer.

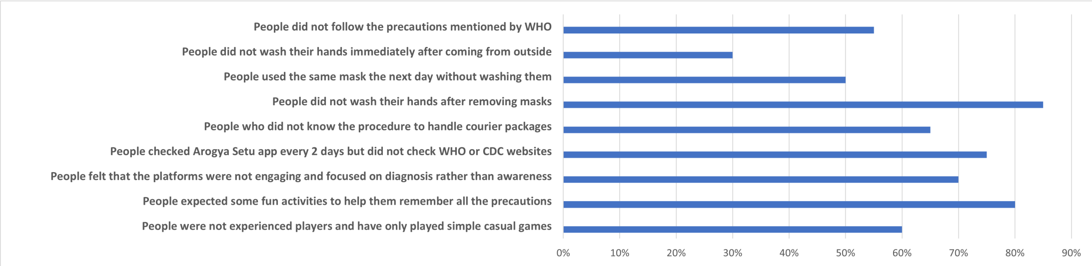


Figure 6.2: Overview of the initial survey conducted in the neighbourhood. 55% of the people did not follow COVID-19 precautions given by WHO, 75% of them did not check WHO and CDC websites regularly, 70% of them felt that the existing platforms were not engaging enough, 80% of them looked forward to using some fun tasks to learn and memorize information related to COVID-19

When enquired about the reasons behind the lack of coronavirus awareness, people shared the following opinions:

1. Websites like the Centre for Disease Control (CDC) and WHO contained information about the virus and the safety measures, but many people found them monotonous. They said it was difficult for them to visit each site and read information related to the virus. Instead, they preferred a platform that could combine entertainment with the necessary information from trusted sites like CDC and WHO.

2. Many people said they did not follow the precautionary measures properly because they kept forgetting about them over time. They felt that the current apps only focused on diagnosing the virus.

3. Additionally, when asked about their preferred method to learn this information, most of them suggested activities that were fun and engaging.

After the users shared their opinions about the existing platforms, their opinions on a game-based learning platform were asked. This idea was proposed by us to the users after they suggested the need for a fun-based learning platform, to know how they feel about the idea. Almost everybody in the study felt that it was a great idea as it would be interactive and the incentives in the game would motivate them to play.

6.2.2 Define

In the discover phase, we collected some opinions of the people related to the existing COVID-19 awareness platforms along with their needs and requirements. In this phase, the main problems of the people were identified from the responses collected. Another iterative feedback session was done with a different set of participants to confirm the problems faced by the first set of users. The main problems were identified after the feedback session. From their responses and from brainstorming among the research team, the following problems were identified related to the existing COVID-19 platforms:

1. They lack the fun and engagement factors while providing COVID-19 awareness.

2. They do not provide a mechanism to increase the memorability of the user

3. The existing platforms do not test the user knowledge over time.

Unlock Me game is designed to address these problems.

6.2.3 Develop

The third phase focuses on prototyping and developing possible solutions to the problem statement. Specific parameters like fantasy, challenge, goals, sensory stimuli, mystery, and control were considered to create an interactive human-centric game and address the lack of fun and engagement factors. These parameters have been considered taking into account the input-process-output model of game design and development [132]. Fantasy involves using an imaginary environment to evoke mental images of situations that do not exist in the real world. Games with fantasy context have proven to show greater interest while learning. Challenge involves the use of progressive levels of difficulty. Goals refer to defining a clear set of instructions before and during the gameplay for the players. Sensory Stimuli consists of the use of attractive graphics and sound effects. Mystery refers to the use of surprises and information that cause curiosity among the players. Control refers to providing the authority to regulate an action in the game to the players. An overview of these parameters have been mentioned in Table 6.2. A detailed explanation of the game ideation, design, and development process has been discussed in Section 6.3.

Throughout the game development process, users were involved to provide their opinions on different features developed in every version of the game. User inputs through questionnaires and video interviews were taken. These consolidated inputs were used to choose and build upon the game features. Each time some features were added to the game, the update was shared with users to get their feedback. A technique similar to Microsoft's "activity-based planning" was used. This technique involves studying what users do to perform certain activities like reading a newspaper, reading the information displayed on LED billboards, distributing information through LED billboards, etc. The study results were used to choose and build upon the game features. A variety of methods ranging from asking users to using an instrumented version of the software to observe user interactions with the game were used. This instrumented version of the software records user actions through event-based logging(game analytics). Game analytics was used to validate the users' feedback and balance the game so that the next set of users do not face similar difficulties. Each time the developer completed a feature, the users tested it. The data was analyzed and the findings were fed back into development. Through this iterative process the usability of our game improved significantly. It also helped to enhance the user experience. Table 6.1 shows the details of each iteration of game development. In this table, *Iteration* is the iteration number, *Duration* is the number of days over which the iteration was complete, *Number of Participants* is the number of participants of varying age groups who participated in the study, *Method of Feedback* is the method used by researchers to elicit user feedback . Some of the useful feedback received were as follows:

1. Users reported difficulty crossing a particular level in one of the iterations

2. Users felt that the initial instructions given to them in the game tutorial were not informative and didn't explain the rules clearly

3. Users felt a lag in the game character movement in the first version of the game.

Each iterative development session included gameplay preceded by instructions and rules, user observations, and detailed feedback. These sessions were conducted online on Google Meet with a moderator to help the users in case they get stuck somewhere. The researchers and participants were involved, any inconsistencies with the testing procedure or task flow were corrected and implemented before the next round of data collection.

6.2.4 Deliver

For the product to have a significant impact on the users, it is essential to evaluate it. A mix of the traditional evaluation model that is MEEGA+, and in-game analysis were used to evaluate the game efficiently. The MEEGA+ model was used for data collection in the form of a post-game questionnaire [135]. Even though most of the games use a questionnaire-based evaluation model, it could be possible that the users might not give honest feedback. Eventually, the findings may be faulty. To efficiently evaluate the learning outcomes and the quality of the game, a method is used that will make this process more reliable. This method of indirectly inferring user's feedback on the game has been done using in-game analytics. This method captures users playing behavior (or the way the user perceives the game) through specific events. These events have been carefully designed and raised when the player accomplishes something or interacts with various elements in the game. The game servers receive the curated analytics and provide an interface to view and download them through a dashboard for further processing. The observations from the iterations in develop phase were also validated using the method of in-game analytics (Table 6.1).

Iteration	Duration(days)	Number of Participants	Type of Participants	Method of Feedback
1	3	5	Researchers	Think-Aloud and Online Interview
2	10	7	Researchers and Junior Professors	Video Interview and Game analytics
3	4	5	Researchers	Video Interviews and Observation
4	4	5	Researchers and Senior Professors from civil engineering	Think-Aloud, Video Interviews and Focus Group
5	10	30	Users from diverse backgrounds	Play Store Release and Game Analytics
6	3	5	Researchers	Internal testing
7	2	20	Researchers and University Students	Video Interviews, Moderated Online Sessions and Game Analytics
8	3	100	University Students	Moderated Online Sessions, Group Discussion, Google Forms
9	14	30	Research Groups, School Students and Elderly	Online Moderated Sessions, Play Store Release, Game Analytics

Table 6.1: User involvement in various iterations in game development. These iterations helped us to uncover major usability issues.

S.No	Parameter	Description
1	Fantasy	Involves usage of imaginary game setup and animated characters
2	Mystery	Optimal level of informational complexity
3	Goals	Aims at providing clear instructions and set of goals to be achieved in each level
4	Stimuli	Involves usage of dynamic graphics, sound effects to grab attention
5	Challenge	Aims at creating and maintaining an optimal level of difficulty throughout the game
6	Control	Aims at providing a sense of control in the hands of users to choose the challenge

Table 6.2: Game Parameters used to develop a game

6.3 Design and Development of *Unlock Me*

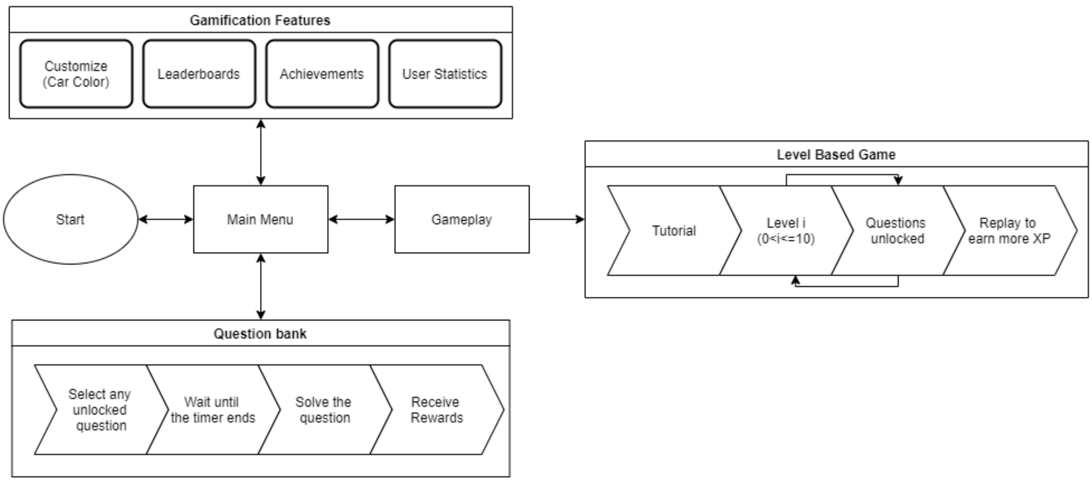


Figure 6.3: Flowchart illustrating the game flow and game play,

It becomes crucial for the player to enjoy and learn by performing simple actions through the interface in an educational game. Thus having an effective interface is of utmost importance. Figure 6.3 highlights the game flow and game play. The game *Unlock Me* has been designed considering the input-process-output model parameters of game design. These parameters, when

S.No	Parameter	Usage in the game
1	Fantasy	Player representing a police personnel and catching the people flouting COVID-19 norms
2	Mystery	Score earned at the end of the game and the quiz in the game can cause curiosity among players
3	Goals	Guided tutorial/Level 0 where players are given clear set of instructions and goals are provided before start of each level
4	Stimuli	Use of 3D Objects and perspective camera for graphics and inclusion of special sound effect throughout the game
5	Challenge	Level of difficulty keeps progressing with the levels
6	Control	Authority given to players to control their actions like catching violators and using sanitizers

Table 6.3: Game parameters and their implementation in the game to address the lack of fun and interactivity problem

implemented, result in an effective instructional game [132]. Table 6.3 describes the way these parameters have been incorporated in the game.

For each parameter, a feature was added that could potentially solve the problems associated with the usability of the existing COVID-19 interfaces. These features have been described in Table 6.4. Each feature has been added in compliance with the standard Human-Computer Interaction (HCI) principles of designing an interactive interface [178].

In this section, the terminologies used in the game, game plot, user scenario, and implementation of the game have been explained.

Challenge	Feature	Design Principle	Significance
Lack of fun and engaging factors	HP bar	Visibility	Simulates health for a virtual player character teaching the players about the spreading of COVID-19.
	Tactile feedback	Feedback	Involves the player deep into the game by simulating sensations of touch using vibrations. This is calibrated in the game to know something big has happened like explosions.
	Leaderboards	Feedback	Users can view where they stand in comparison to other players thereby increasing competitiveness among players.
	Achievements	Feedback	Increases player retention, engages the user to reach a target and provides a feeling of satisfaction to the user.
	Customize	Visibility	Enhances player engagement as the user can change skins and beautify the game as per their liking.
	Statistics	Feedback	Provide detailed analysis of user's performance. Visually represent the data with graphs and diagrams that can be easily understood by the user.
Lack of testing user knowledge over time	Sanitizer	Affordance	Eases the gameplay where user might be struggling in game. Educates users about the importance of sanitizer.
	Quiz	Feedback	Would increase the memorability and the learning outcome of the users eventually.
Lack of providing memorability of useful information	Mythbusters	Visibility	Provides facts related to COVID-19 awareness in detail. Unlocks after completing half of the questions in question bank
	Info Block	Affordance	Shows COVID-19 norms during the game, that must be followed by the people.

Table 6.4: Overview of the features in the game and the challenges they are solving.

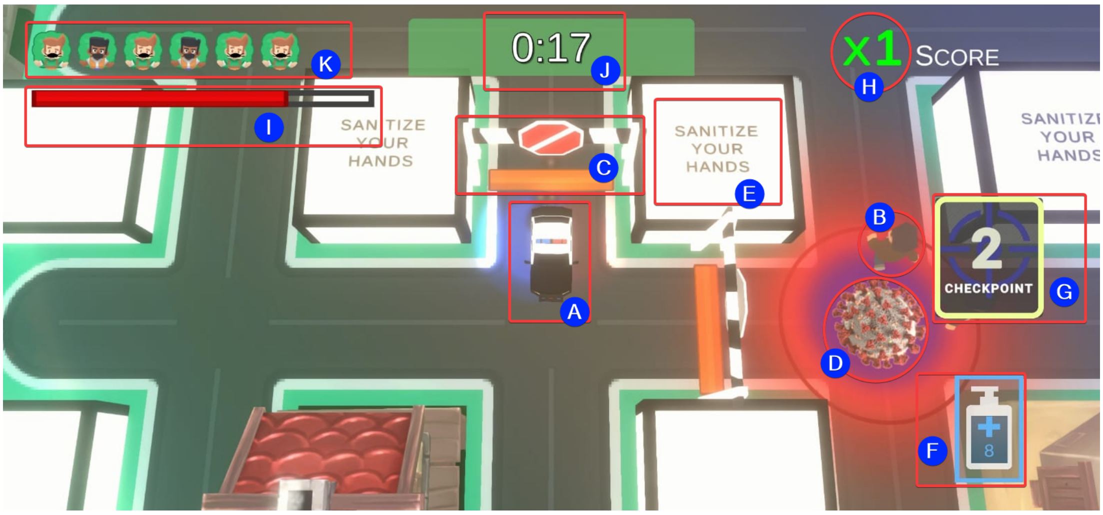


Figure 6.4: Game plot containing game elements - [A] Player [B] Violator [C] Inserted Checkpoint [D] COVID-19 bomb [E] Info Block [F] Sanitizer [G] Number of checkpoints available [H] Score Multiplier [I] Health bar [J] Timer [K] Number of violators left to be caught

6.3.1 Game Terminology

1. Game map : The navigable game area that is available during each level.

2. Characters : The game has one player and multiple non-player characters. The player is police personnel whose job is to patrol the game map searching for people who are violating COVID-19 lockdown guidelines (Fig. 6.4 [A],[B]). Violators are the non-player characters, and it is vital to catch these violators to stop the spread of COVID-19 (Fig. 6.4 [B]).

3. Timer : This is a countdown, which indicates the time remaining until the level ends. The duration of this timer depends on the difficulty of the level and size of the game map (Fig. 6.4 [J]).

4. Check points : These are barriers for the violators placed by the player during the game. Checkpoints are used to catch the violators to avoid direct contact of the player with violators (Fig. 6.4 [C]).

5. COVID-19 bombs : These are game obstacles created by violators and infect the player if they contact them. These bombs are represented by COVID-19 symbols (Fig. 6.4 [D]) and express that coronavirus can stay on surfaces

or could be airborne. Even when the COVID-19 infected violator has left the area, the virus can persist and infect the player if the player comes in contact with the COVID-19 bomb.

6. Levels and Survival mode : There are 11 stages in the game of *Unlock Me*, the first being level 0, a guided tutorial. In levels 1 to 9, the violators are to be caught by the player using checkpoints. Level 10 is the survival mode where the player aims to evade the violators without using checkpoints.

7. Experience points (XP) : This is a measurement to quantify a player's experience in the game (Fig. 6.6b). XP is calculated based on the performance of the player using this formula–

$$XP = (n * m * b) + \sum(t) + e \tag{6.1}$$

where,
n = *number of violators caught*
m = *A multiplier that increases if you catch more violator's in shorter time frame*
b = *Base score for catching the violator's*
t = *The time remaining after each violator is caught*
e = *A bonus that is awarded for level 10 - survival mode.*

After gaining enough XP, the player level increases, which unlocks a question from the question bank. The amount of XP required to make the player level up is calculated using this formula-

$$XP\,required\,to\,level\,up = (100 * current\,XP\,level) + 1000 \tag{6.2}$$

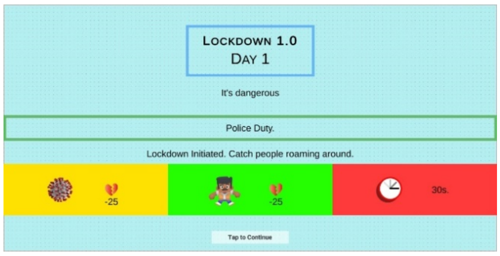
current XP level = The XP level at which the user lies in, based on the total XP gained in the game.

8. Health Points (HP) : This is a value representing the player's health at a game level. Players lose the level if no HP remains. The fear of losing HP can make the user learn about physical distancing and the value of using preventive measures(Fig. 6.4 [I]).

9. Question Bank : This part of the game contains a set of COVID-19 related questions of different difficulty levels. This feature encourages deeper

(a) Main Menu of Unlock Me showing the navigation options - [A] Tutorial [B] Level based game [C] Question bank [D] Myth-busters [E] Customization [F] User Statistics [G] Leader-boards and Achievements

(b) Level menu



(c) Start screen of level-1

Figure 6.5: Screenshots of Gameplay

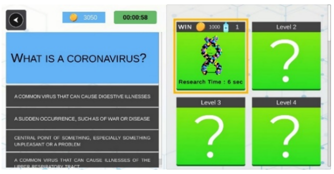
learning of the COVID-19 precautions [179]. Inside the question bank, players can answer a question related to COVID-19 as soon as they reach a new level (Fig. 6.6c). These questions are repeated like flashcards to help users remember the information easily.

10. Info blocks : These are cubes present in the game level's background that show COVID-19 norms. Some basic prevention measures are displayed during the game, analogous to billboards, to make the player aware of COVID-19 norms known as an info block. (Fig. 6.4 [E])

11. Myth-busters: This section of the game contains clarifications and facts from WHO about some common misconceptions related to COVID-19. (Fig. 6.5a [D])

(a) Adding Checkpoints

(b) End screen with score board



(c) Question Bank with COVID-19 questions

Figure 6.6: Screenshots of Level 1 of *Unlock Me*

6.3.2 Game Plot

The player's main aim is to catch the violators within the given time frame by creating checkpoints in the game map(Fig. 6.6a). While patrolling, the player might encounter COVID-19 bombs. The repercussions of encountering the violators will make the players realize the importance of social distancing. The COVID-19 bombs will educate the users about the infection left over the surfaces by covid infected people and make the users understand the significance of disinfection of surfaces.

The game has a feature of collecting sanitizers that players can use to regain HP and receive a temporary shield that protects the player from violators and COVID-19 bombs. Throughout the game, COVID-19 precautionary measures are displayed in the info blocks to help increase the memorability of these measures.

S.No	COVID-19 norms	Game Feature	Description
1	Social Distancing	Violators & info blocks	Encountering with the violators will reduce the health points of the player, thus making the players realize the importance of social distancing. The info blocks remind the users to maintain social distancing in public places.
2	Use of sanitizers	Sanitizer power up & info blocks	By using sanitizers, the players can protect themselves from losing health points after encountering a COVID-19 Bomb or a violator. The info blocks keep reminding the players to use sanitizer after coming back home from outside.
3	Wearing Masks	Question Bank, info blocks & Myth-Busters	Question bank contains a few questions related to the correct usage of masks. Info blocks remind the players to use masks when in public places. Myth-Busters clear some pre-conceived notions about usage of masks like usage of masks could lead to CO2 intoxication.
4	Vaccination	Question Bank & Myth-Busters	Questions related to the importance, effectiveness and duration of the vaccines have been asked in the question bank section. Myth-Busters contain clarifications to the common misconceptions related to the vaccines like COVID-19 Vaccines affect ones DNA
5	Cleaning and disinfection	COVID-19 Bombs & Question Bank	Since COVID-19 Bombs are present on the ground or any surface, it warns the players to not touch any surface outside and if they do,they will have to use a sanitizer to regain their HP. Questions related to the use of recommended disinfectants are given in the question bank section.

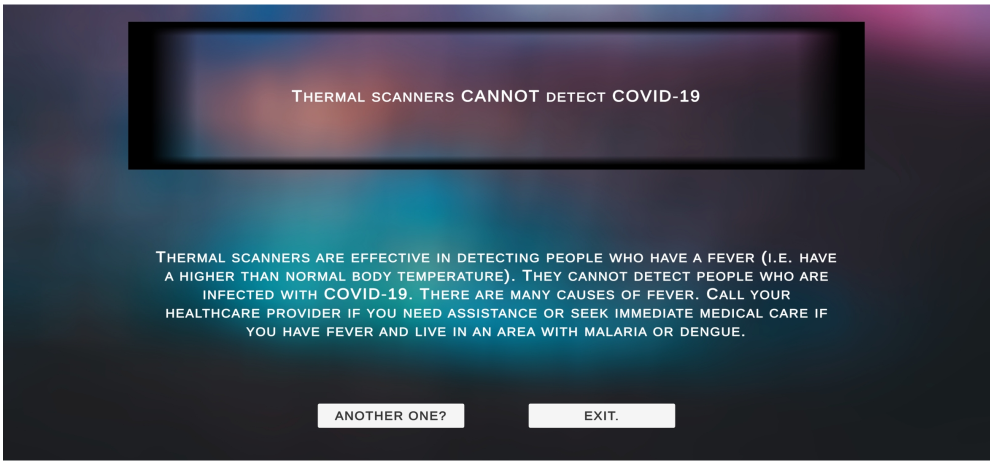
Table 6.5: Game features representing COVID norms

The game starts with level 0 (tutorial), which gives guided gameplay. After completing the tutorial, the player can begin playing from level 1, and once the player completes the level, some XP is generated, and the next level is unlocked. An example of the scorecard with XP generated after playing level 1 is given in Fig 6.6b, where the player gains 765 XP out of 1000 XP. With the XP gained, if the user levels up, a question is unlocked from the question bank, which the player can solve. A screenshot of the questions in the question bank is given in Fig 6.6c. Solving these questions helps the players gain coins that, combined with the coins earned during gameplay, can allow the players to customize the appearance of the police car (depicted in Fig. 6.7b). The question bank has been added to encourage in-depth learning and to test the player's knowledge about COVID-19. After half of the questions have been solved, the Mythbusters section is unlocked (Fig. 6.7a). This section helps the players to clear some misconceptions related to the vaccination against COVID-19, treatment, and diagnosis of COVID-19. This section is beneficial in removing the fears of the players to follow specific precautionary measures. Some of the common misconceptions are wearing a mask causes CO2 intoxication, vitamin C supplements can help prevent COVID-19, and COVID-19 vaccines change a person's DNA and cause infertility. The Myth-Busters contained in this section have been cited from the WHO website [180]. During the gameplay, the players can access the statistics section, placed on the left side of the main menu (Fig. 6.5a [F]). Statistics contain the player's progress, the number of levels cleared, the number of questions answered, and the overall score (Fig. 6.7c).

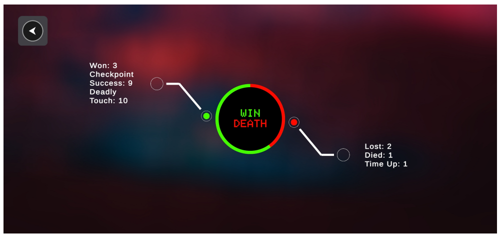
The players have been educated about the various COVID-19 norms, precautionary measures, and information throughout the game. These features and the COVID-19 norms that the players are getting educated about are mentioned in Table 6.5. Also, a list of the game's main features and their significance in addressing the player problems have been given in Table 6.4.

6.3.3 Game Mechanics

To develop the game prototype, Unity3D game engine was utilized. To simulate the objects in the scene and to ensure that they respond to collisions and forces, Unity provides a physics engine. 2D physics to make the mechanics of the games.



(a) Myth-busters



(b) Customization of the Police Car

(c) User Statistics

Figure 6.7: Screenshots of Game Features

Initially, the levels were made using 2D graphics called sprites created in GIMP (GNU Image Manipulation Program), a freely distributed program used for image retouching and composition. To enhance the feel of the environment from a player's perspective, the 3D objects were placed and a perspective camera was used. In terms of game development, a camera is a system that shows the player's action from different angles. There are two types of cameras called orthographic and perspective camera. A perspective camera is used with 3D objects capturing how the real world is seen. In the real world, if a long road (from our perspective) is looked at, it appears to be tapering at the end. A similar view is represented in the game using the perspective camera. On the other hand, an orthographic camera is commonly used with 2D objects, and it has no perspective.

A perspective camera is used with a heads-up overlay display (HUD) canvas displaying the various stats and information. This camera is placed at an angle of 15 degrees tilted along the x-axis. This implementation gives depth to the 3D visuals and enhances the game's visuals, making the interface attractive and fun for the players. One of the main problems identified during the user survey was that most people didn't find the existing platforms fun and interactive. Hence, improving the graphics and flow of the game helps address that problem. 2D and 3D lights were used along with particle effects to lighten the scene. After putting additional post-processing, global and local volumes enhanced the colors and made the scene more vibrant. The effects used were bloom, depth of field, color adjustments, vignette, white balance, shadows, mid-tones, and highlights. The camera transitions from perspective to an orthographic camera with a 0-degree tilt and larger Field of View (FOV) while placing checkpoints in the game to avoid misclicks and maximize the area visible. This camera transition is done using a combination of "Cinemachine," a tool for handling cameras, and by a C# script. C# scripts have been used throughout the development of the game. Unity analytics were used to track the actions performed by each user while playing the game. The in-game analysis is obtained from the Unity server on which the game data is sent for storage by the player's devices. Each event action that the user performs during the game is sent to the server as custom events.

6.4 *Unlock Me* Evaluation: As a game and As a COVID-19 learning tool

A two-fold evaluation process was adopted for *Unlock Me*, where the former considers the efficacy of it as a learning tool and the latter as a game. Each of these is described below:

1. **COVID-19 Learning Tool:** To evaluate the impact of the game in enhancing COVID-19 learning among the players.

2. **User-centered Game:** To evaluate the quality of the game.

Since COVID-19 is utilized as a use case so the efficacy of our tool is evaluated with this use case. The evaluation was conducted in two phases. In the first phase, a large scale evaluation was performed on the initial prototype of *Unlock Me* with 141 university students (90 male and 51 female students) having average age of 21 years. In the second phase, the attempt was made to check the robustness of this approach so the evaluation study was extended for different age groups. The details for these diverse age groups is given in Section 6.2. But due to COVID-19 restrictions and miscellaneous challenges like unable to provide remuneration, the game was publicized in research groups. Even though we tried but given the time duration that we had and the lack of resources, there were few downloads. Hence the dataset for second evaluation study is limited.

All participants were required to download and install *Unlock Me* from the Google play store. These participants were also briefed about the evaluation process, as shown in Fig.6.8. The procedure and metrics for the evaluation of

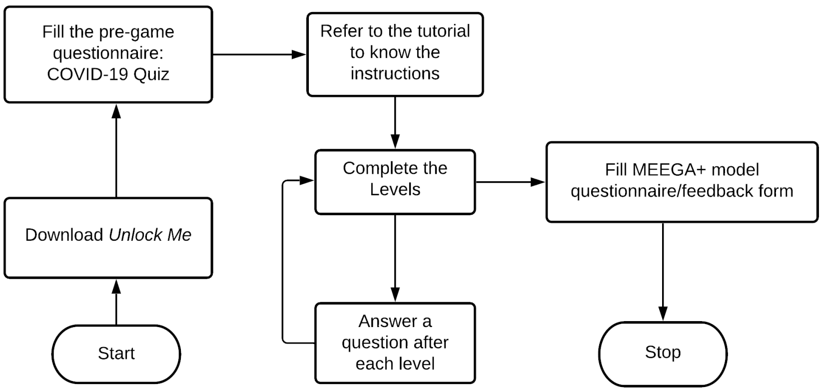


Figure 6.8: Game Evaluation Process

each of these goals are discussed in the following sections.

6.4.1 Evaluation as a COVID-19 Learning Tool

Our main motive to design the *Unlock Me* game was to help the players learn about COVID-19 and implement the knowledge gained in their daily lifestyle. To realize this objective, a pre-game COVID-19 assessment was conducted and the scores obtained were compared with the scores of the questions during the gameplay. In this section, the procedure (Section 6.4.1) and metrics (Section 6.4.1), used for evaluation of *Unlock Me* as a learning tool are presented.

Procedure

To evaluate the amount of COVID-19 knowledge gained by the participants through the game, two types of assessments were conducted:

1. Pre-game Questionnaire: In the pre-game assessment, participants were given an online questionnaire consisting Multiple Choice Questions

(MCQs), where they were assessed on their existing awareness about COVID-19.

2. In-game Questions: In the in-game assessment, participants answered COVID-19 related questions (MCQs) from the question bank while playing the game.

The questions that were selected for both these modes of assessment were divided into three categories: easy, medium, and hard. These categories were made as per the difficulty level of the questions. The easy questions were presented to the users first to establish familiarity with the experimental procedure. Presenting easy questions first is following the law of initial values [181]. Medium and hard questions followed this. Easy questions were assigned 4 points, medium 5 points, and hard 6 points in both the modes of assessment.

Metrics

For both the evaluation modes, the Average Score per Difficulty level (ASD) has been calculated. However, the ASD differs for pre-game and in-game assessment. The number of attempts allowed during the pre-game assessment is only one, while the player can take multiple attempts to answer the questions correctly during the in-game assessment. In the second evaluation phase, the user was allowed to take multiple attempts for pre-game assessment as well so the same formula for calculating ASD was used for pre-game and in-game assessment.

1. The pre-game COVID-19 assessment ASD score was calculated using this formula:

$$ASD(\textit{pre-game}) = \frac{\sum_{i=1}^{Q}\left(\frac{\sum_{i=1}^{N} S}{N}\right)}{Q} \tag{6.3}$$

$Q =$ *Number of questions per difficulty level*

$S =$ *Score obtained for each player per question*

$N =$ *Number of players*

2. In-game Questions: The average scores of the questions for each difficulty level are calculated by taking the cumulative average scores of all the

questions that belong to that particular difficulty level. The formula for calculating the ASD during the gameplay is given below:

$$ASD(in\text{-}game) = \frac{\sum_{i=1}^{Q}[(\sum_{i=1}^{i=N}\frac{S}{T}) * \frac{1}{N}]}{Q} \tag{6.4}$$

$Q =$ *Number of questions per difficulty level*

$N =$ *Number of players*

$S =$ *Maximum score per question according to its difficulty level*

$T =$ *Total number of tries taken to solve the question correctly*

6.4.2 Evaluation as a User-centered Game

To efficiently evaluate the quality of the game, an indirect inference from our in-game analysis was drawn and a comparison was made with a traditional game evaluation model called MEEGA+. This section presents the procedure (Section 6.4.2), and metrics (Section 6.4.2) used to evaluate *Unlock Me* as a User-centered game. This comparison has been made based on three main factors: usability, player experience, and learning outcomes. These factors have been described below:

1. Usability: Usability refers to “ensuring that interactive products are easy to learn, effective to use, and enjoyable from the user’s perspective“ [178]. Hence, to measure usability, the MEEGA+ model provides us with sub-dimensions: Aesthetics, Learnability, Operability, Accessibility, and User Error Protection. Aesthetics measures the game’s overall visual appearance comprising color, movement, fonts, and layout. Learnability is the measure of how easily players can learn the game. Operability refers to the ease with which players play the game, and accessibility refers to the game’s availability to different kinds of people irrespective of age, game experience, region, etc. User error protection is the ability of the game to help players rectify their errors during gameplay.

2. Player Experience: Player Experience refers to the desirable and the undesirable aspects of an interface from the user’s perspective. The sub-dimensions of player experience are confidence, challenge, satisfaction, fun, focused attention, and relevance. Confidence refers to the feelings of

mastery or self-efficacy achieved by playing the game. Challenge refers to the progressive difficulty levels used in the game. Satisfaction refers to the fulfillment of a player's needs after playing the game. Fun is a feeling of enjoyment achieved by fulfilling the challenges of the game. Focused attention refers to the involvement of the player in the game. Relevance refers to the extent to which the game has been designed to fulfill player requirements.

3. Learning Outcomes: Learning Outcomes refers to the ability of the interface to induce some skills or knowledge about a specific topic. The sub-dimensions of learning outcomes are short-term learning and learning goals. Short-term learning focuses on the direct knowledge gained after playing the game. Evaluation of learning goals involves addressing the problems related to learning. It mainly consists of the assessment of levels of awareness, remembering, understanding and application.

Procedure

Game-related data was collected to evaluate the quality of the game through the following techniques:

1. After downloading the game, the participants were asked to play the game and complete all the levels. The participants had to play the game and answer the COVID-19 questions associated with each level in the game. In-game analysis used the anonymous data collected through this gameplay.

2. MEEGA+ Model: The MEEGA+ model uses a post-game questionnaire to assess the game's impact on the players [135]. The participants were provided with the MEEGA+ questionnaire (in the form of a feedback form). The form had questions in a 5-point rating scale format (1-strongly disagree,2-disagree, 3-neutral, 4-agree, 5-strongly agree).

Metrics

An overview of the factors used in the MEEGA+ and the in-game analysis has been given in Table 6.6. The following factors were evaluated:

Factor	Sub-factors	MEEGA+ Question asked	Validating factors
Usability	Aesthetics	The game design is attractive (interface, graphics, text, font)	-
	Learnability	I needed to learn a few things before I could play the game	-
		Learning to play this game was easy for me	Percentage of players who cleared levels 0 & 1 in the first attempt
	Operability	The game is easy to play	Percentage of players who cleared levels 1,2 & 3 in the first attempt
	Accessibility	The game allows customizing the appearance (font & colour) according to my preferences	-
	User error protection	When I make a mistake, it is easy to recover from it quickly	Percentage of users whose average sanitizer usage was greater than the average usage
Player Experience	Confidence	The contents and structure helped me to become confident that I would learn with this game	-
	Challenge	This game is appropriately challenging for me and provides challenges at an appropriate pace	Increase in anxiety level and completion level
	Satisfaction	Completing the game tasks gave me a satisfying feeling of accomplishment	-
	Fun	I had Fun playing the game	-
	Focused Attention	I was so involved in the gaming task that I lost track of time	-
	Relevance	I prefer learning with this game to learning through other ways (e.g., other teaching methods like COVID-19 Awareness platforms and campaigns)	-
Learning Outcome	Perceived Learning	The game contributed to my learning about COVID-19 and I feel satisfied with it	Score comparison from pre-game assessment and questions answered during gameplay
		I was able to remember information related to COVID-19 through this game	-

Table 6.6: MEEGA+ factors used for evaluation of the quality of game

1. Metrics from the in-game analysis: The MEEGA+ factors that have been measured using in-game analysis are stated below:

   (a) Learnability: Since learnability is how easy the game is for the players to learn, this was calculated by the percentage of players who cleared the initial level in the first attempt itself. The initial levels considered were levels 0 and 1. Since level 0 is a guided gameplay level, the players would learn about the rules and controls of the game in this level, and level 1 is the basic level to help players adapt to the environment. If a player clears both level 0 and level 1 in the first attempt, it indicated that the player has not only followed the instructions given in level 0 but has also implemented them in level 1 correctly. The formula that was used to calculate the percentage of players who cleared levels 0 and 1 in the first attempt is given below:

   $$\textit{\% of players who cleared levels 0 \& 1 in the first attempt} = (\sum_{i=0}^{1} \frac{W}{F})*(100) \tag{6.5}$$

   $W =$ *Total wins for levels 0 & 1*

   $F =$ *Number of players that attempted the level*

   (b) Operability: Operability is the measure of how easy it is for the players to operate and play *Unlock Me*. To know the ease of gameplay, we measured the percentage of players who cleared the initial easy levels of the game in the first attempt. Levels 1, 2, and 3 were taken because they belong to the easy category, and from level 4, the difficulty increases and the success rates drop. If a player clears levels 1, 2, and 3 continuously in the first attempt, it indicates that the player can use the controls and play the game with ease. The formula used to calculate the percentage of players who cleared levels 1, 2, and 3 in the first attempt is given below:

   $$\textit{\% of players who cleared levels 1, 2 \& 3 in the first attempt} = (\sum_{i=1}^{3} \frac{W}{T})*(100) \tag{6.6}$$

   $W =$ *Total wins for levels 1,2,3*

   $T =$ *Number of players that attempted the level*

(c) User Error Protection: The game can provide a mechanism to rectify the error made by the player during gameplay. In *Unlock Me*, after making an error of either coming in contact with the violators or the COVID-Bombs, players can use a sanitizer to regain HP. Here, we calculate the percentage of players who have used sanitizers greater than the average usage at all levels. Suppose a player uses all the sanitizers provided to them at each level to rectify their errors. In that case, it indicates that the game is helping them revive their error through a particular mechanism. The formulas used to calculate the percentage of players using sanitizers are given below:

$$PPS = \left(\frac{Number\ of\ players\ whose\ usage\ is\ greater\ than\ average}{Total\ number\ of\ players}\right) * (100) \tag{6.7}$$

$$Average\ usage\ rate\ of\ sanitizers = \left(\frac{\sum_{i=0}^{N}[Number\ of\ sanitizers\ used]}{[Number\ of\ sanitizers\ provided]}\right) * (100) \tag{6.8}$$

$PPS =$ *% of players who used sanitizers greater than average usage*

$N =$ *Total number of players*

(d) Challenge: The challenge is how the players interpret the level in the game to be easy, medium, and hard. The players should feel an optimal level of difficulty throughout the game. To find if the game was challenging enough for the players, we calculated the percentage of players who achieved an overall success rate less than the average success rate of the players throughout the game. If a player has a success rate less than the average, it indicates that the player finds the game challenging overall. The formulas for measuring the percentage of players with fewer success rates and the average success rate are given below:

$$PPSR = \left(\frac{Number\ of\ players\ whose\ success\ rate\ is\ less\ than\ average}{Total\ number\ of\ players}\right) * (100) \tag{6.9}$$

$PPSR = \%$ *of players who had an overall success rate lower than avg*

$N =$ *Total number of players*

$W =$ *Total wins throughout gameplay per player*

$T =$ *Total attempts made by the player*

Additionally, to know if the users found the game challenging, we evaluated the player's anxiety using the State Anxiety questionnaire [182]. State anxiety refers to the change in behavior of a person due to a particular situation or event. State anxiety is tested before and after the game-play to know the game's impact on the individual's anxiety levels. The state anxiety questionnaire contains 20 statements to which the players have to respond on a scale of 1 to 5. (1-strongly disagree, 2-disagree, 3- agree, 4- strongly agree). To calculate the average state anxiety score before and after the game, the following formula is used:

$$Average\ State\ Anxiety\ Score = \frac{\sum_{i=0}^{N}(\sum_{i=0}^{20} S)}{N} \tag{6.10}$$

$N =$ *Total number of players*

$S =$ *Score per statement in the questionnaire*

2. Metrics from the MEEGA+ model: The MEEGA+ Model does not use a particular set of formulas, but we have calculated the percentage of players who have agreed to a specific factor in the game. For example, if players were asked in the questionnaire, "I had fun playing the game," then the percentage of the players who gave 4 or 5 out of 5 was considered.

6.4.3 Results

In this section, we present the results for both phases of evaluation across all three age groups. First, the evaluation of the COVID-19 learning among players is presented, followed by the results of the evaluation of the quality of the game.

Evaluation as a learning tool

University students(17-21 years):The comparison between ASD for pre-game questionnaire and in-game questions is represented in Fig. 6.9.

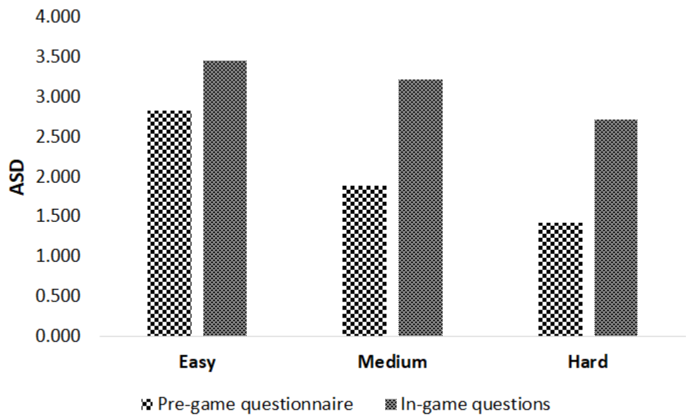


Figure 6.9: Average score comparison of players in the pre-game assessment with average score of players while answering the questions in the game. The average score was calculated for each difficulty level (Easy, Medium and Hard). There was a 22% increase in average scores for easy, 71% increase for medium and 90% increase for hard difficulty levels during the gameplay compared to their scores before playing the game.

1. Pre-Game Questionnaire: The average scores that we calculated were 2.18 out of a maximum score of 4 per question in the easy difficulty level. Medium difficulty level had an average score of 1.87 out of a maximum score of 5 per question. The hard difficulty level had a average score of 1.42 per question out of a maximum score of 6.

2. In-Game Questions: The average score during gameplay for easy difficulty level was found to be 3.44 per question out of a maximum score of 4. The medium level of difficulty had an average score of 3.20 per question out of a maximum score of 5 and the hard level of difficulty had an average score of 2.70 per question out of a maximum score of 6 points.

It can be observed that the average score per difficulty level keeps increasing. There has been an increase in the average score by 22% in the in-game questions compared to the pre-game questionnaire for the easy questions. For the medium questions, there has been an increase in average score by 71%, an increase of 90% for the hard questions, and an overall increase of 53%. The gap between the average scores of the pre-game assessment and the In-game assessment widens with the increase in difficulty level. The increase in gap

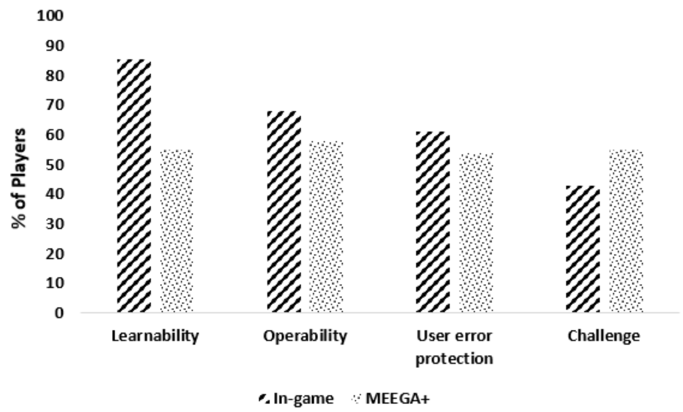


Figure 6.10: Phase 1: Factors evaluated using In-game analysis and MEEGA+ Model (a) The learnability of the game calculated through In-game analysis was 30.71% more compared to the MEEGA+ results (b) Operability calculated through In-game analysis was 10% more than the MEEGA+ results (c) User error protection calculated through In-game analysis was 7% more compared to the MEEGA+ results (d) Challenging level of the game to the player calculated through In-game analysis was 12% compared to the MEEGA+ results (e) There was increase in the average anxiety scores by 5.27% post-game compared to pre-game

shows that after attempting the questions multiple times and reading about the information related to those questions, the player learned about the information and eventually was able to answer the question. Combining all the difficulty levels, we can say that there has been an increase of player knowledge related to COVID-19 by 53%.

School children(10-16 years):The participants were 25 school students having majorly moderate game playing experience (Low=5, Moderate=18, High=7). This school was within the campus of our university. Around the time of data collection, these participants were having their sessional exams so could spare only 30 minutes of time. They were willing to play the game but did not want to fill the long questionnaire. But we did take notes for few of them based on their experiences after playing the game. This group of participants obtained an ASD of 3.27 in the easy difficulty level, 3.07 in the moderate difficulty level and 2.23 in the hard difficulty level of the game. This metric was calculated

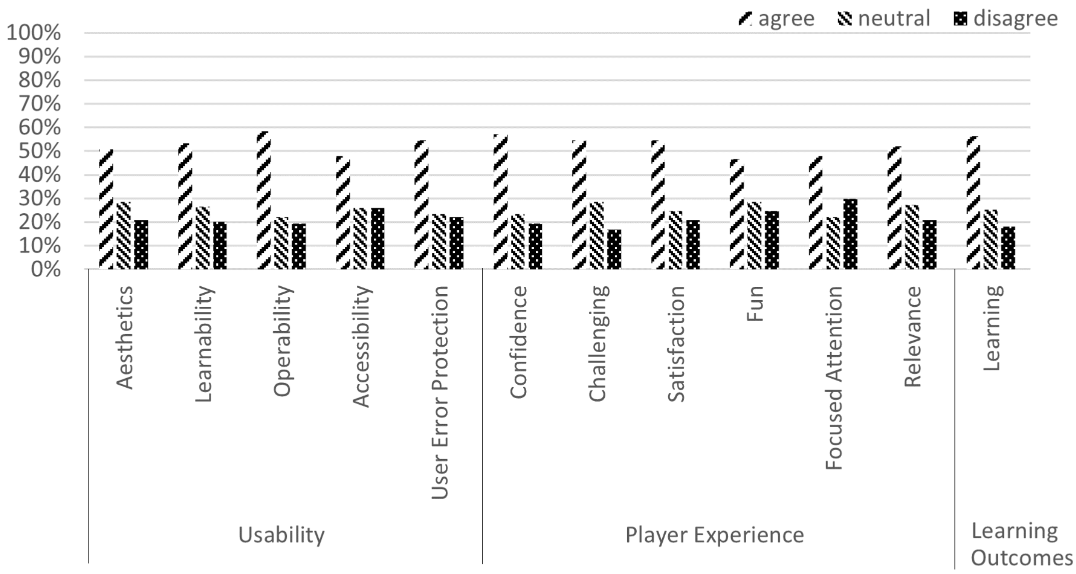


Figure 6.11: Phase 1Data Analysis of the responses received from the MEEGA+ Model questionnaire. The average number of player who found the game to be usable with good player experience and learning outcomes were 52.40%. While, 25.40% of the players were neutral about the game and 22% of the players did not find some of the elements of the game to be interactive or relevant.

from in-game analysis.

Working Professionals(30-40 years) and Elderly(41-50 years): Our third group for evaluation study comprised of 10 working professionals (downloads received from play store upon posting in research groups)in the age group of 30-40 years and 10 senior people in the age group of 41-50 years. As mentioned in the primer of Section 6.4, we got limited participants in this category. The elderly people refused to fill the MEEGA+ form and pre-game assessment form but were ready to tell us their experiences. They were less digitally skilled. They liked our idea of game. They could play the game upto a few levels only. The ASD for this group was 3.27 in the easy category, 3.12 in the moderate category and 2.65 in the hard category. This metric was calculated from in-game analysis.

Evaluation as a User-centered Game

University students(17-21 years): The overall results of the MEEGA+ model can be seen in Fig. 6.11, and the comparative analysis between the MEEGA+

model results and the in-game statistics can be seen in Fig. 6.10. The findings for the factors and the sub-dimensions under the MEEGA+ model have been stated below:

1. Usability:

   (a) Aesthetics: 51% of the players felt that the game was attractive and contained well-blended text font and colors.

   (b) Learnability: Using the in-game analysis, we have found that 85.71% of the players cleared the levels 0 and 1 in the first attempt 6.10. From the MEEGA+ model results in Fig. 6.11, 55% of the players felt that the game was easy to learn. The percentage of players who have learned the game quickly is more than the percentage of players who agree that the game was easy to learn. Although a majority of the MEEGA+ responses hold when compared to the in-game analysis, it can be seen that some of the players either could not realize that the game was easy to learn or did not report their experience. From this result, we can say that the game is easy to learn and has a good learnability factor.

   (c) Operability: From in-game analysis, it was found that 68.01% of the players cleared levels 1, 2, and 3 in the first attempt itself (Fig. 6.10). Since most of the players were able to clear levels 1, 2, and 3 in the first attempt, we can say that the game is easy to play and has a good operability factor. Later, when players were asked if the game was easy to play (operability), 58% of them agreed that the game was operable, and if we compare this with our in-game analysis, we can see that 68% of the players cleared levels 1, 2, and 3 in the first attempt (Fig. 6.10). The number of players who cleared levels 1, 2, and 3 in the first attempt is more than the number of players who agreed that the game was operable. Hence, it can be said that the game is operable and easy to use.

   (d) Accessibility: 48% of the players agreed that the game was accessible as it allows for customization of font and text according to their preferences. We can say that the game is moderately accessible as the number of players who agreed to the accessibility of the game was less than 50%.

(e) User Error Protection: From in-game analysis, it was found that the average percentage of sanitizer usage was 73%, and 61% of the players exceeded the average sanitizer usage throughout gameplay (Fig 6.10). Also, 54% of the players agreed that the game provided them a mechanism to recover from an error. Since most of the players who used sanitizers greater than the average usage and felt that the game had the means to rectify the mistakes, we can say that the game has a good user error protection factor.

2. Player Experience:

   (a) Confidence: 57% of the players felt confident and had a feeling of mastery after playing the game, which indicates that the game induces confidence among most of the players.

   (b) Challenge: Using the in-game analysis, we found that the average overall success rate was 36% and 43% of the players had a success rate less than the average success rate (Fig. 6.10). Since most of the players have a success rate greater than the average, we can say that the game is not very challenging to the players and might have found it to be of moderate difficulty. From the state anxiety questionnaire results, we can see that there has been a slight increase in the average state anxiety scores after the game compared to the pre-game state anxiety scores. The average score for the pre-game questionnaire was 48 out of 60, and the average post-game anxiety score was found to 50 out of 60. Although, in the MEEGA+ model, 55% of the players said they found the game challenging. We can infer that 12% of the players, although cleared most of the levels successfully (having high success rates), felt that the game was slightly challenging to clear.

   (c) Satisfaction: 55% of the players felt satisfied with the game and had a feeling of accomplishment of learning some information. Since most of the players felt satisfied with the information provided, we can say that the game has a good player satisfaction factor.

(d) Fun: 46.75% of the players found the game to be fun to play. The number of players who found the game to be fun is slightly less than 50%. Hence, we can say that the game has a moderate level of fun factor.

(e) Focused Attention: 48% of players said they were so immersed while playing the game that they lost track of time.

(f) Relevance: 52% found the game to be relevant in solving their problems, and they preferred learning through the game-based method rather than the traditional COVID-19 Awareness platforms. We can say that the game is relevant and is in line with the problem statement.

3. Learning Outcome: A detailed account of the game as a learning tool for the players was given earlier. But, when we asked the players if they gained knowledge throughout the game, 56% of the players felt that the game, *Unlock Me*, helped them learn about COVID-19 and 57% of the players felt that the game helped them remember information related to COVID-19.

When the average of the scores was taken for all sub-dimensions under the MEEGA+ questionnaire, it was found that 52.40% of the players found the game to be usable, had a good player experience, and felt that the game, *Unlock Me* helped them gain awareness about COVID-19.

School Children(10-16 years): In order to accommodate for the time constraint of these school children, we cut short the MEEGA+ questionnaire to include only those items that could be validated by our in-game analysis namely Learnability, Operability, User error protection and Challenge. The results of this have been presented in Fig. 6.12.

Working Professionals(30-40 years) and Elderly(41-50 years) Figure 6.13 shows the results. The learnability and operability of this group was less compared to the first two groups. The user error protection was also less thus explaining that these users did not use sanitizers effectively. The challenge was more than the other two groups. This directly relates to the fact that participants in this group have less experience of playing video games hence might have found it difficult to cope up with the increasing difficulty of levels. Additionally, there was only one elderly person who reached upto the last level.

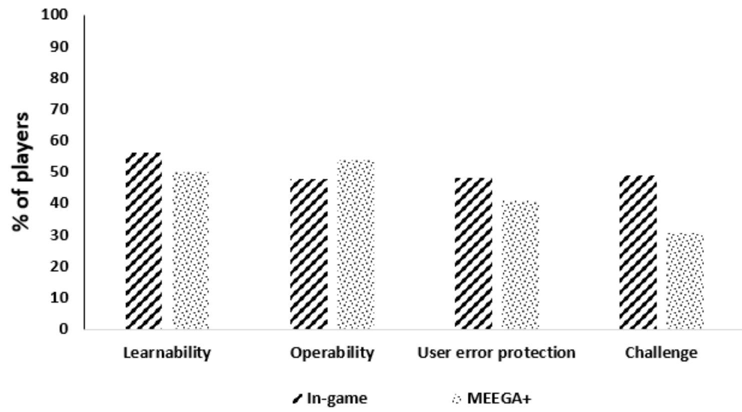


Figure 6.12: Factors evaluated using In-game analysis and MEEGA+ Model in the age group 10-16: The learnability of the game calculated through In-game analysis was 6.12% more compared to the MEEGA+ results. Operability calculated through In-game analysis was 6.15% less than the MEEGA+ results. User error protection calculated through In-game analysis was 7.20% more compared to the MEEGA+ results and Challenging level of the game to the player calculated through In-game analysis was 18.17% more compared to the MEEGA+ results.

6.5 Discussion

In this section, the limitations of this chapter and the possible approaches that could be used to address these limitations in the future is covered. The lessons learned from the entire research study and the ways in which this study can be helpful for other researchers is also described.

Unlock Me, a single-player game, was developed by considering the parameters like fantasy, mystery, goals, stimuli, challenge, and control. However, social interaction can be considered in the future to encourage coordination among other players digitally. For implementing social interaction, a multiplayer mode can be added. A study conducted on a serious game in both multiplayer and single-player modes indicates that the multiplayer game had a high player engagement rate

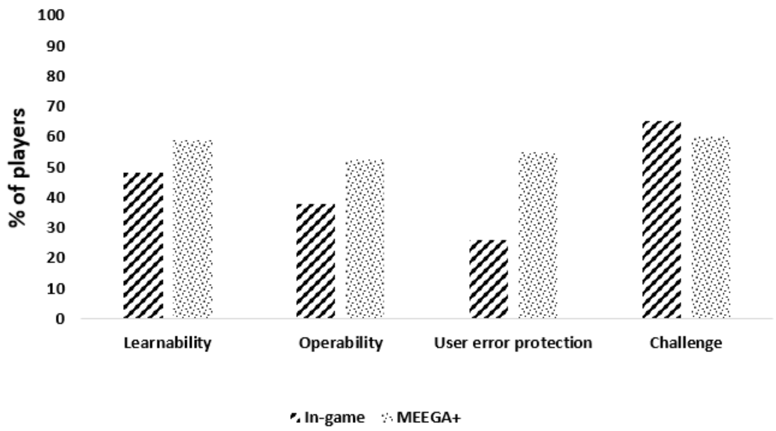


Figure 6.13: Factors evaluated using In-game analysis and MEEGA+ Model for working professionals and elderly. The learnability of the game calculated through In-game analysis was 10.62% less compared to the MEEGA+ results. Operability calculated through In-game analysis was 14.60% less than the MEEGA+ results. User error protection calculated through In-game analysis was 28.92% less compared to the MEEGA+ results and Challenging level of the game to the player calculated through In-game analysis was 5.21% more compared to the MEEGA+ results.

compared to the single-player mode [183]. Introducing a multiplayer mode would engage the players extensively and induce a competitive spirit among each player to compete against the other player. This feature can also act as another correlate of stress- social cognitive stress with high external validity as proposed by [57].

Currently, we have evaluated the game over a short time to test the COVID-19 learning of the player from the game. The average scores of the pre-game questionnaire were compared with the in-game questions based on easy, medium, and hard difficulty levels. The results for 17-21 years age group show an increase in learning by 53% over a short period of the experiment. However, to evaluate the long-term learning of the player, we can conduct another quiz with the same set of participants after a few weeks. The quiz can be conducted in the form of a simple online questionnaire. Then we can measure the overall memorability of the COVID-19 content among the players over a long period.

In-game analysis Authors in [184] have used in-game metrics for measuring overall game performance and maximum level reached. They have inferred the playing behaviour with an objective to measure the game's effectiveness in improving the learning on rational numbers. While calculating the MEEGA+ factors through the in-game analysis, the gaming behavior of the players was used to make an indirect inference. Quantitative measurement of these factors has not been done before. However, an attempt was made in this chapter to empirically represent the factors of the MEEGA+ model through in-game analysis. Another set of heuristics has to be derived to evaluate remaining factors in MEEGA+. The results of MEEGA+ and in-game analysis were extensively compared in the first evaluation phase. In the second evaluation phase, we could get only limited time to interact with the participants. Nevertheless, in-game analysis proved helpful in evaluating player experiences.

Additionally, in-game analysis has been used to improve the game design. As indicated in Table 6.1, this technique was used to understand game progression and analysis of the difficulty curve, which leads to game balancing. This technique is instrumental when the players are remote and direct observation is not feasible. It can reduce the time and effort of evaluating the user experience. In the future, to complement our study of the sub-dimensions under player experience, we can record the user's affective state with physiological signals.

Tutorial Through this study in this chapter, it was realized that the game's tutorial plays a crucial role in creating the desire for the users to play. It also helps the less experienced game players to master the game. At the same time, it turned out to be annoying for more experienced game players because they thought it was too simple. Most users reported that the tutorial was straightforward compared to the levels. The tutorial was made after most levels in the game were complete. Level 10 in *Unlock Me* is a survival mode that is different from the rest of the levels. The current implementation of Unlock-Me does not have instructions to clear level 10 in the tutorial. The knowledge of PLAY and GAP (Game Approachability Principles) can prove useful to design effective tutorials. [185]. Think aloud technique works with a small set of users. In our case, most of the users were remote; hence we relied on in-game analytics, which has the added benefit of being an unobtrusive measurement.

Role of game in learning The players get a sense of accomplishment upon completing levels in the game. The user receives both auditory and visual stimuli, which are far better than reading a piece of text in an article or a form. The player engrossed in the game pays more attention to anything that appears on the screen even if it is useless as (example ad) this can be used to communicate important things. Games have a positive impact on problem-solving skills and overall attentiveness. There are various peripheral activities associated with playing a game like pace, flow, immersion, engagement, game mechanics, sound track, sound effects, camera angle, narrative, and emotional connection.

6.6 Conclusion

This chapter describes the design and development of a freely available COVID-19 based mobile awareness game to help educate people about the virus. The game is conceptualized by adapting a user-centered methodology. This methodology includes defining the problem statement, ideation, design and deliver phases used to develop the game. Various features were included in the game to educate the user of certain COVID-19 norms and information. To evaluate the game, a survey was conducted among diverse age group participants and positive results were obtained. A two-fold evaluation method was used to evaluate the game as a learning tool and as a user-centered interface. Multiple factors of the standard game evaluation model MEEGA+, were measured using in-game analysis to assess the validity of this model.

Chapter 7

Utilizing WiFi sniffer traffic to classify stress related activities

7.1 Introduction

Stress can arouse in reaction to provocative stimuli. It can also occur when a person is challenged and unable to cope with excessive demands. In either case, the body's defensive mechanism gets activated. A frequent and long-term activation can cause chronic stress and health problems like poor digestion, increased weight, or sleep-related problems. Thus, stress must be detected timely, and the individual must be aware of its ill effects on health. The different methods to estimate stress are (i) subjective assessment using a self-reported stress questionnaire [186], (ii) objective assessment using an individual's cortisol hormone [187], or (iii) physiological assessment using custom-made or off-the-shelf wearable devices [188]. Method (i) is the most frequent subjective assessment method used in stress related studies. Method (ii) requires clinical intervention as the samples of the cortisol hormone need to be sent to the laboratory for assessment. Method (iii) is mostly performed in laboratory settings [189].

In order to complement the study on stress detection, unobtrusive monitoring of stress and its related real life indicators is a must [1]. A major real-life stress indicator is disturbed sleep. Stress manifests during the day and the night while a person is sleeping. It is well known that prolonged stress can harm sleep quality

[1]The content presented in this chapter is a revised version of my publication [46].

[190, 191, 192, 193]. Thus, disturbed sleep is also one of the ill effects of stress; hence, it must be measured and should drive the interventions to motivate people to sleep better.

The quality of sleep is essential to know that the sleep was disturbed or not. There has also been significant work on identifying sleep quality using phone activity. However, this has been done using smartphone sensors that drain much battery and hamper the person's privacy. Since sleep is also a human activity its non-intrusive measurement becomes important.

Previous research has indicated a correlation between smartphone usage before bedtime and sleep quality [194, 195]. Thus monitoring of smartphone usage gives a lens to look at sleep quality supported with ground truth from questionnaires. Several authors have approximated sleep duration by identifying smartphone usage patterns, but their work requires installing an app [196, 104]. These authors estimate different activities like: a person is using or not using the smartphone, sleeping or awake. The installed application has access to embedded sensors that infer the above activities. A major drawback with the above approach is that, users are hesitant to install different apps on their phones, and the sensors drain battery life. Thus, data collection becomes tough. Hence, in this chapter a device-agnostic solution to differentiate activities is proposed.

The device-agnostic approaches for detecting human activities utilized in the existing body of research rely on signals received from existing WiFi infrastructure, primarily Channel State Information (CSI) [197]. While efficient, CSI-based activity recognition either needs specialized modifications at the WiFi Access Points or needs to deploy USRP devices. Once the CSI data is collected, sophisticated signal processing approaches with optional conjunction with machine learning approaches are applied to recognize human activities. Either way, even though largely being device-agnostic, these approaches still need infrastructure changes.

The aim of this chapter is to develop a true device-agnostic system for human activity detection that does not encompass any client or access point changes. Another objective is to make the approach comfortable and affordable for the elderly population and those who suffer with chronic health condition. Thus, the mandate of wearing sensors on the body or requiring the installation of special sensors in the environment is relaxed. For this purpose, WiFi MAC-layer traffic collected by a passive sniffer is leveraged. Sniffers are easy to be configured with most commercial grade access points [198], end-user laptops [199], or even smartphones [200].

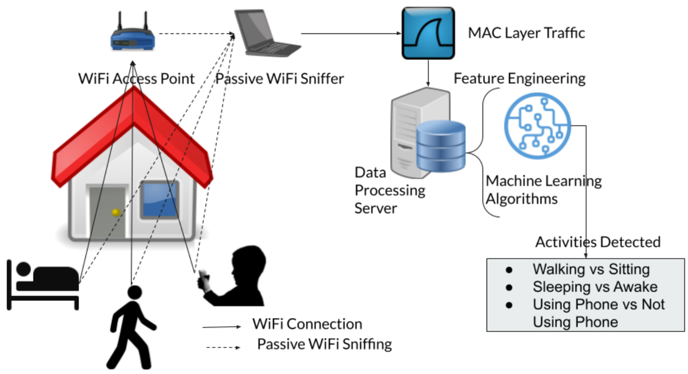


Figure 7.1: **System Architecture**: Human activities are detected in a home setting with off-the-shelf WiFi Access Point. A passive WiFi sniffer is deployed near the access point that captures MAC-layer traffic which is processed at a Data Processing Server. Post feature engineering, machine learning algorithms are executed to detect six different activities.

The initial proof-of-concept system, shown in Fig. 7.1, extracts special signatures from this traffic to distinctly identify the following simple human activities - Walking, Sitting, Sleeping, Awake, Using Phone and Not Using Phone. With data collected for 15+ hours, following machine learning models are trained: Logistic Regression, Support Vector Classifier (SVC), Random Forest, Gradient Boost, k-Nearest Neighbor (k-NN), and Convolutional Neural Network (CNN). First these models are evaluated for simple binary classification of the following sets of simple activities – (a) Walking vs Sitting, (b) Sleeping vs Awake, and (c) Using Phone vs Not Using Phone, and then these activities are interleaved. The evaluation reveals that WiFi MAC-layer traffic has special signatures for each of these activities. For the former case, the off-the-shelf machine learning algorithms can achieve up to 93% detection accuracy; while for the latter case they can achieve up to 82% detection accuracy. Lastly, the scenarios for complex activities are recreated that inherently leverage the simple activities. These include walking while using the phone vs walking without using the phone and sitting while using the phone and sitting without using the phone. The trained models achieve up to 82.86% detection accuracy for these complex activities. To the best of our knowledge, this is the first work on human activity recognition that is device-agnostic and is able to detect six distinct activities.

7.2 Activity Detection with Sniffer Traffic

7.2.1 System Details

System Overview: The current proof-of-concept system primarily works in a home environment, as shown in Figure 7.1. However, it can operate in production or enterprise WiFi environments as well. It is assumed that the home has a WiFi access point deployed (that most homes have in the present times). To maximize the likelihood of recording all MAC-layer traffic destined to the access point and transmitted by the access point, a sniffer is placed nearby the access point. The sniffer passively listens to the same frequency channel on which the WiFi access point is operating. The sniffer does not inject any active traffic into the traffic. MAC-layer traffic captured by the sniffer is sent to a data processing server for feature engineering and applying the machine learning algorithms.

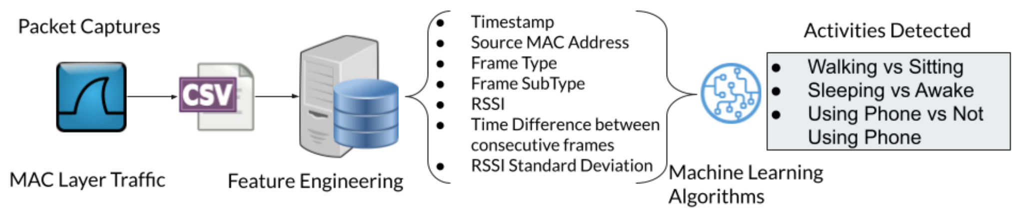


Figure 7.2: **System Flow**: Wireshark packet captures are converted to Comma-Separated-Value formatted file (CSV). Required features are extracted from the CSV and given as input to the Machine Learning algorithms to detect human activities.

Technical Details: The model of the access point used is Optronix RL821GWV that operates in 2.4 GHz and supports IEEE 802.11a/b/g/n. A USB WiFi adapter, TP-Link TL-WN722N that supports IEEE 802.11n is used as the passive sniffer with the help of monitor mode configuration. It is connected to a Lenovo Ideapad 130 Core i5 8th Gen - (8 GB RAM/1 TB HDD/Ubuntu 20.04.3 LTS) with Intel Core i5-8250U CPU. MAC-layer traffic is captured with command line utility `tshark`, version 3.2.3. The data processing server used is a Lenovo Ideapad 130 Core i5 8th Gen - (8 GB RAM/1 TB HDD/Ubuntu 20.04.3 LTS) with Intel Core i5-8250U CPU. The phone used for the data collection is a Xiaomi Redmi Note 5 Pro with Android Version 9.

7.2.2 Activities Detected

1. Walking and Sitting – Walking is defined as the movement of a subject carrying a phone or any other device connected to WiFi. The movement should be within the WiFi range in order to detect activities. The subject moved into the house for an average distance of 40 sq. meters. Sitting is defined as sitting while carrying your phone or any device connected to WiFi either in your pocket or hand and the device should be connected to WiFi.

2. Sleeping and Awake – Sleeping is defined as when the subject is asleep and not using the phone at all. However, the phone is connected to WiFi and kept by the side of the subject. A phone may or may not receive any notifications in this period of sleep. Awake is defined as when the subject is awake but not using the phone actively. The phone is in the pocket. In this period, the phone may be receiving external notifications that are app-based.

3. Using a Phone and Not Using a Phone— Using a phone is defined as the activity when a subject is actively using the phone for different apps installed on the phone. However, "Not Using the Phone", is defined as the subject not performing any activity on the phone. In both cases, the phone is connected to the WiFi network.

7.2.3 Dataset and Machine Learning Algorithms

Dataset Details: With sniffer we collect the MAC-layer traffic in the form of Wireshark Packet Captures. MAC-layer traffic follows the format of standard `IEEE` 802.11 WiFi frames [201]. Major frame categories are Data, Control, and Management Frames. These packet captures are then converted to a CSV formatted file, as shown in Figure 7.2. We filter the packet capture for following fields – (1) Timestamp, (2) Source MAC Address, (3) Frame Type, (4) Frame SubType, and (5) RSSI. While the above fields are captured in the packet capture, we also calculate two more fields that are fed as input to our machine learning models. These are – (6) Time Difference between Consecutive Frames – this feature is calculated by the difference in timestamp of two consecutive frames and (7) RSSI Standard Deviation – this denotes the variation in RSSI and is measured as the standard deviation of RSSI.

Our dataset is collected for three different scenarios:

a) Simple Activities: Here the activities Walking vs Sitting, Sleeping vs Awake, and Using Phone vs Not Using Phone are considered in a mutually exclusive way *i.e.* walking means that the subject is walking without using the phone and using phone means that the subject is actively using the phone without walking. The duration of data collection is as follows – Walking vs Sitting - 1 hour each, Sleeping vs Awake – 6 Hours each, and Using Phone vs Not using Phone - 2.5 hours each

b) Interleaved Activities: This scenario is closer to real-world where human activities are interleaved. For example, the subject follows activities in a particular order like – sitting → walking → sleeping → sitting → using a phone. The duration of data collection is as follows – Walking and Sitting – 1 hour each, Sleeping and Awake – 6 Hours each, and Using Phone and Not using Phone – 2.5 hours each.

c) Complex Activities: Here, the exclusivity is a bit relaxed where we move one more step closer to the real world.The subject is sitting while using the phone, sitting while not using the phone, walking while using the phone, or walking while not using the phone. The duration of data collection is 15 minutes for each of the 4 cases.

Machine Learning Algorithms: Six machine learning algorithms are leveraged – Logistic Regression, Support Vector Classifier, Random Forest, Gradient Boost, K-Nearest Neighbors, and Convolutional Neural Network. Each algorithm operates on the same CSV dataset, as shown in Figure 7.2. The performance of each of these algorithms is compared. The datasets are divided into training (70%) and test sets (30%) and supervised learning is performed. Our goals of classification depend on the three scenarios mentioned above – (a) Simple activities, (b) Interleaved activities, and (c) Complex activities. For scenario (a), binary classification is performed where two labels are considered – '0' and '1'. '0' signifies activity 1 and '1' signifies activity 2. For scenario (b), a multi-class classification is done where six different activities are classified – '0', '1', '2', '3', '4', and '5'. For scenario (c), again a multi-class classification is performed where classification is done for 4 different activities – '0', '1', '2' and '3'.

Feature Engineering: With the help of data analysis and domain knowledge, the following activity specific insights are identified that can be learned from the MAC-layer traffic. These are explained below:

1. Walking vs Sitting: The feature set for training the model in this case consists of Frame SubType, RSSI, Time Difference between consecutive frames and RSSI Standard Deviation. Main distinguishing feature for these activities is RSSI and its variation. It follows the inverse square law of wireless networks – as a client moves farther and closer to the access point to which it is connected, the RSSI will decrease and increase, respectively,

2. Sleeping vs Awake: The feature set for training the model consists of Frame SubType, RSSI and Time Difference between consecutive frames. The main feature is the time difference between consecutive frames. When a subject is sleeping the phone usage is minimal or no usage at all; for example, extremely few notifications, no phone calls, or texts. As a result of which, the WiFi traffic is considerably reduced; while the case is reversed when the person is awake. When the person is awake and not using the phone actively, it will still be receiving WiFi traffic due to more app-based notifications.

3. Using Phone vs Not Using Phone: The feature set for training the model is same as the prior one *i.e.* it consists of Frame SubType, RSSI and Time Difference between consecutive frames. Unlike, sleeping when phone usage is minimal when a person is awake and actively using the phone all types of MAC-layer frames are frequently exchanged between the access point and the client. Therefore, both the subtype of the frame, for example, a probe

request or a data frame, and the time difference between consecutive frames constitute significant features to identify if a subject is actively using the phone.

For Interleaved activities the feature set is same as Simple activities. However, for Complex activities the feature set for training the model consists of Frame SubType, Time Difference between consecutive frames and RSSI Standard Deviation. In this case, the feature set is carefully chosen so as to cover walking and sitting while using and not using the phone.

7.2.4 Performance Metrics

We consider following metrics to evaluate the trained models. Acronyms used are – True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

1. **Accuracy**: It is defined as the proportion of correct classifications in all activities. It is calculated as the fraction of the number of correct assessments to the number of all the assessments.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7.1}$$

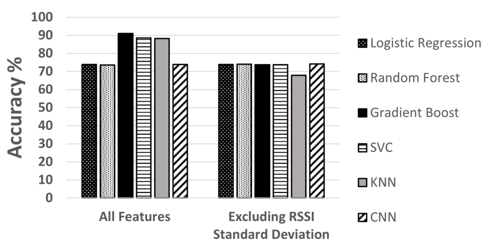
2. **F1 score:** It is defined as the harmonic mean of Recall and Precision.

$$F1\_score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{7.2}$$

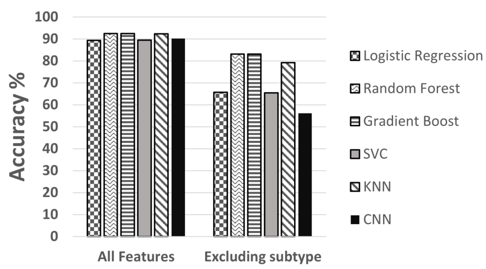
Here, Precision is defined as the fraction of true positives among all of the examples which are predicted to belong to a certain class – $Precision = \frac{TP}{TP+FP}$ and Recall is defined as the fraction of true positives among all the actual examples that truly belong to a certain class – $Recall = \frac{TP}{TP+FN}$.

7.3 System Evaluation

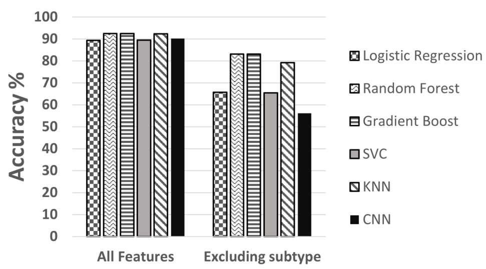
The training accuracy ranges from 74.36% to 92.26% and test accuracy ranges from 73.59% to 91.04%. Since the train and test accuracy are close to each other, hence the models are a good fit. As shown in Fig. 7.3a, the test accuracy of the



(a) Walking vs Sitting



(b) Sleeping vs Awake



(c) Using Phone vs Not

Figure 7.3: Accuracy of simple activity detection with a demonstration of most significant features. When a significant feature is missing, detection accuracy drops by up to 20%. Most significant feature for (a) is RSSI Standard Deviation, (b) is Time Difference between consecutive frames, and (c) is Frame Subtype.

classifiers to differentiate between Sitting and Walking activities is around 91.04% in the best case which is with Gradient Boost. To identify the most significant feature from its training set, we considered the hypothesis that there should be large variations in RSSI when the subject is walking vs when it is sitting. This hypothesis was validated when we removed RSSI deviation from the feature set. The accuracy drops to around 70%.

Next, for Sleeping vs Awake activities the training accuracy ranges from 81.21% to 87.04% and test accuracy ranges from 81.90% to 88.54%. In this case also the train and test accuracy are almost similar, hence our model is a good fit. As shown in Fig. 7.3b, the test accuracy of the classifiers is around 88.54% in the best case which is Gradient Boost. To identify the most significant feature from its training set, we considered the hypothesis that when a subject is sleeping its phone is likely to transmit/receive less WiFi traffic and if it is awake it's phone is likely to transmit/receive more WiFi traffic. Hence, there will be a difference in the time interval between consecutive WiFi frames. This hypothesis was validated when we removed Time Difference from the feature set. The accuracy drops around 5% for Gradient Boost and around 25% for k-NN .

Lastly, for Using Phone vs Not Using Phone the training accuracy ranges from 90.31% to 93.63% and test accuracy ranges from 56.04% to 92.49%. The average test accuracy is less because of CNN. As shown in Fig. 7.3c, the test accuracy of the classifiers to differentiate between Sleeping and Awake activities is around 92.49% in the best case which is for Gradient Boost and Random Forest. To identify the most significant feature from this set, we considered the hypothesis that there will be different type of MAC-layer frames when the subject is using phone vs when it is not using phone. Here, the frames ultimately translate to the app-based traffic. For example, the phone exchanges a lot of data frames when it is being actively used while it exchanges a lot of control and management frames when it is not actively used. Hence there will be a difference in the Frame SubType for the two classes. This hypothesis was validated when we removed Frame SubType from the feature set. The accuracy drops around 10% for both Gradient Boost and Random Forest.

Table 7.1 shows the average F1 Score, Recall, and Precision for all models.

F1 Scores are close to 1 for best performing models signifying that the models are able to avoid false positives and false negatives.

Table 7.1: Performance metrics for simple activity detection: F1 scores attempts to strike a balance between precision and recall when the data is not balanced.

Activity	**Metrics**	**Machine Learning Model**					
		Logistic Regression	Random Forest	Gradient Boost	SVC	KNN	CNN
Using Phone and Not Using Phone	F1 Score	0.88	0.89	0.89	0.85	0.89	0.91
	Recall	0.88	0.86	0.87	0.81	0.86	0.85
	Precision	0.90	0.93	0.90	0.90	0.93	0.97
Sleeping and Awake	F1 Score	0.84	0.88	0.89	0.84	0.88	0.82
	Recall	0.82	0.88	0.89	0.82	0.87	0.98
	Precision	0. 87	0.89	0.89	0.87	0.89	0.70
Walking and Sitting	F1 Score	0.75	0.89	0.89	0.69	0.87	0.82
	Recall	0.75	0.88	0.87	0.64	0.86	0.79
	Precision	0.76	0.90	0.91	0.74	0.88	0.85

7.3.1 Interleaved Activities

As shown Fig. 7.4 the accuracy ranges from 50% to 81.9%. Logistic regression does not perform as good as other models because it assumes linearity between dependent and independent variable. SVC algorithm does not perform well because it requires choosing a good kernel function which in

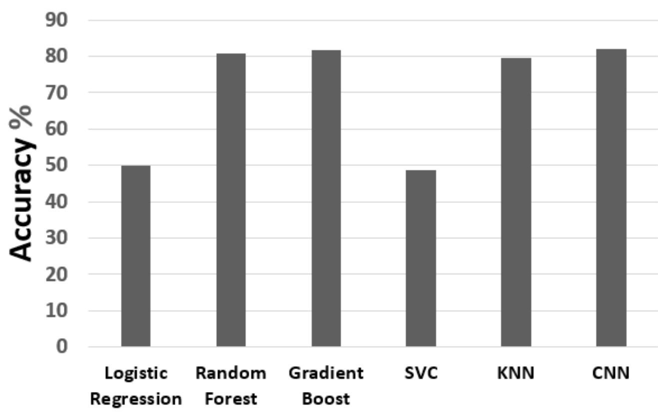


Figure 7.4: Accuracy of interleaved activity detection. The detection accuracy of Logistic Regression and SVC is ≈30% less than other algorithms.

turn involves hyperparameter tuning. Table 7.2 shows that the F1 score for classifying interleaved activities varies from 0.76 to 0.98 signifying a low rate of misclassifications.

7.3.2 Complex Activities

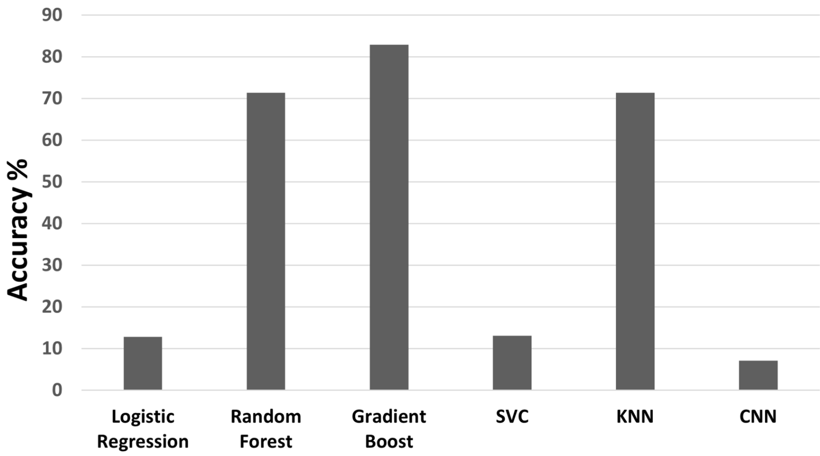


Figure 7.5: Accuracy of complex activity detection. Gradient Boost gives the best fit.

Table 7.2: Performance metrics for interleaved and complex activity detection

Activity	Performance Metrics	Machine Learning Model					
		Logistic Regression	Random Forest	Gradient Boost	SVC	KNN	CNN
Interleaved	F1 Score	0.76	0.94	0.95	0.92	0.87	0.98
	Recall	0.73	0.94	0.96	0.94	0.86	0.95
	Precision	0.80	0.94	0.95	0.91	0.88	0.97
Complex	F1 Score	0.03	0.71	0.97	0.04	0.81	0.96
	Recall	0.01	0.71	0.96	0.02	0.95	0.97
	Precision	0.12	0.71	0.98	0.76	0.71	0.93

The training accuracy for complex activity detection ranges from 10.97% to 99.16%. The test accuracy for this scenario ranges from 7.12% to 82.86%. Note that there is a difference of just 1.46% between train and test accuracy for

Gradient Boost model. From Fig. 7.5, it is evident that Gradient Boost has the best test accuracy, while Logistic Regression, SVC, and CNN perform poorly due to over-fitting and poor training accuracy.

Table 7.2 shows the F1 score varies from 0.03 to 0.97 signifying Gradient Boost performs correct classification.

7.4 Conclusion

The major contributions of this work are:

1. The first proof-of-concept system is presented to distinctly identify six simple human activities viz. Walking, Sitting, Sleeping, Awake, Using Phone and Not Using Phone. First, a dataset is collected for each of these activities exclusively and WiFi-MAC layer signatures are identified as features to recognize these activities. Next, a real-world scenario is recreated by introducing the activities in an interleaved manner. Lastly, leveraging the simple activities, scenarios of complex activities are created that involve using and not using the phone while walking and sitting [Section 7.2].

2. Six off-the-shelf machine learning algorithms – Logistic Regression, Support Vector Classifier, Random Forest, Gradient Boost, k Nearest Neighbors and Convolutional Neural Network are applied to the dataset collected. The performance of each of these algorithms is compared for each individual activity. The best-case results are as follows – (a) Classification accuracy for simple activities - Walking vs Sitting: 91.04%, Sleeping vs Awake: 88.54%, and Using Phone vs Not Using Phone: 92.49% , (b) Classification accuracy for interleaved activities is 81.9%, and (c) Classification accuracy for complex activities is 82.86% [Section 7.3].

Through this chapter, a proof of concept is done to establish if a device-agnostic approach can be developed to detect human activities. Sleep is also a human activity and there is a lot of correlation between sleep and stress. The work done in this chapter is our initial work and there are plans to expand this work by introducing a device-agnostic technique to detect sleep related stress and thereafter propose strategies for people to help them sleep better using an easy to use human computer interface.

Chapter 8

Conclusion and Future Research Directions

This thesis aimed to induce, detect and classify stress under various contexts and with different stress-inducing stimuli. With the advancement of technology, many psychology-based experiments to detect stress have been digitized. The major problem with these experiments is the accurate annotation of the collected user data. This annotation is referred to as ground truth which means the labeling of the participant's experience upon exposure to a stimulus. The problem of accurate ground truth labeling arises because, in most existing studies, the ground truth is merely a self-reported label. This thesis also considers the temporal variations of the participant's experience under the influence of stressful stimuli. Another focus of this thesis is to improve user engagement with the stress stimulus. The more the engagement, the closer the accuracy of the annotations. Thus, this thesis delved into multiple approaches to designing stress induction stimuli, multiple methods of detecting and validating the effect of the designed stimulus, and different techniques for classifying whether the person is stressed or not upon exposure to the stress stimulus.

8.1 Significant Contributions

Against the above backdrop, the following are the significant contributions through this thesis:

1. A data set of recorded EEG signals was created while the participants

were exposed to an emotional stressor. This emotional stressor was in the form of COVID news containing negative videos and movie clips containing emotionally inductive scenes. The EEG signals were acquired while participants watched the video clip as a stress elicitation material to infer the stressed state of the subject. Four groups of frequency-domain features were extracted from EEG signals. Using state-of-the-art machine learning and deep learning algorithms, these features were used to classify participants into stress and non-stress [Chapter 4].

2. An innovative stressor Color Word and Memory Test (CWMT) in the form of gamified mobile application was developed. This was inspired by well-known Stroop Test stressor. A data set of recorded EEG signals was created which were acquired while participants attempted Color Word and Memory Test (CWMT). A system to capture the complexity of EEG signals extracted from different brain regions was proposed to understand the region-specific neural changes due to stress. Different experiments were conducted to find significant difference between sub-groups i.e stressed vs not stressed using fractal dimensions extracted from EEG signal from different brain regions. Additionally, different statistical analysis techniques were presented to ascertain the validity of using HFD on EEG signals, in identifying biomarker for stress [Chapter 5].

3. An instrumented version of the educational game *Unlock Me* was created to record user actions through event-based logging. This instrumentation was done to validate the users' feedback based on self-report. The instrumented system was used to obtain in-game analytics which was in turn utilized to validate user experience after gameplay. This method can prove helpful in data-driven validation of the player experience. It does not require direct interaction with the players but still can be used to improve the game design. This improvement happened by understanding game progression and analyzing the difficulty curve, which led to game balancing. It can reduce the time and effort of evaluating the user experience [Chapter 6].

4. A device-agnostic system was proposed using WiFi to identify significant

mac layer signatures representing different human activities. A dataset was collected for each of these activities exclusively and significant WiFi MAC layer signatures are identified as features to recognize these activities. Next, a real-world scenario is recreated by introducing the activities in an interleaved manner. Lastly, leveraging the simple activities, scenarios of complex activities are created. Six off-the-shelf machine learning algorithms – Logistic Regression, Support Vector Classifier, Random Forest, Gradient Boost, k Nearest Neighbors and Convolutional Neural Network were used for classifying distinct human activities from the dataset collected. The evaluation reveals that WiFi MAC-layer traffic has special signatures to detect human activities and a detection accuracy of 92.49% was achieved in the best case [Chapter 7].

8.2 Future Research Directions

The several ways in which the work presented in this thesis can be extended are:

1. While calculating the MEEGA+ factors through the in-game analysis, the gaming behavior of the players was used to make an indirect inference. Quantitative measurement of these factors has not been done before. However, an attempt to empirically represent the factors of the MEEGA+ model through in-game analysis was done in this thesis. In order to calculate all factors and subdimensions under MEEGA+, other in-game measures need to be curated and analysed.

2. To establish the viability of the device-agnostic approach, its performance has to be evaluated in different WiFi networks.

3. We developed a proof of concept in Chapter 7, where we have identified unobtrusive methods of classifying a person as sleeping or awake. We will be extending this system to identify the activities that happen in a disturbed sleep.

4. To ensure environmental portability the machine learning models need to be agile to different environments. Thus to achieve this, more sophisticated machine learning approaches need to be explored.

5. A gamified mobile application to record the activities a person performs during the day can be used to annotate the ground truth and thus make a context-aware stress detection system. Further, this app can be used for stress management by tracking stress regularly and giving appropriate feedback to the user.

6. Since the non-linear feature, HFD can find a significant difference between changes in EEG signals for stressed and non-stressed sub-groups under time-based stress stimulus. It can be validated further with other stress-inducing stimuli and other stress detection methods apart from EEG.

7. Exploring other non-linear features like maximal Lyapunov exponent, correlation dimension or nonlinear interdependence.

8. One should further check if the changes of HFD coincide with corresponding relative changes of the power spectra, viz. the ratio between the power of slow and fast frequency bands, or, alternatively, the slope of the power spectra.

9. Various linear and non-linear features of finding changes in neural activity upon exposure to a stress stimulus can be undertaken and analyzed to present the most significant features for inferring the difference between stressed and non-stressed states.

Bibliography

[1] Kavallur Gopi Smitha, Ng Yu Xin, Seah Sze Lian, and Neethu Robinson. Classifying subjective emotional stress response evoked by multitasking using eeg. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3036–3041. IEEE, 2017.

[2] Eric Patterson. Stress facts and statistics.

[3] Sonal Khetarpal. Rising stress levels of india inc: Overwork, financial woes, remote work take toll on mental health.

[4] Mehdi Boukhechba, Alexander R Daros, Karl Fua, Philip I Chow, Bethany A Teachman, and Laura E Barnes. Demonicsalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health*, 9:192–203, 2018.

[5] Krešimir Ćosić, Siniša Popović, Marko Šarlija, Ivan Kesedžić, and Tanja Jovanovic. Artificial intelligence in prediction of mental health disorders induced by the covid-19 pandemic among health care workers. *Croatian medical journal*, 61(3):279, 2020.

[6] Yunji Liang, Xiaolong Zheng, and Daniel D Zeng. A survey on big data-driven digital phenotyping of mental health. *Information Fusion*, 52:290–307, 2019.

[7] Justin Hunt and Daniel Eisenberg. Mental health problems and help-seeking behavior among college students. *Journal of adolescent health*, 46(1):3–10, 2010.

[8] Neil Schneiderman, Gail Ironson, and Scott D Siegel. Stress and health: psychological, behavioral, and biological determinants. *Annual review of clinical psychology*, 1, 2005.

[9] Khalid Masood, Beena Ahmed, Jongyong Choi, and Ricardo Gutierrez-Osuna. Consistency and validity of self-reporting scores in stress measurement surveys. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4895–4898. IEEE, 2012.

[10] Syed Anwar, Sanay Saeed, Muhammad Majid, Saeeda Usman, Chaudhry Mehmood, and Wei Liu. A game player expertise level classification system using electroencephalography (eeg). *Applied Sciences*, 8(1):18, 2018.

[11] Sheldon Cohen, T Kamarck, R Mermelstein, et al. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10, 1994.

[12] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage Publications Sage CA: Los Angeles, CA, 2006.

[13] Charles D Spielberger. State-trait anxiety inventory. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[14] Pedro Sanches, Kristina Höök, Elsa Vaara, Claus Weymann, Markus Bylund, Pedro Ferreira, Nathalie Peira, and Marie Sjölinder. Mind the body! designing a mobile stress management application encouraging personal reflection. In *Proceedings of the 8th ACM conference on designing interactive systems*, pages 47–56, 2010.

[15] Alexandros Liapis, Christos Katsanos, Dimitris Sotiropoulos, Michalis Xenos, and Nikos Karousos. Recognizing emotions in human computer interaction: studying stress using skin conductance. In *IFIP Conference on Human-Computer Interaction*, pages 255–262. Springer, 2015.

[16] Habib Yaribeygi, Yunes Panahi, Hedayat Sahraei, Thomas P Johnston, and Amirhossein Sahebkar. The impact of stress on body function: A review. *EXCLI journal*, 16:1057, 2017.

[17] Beth L Abramson, Terrence D Ruddy, RA Dekemp, Louise A Laramee, JF Marquis, and Rob SB Beanlands. Stress perfusion/metabolism imaging: a pilot study for a potential new approach to the diagnosis of coronary disease in women. *Journal of Nuclear Cardiology*, 7(3):205–212, 2000.

[18] Jane B Allendorfer, Rodolphe Nenert, Kathleen A Hernando, Jennifer L DeWolfe, Sandipan Pati, Ashley E Thomas, Neil Billeaud, Roy C Martin, and Jerzy P Szaflarski. Fmri response to acute psychological stress differentiates patients with psychogenic non-epileptic seizures from healthy controls–a biochemical and neuroimaging biomarker study. *NeuroImage: Clinical*, 24:101967, 2019.

[19] Richa Gupta, M Afshar Alam, and Parul Agarwal. Modified support vector machine for detecting stress level using eeg signals. *Computational Intelligence and Neuroscience*, 2020, 2020.

[20] Jorgen Serup, Gregor BE Jemec, and Gary L Grove. *Handbook of non-invasive methods and the skin*. CRC press, 2006.

[21] Maneesh Bilalpur, Seyed Mostafa Kia, Manisha Chawla, Tat-Seng Chua, and Ramanathan Subramanian. Gender and emotion recognition with implicit user signals. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 379–387. ACM, 2017.

[22] Aamir Arsalan, Muhammad Majid, Amna Rauf Butt, and Syed Muhammad Anwar. Classification of perceived mental stress using a commercially available eeg headband. *IEEE Journal of Biomedical and Health Informatics*, 23(6):2257–2264, 2019.

[23] Anum Asif, Muhammad Majid, and Syed Muhammad Anwar. Human stress classification using eeg signals in response to music tracks. *Computers in biology and medicine*, 107:182–196, 2019.

[24] Richard J Davidson, Paul Ekman, Clifford D Saron, Joseph A Senulis, and Wallace V Friesen. Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I. *Journal of personality and social psychology*, 58(2):330, 1990.

[25] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.

[26] Anum Asif, Muhammad Majid, and Syed Muhammad Anwar. Human stress classification using eeg signals in response to music tracks. *Computers in biology and medicine*, 107:182–196, 2019.

[27] Xiyuan Hou, Yisi Liu, Olga Sourina, Yun Rui Eileen Tan, Lipo Wang, and Wolfgang Mueller-Wittig. Eeg based stress monitoring. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 3110–3115. IEEE, 2015.

[28] Nishtha Phutela, Devanjali Relan, Goldie Gabrani, and Ponnurangam Kumaraguru. Eeg based stress classification in response to stress stimulus. In *International Conference on Artificial Intelligence and Speech Technology*, pages 354–362. Springer, 2021.

[29] Sanay Muhammad Umar Saeed, Syed Muhammad Anwar, Humaira Khalid, Muhammad Majid, and Ulas Bagci. Eeg based classification of long-term stress using psychological labeling. *Sensors*, 20(7):1886, 2020.

[30] Jong S Kim and Jeeyeon Kim. Pure midbrain infarction: clinical, radiologic, and pathophysiologic findings. *Neurology*, 64(7):1227–1232, 2005.

[31] John-Dylan Haynes and Geraint Rees. Decoding mental states from brain activity in humans. *Nature reviews neuroscience*, 7(7):523–534, 2006.

[32] Mary Oliver. Towards an understanding of neuroscience for science educators. *Studies in Science Education*, 47(2):211–235, 2011.

[33] Ssang-Hee Seo, Jung-Tae Lee, and Marius Crisan. Stress and eeg. *Convergence and hybrid information technologies*, 27, 2010.

[34] Raymundo Cassani, Mar Estarellas, Rodrigo San-Martin, Francisco J Fraga, and Tiago H Falk. Systematic review on resting-state eeg for alzheimer's disease diagnosis and progression assessment. *Disease markers*, 2018, 2018.

[35] Robert Horlings, Dragos Datcu, and Leon JM Rothkrantz. Emotion recognition using brain activity. In *Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing*, pages II–1, 2008.

[36] Giorgos Giannakakis, Dimitris Grigoriadis, and Manolis Tsiknakis. Detection of stress/anxiety state from eeg features during video watching. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6034–6037. IEEE, 2015.

[37] JF Alonso, S Romero, MR Ballester, RM Antonijoan, and MA Mañanas. Stress assessment based on eeg univariate features and functional connectivity measures. *Physiological measurement*, 36(7):1351, 2015.

[38] Inga Dziembowska, Paweł Izdebski, Anna Rasmus, Janina Brudny, Marta Grzelczak, and Piotr Cysewski. Effects of heart rate variability biofeedback on eeg alpha asymmetry and anxiety symptoms in male athletes: A pilot study. *Applied psychophysiology and biofeedback*, 41(2):141–150, 2016.

[39] Richard S Lazarus and Susan Folkman. *Stress, appraisal, and coping.* Springer publishing company, 1984.

[40] Hans Selye. *The stress of life.* 1956.

[41] Houtan Jebelli, Mohammad Mahdi Khalili, and SangHyun Lee. Mobile eeg-based workers' stress recognition by applying deep neural network. In *Advances in Informatics and Computing in Civil and Construction Engineering*, pages 173–180. Springer, 2019.

[42] VV Nemets, VI Shmurak, VE Sobolev, VV Garnuk, ED Rovan, and EP Vinogradova. Effects of transient and prolonged uncontrollable stress on animals of dominant and subordinate social status with different types

of stress reactions. *Neuroscience and Behavioral Physiology*, 50(5):618–624, 2020.

[43] Eduardo Dias-Ferreira, João C Sousa, Irene Melo, Pedro Morgado, Ana R Mesquita, João J Cerqueira, Rui M Costa, and Nuno Sousa. Chronic stress causes frontostriatal reorganization and affects decision-making. *Science*, 325(5940):621–625, 2009.

[44] Lars Schwabe, Oliver T Wolf, and Melly S Oitzl. Memory formation under stress: quantity and quality. *Neuroscience & Biobehavioral Reviews*, 34(4):584–591, 2010.

[45] Nishtha Phutela. Measuring stress appraisal through game based digital biomarkers. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*, pages 403–404, 2021.

[46] Harshit Grover, Dheryta Jaisinghani, Nishtha Phutela, and Shivani Mittal. Ml-based device-agnostic human activity detection with wifi sniffer traffic. In *2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 72–77. IEEE, 2022.

[47] Nishtha Phutela, Devanjali Relan, Goldie Gabrani, Ponnurangam Kumaraguru, and Mesay Samuel. Stress classification using brain signals based on lstm network. *Computational Intelligence and Neuroscience*, 2022, 2022.

[48] Nishtha Phutela, Atluri Nikhitha Chowdary, Shreyansh Anchlia, Dheryta Jaisinghani, and Goldie Gabrani. Unlock me: A real-world driven smartphone game to stimulate covid-19 awareness. *International Journal of Human-Computer Studies*, page 102818, 2022.

[49] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[50] Jordan J Bird, Luis J Manso, Eduardo P Ribeiro, Anikó Ekárt, and Diego R Faria. A study on mental state classification using eeg-based brain-machine interface. In *2018 International Conference on Intelligent Systems (IS)*, pages 795–800. IEEE, 2018.

[51] Divya Acharya, Shivani Goel, Harshit Bhardwaj, Aditi Sakalle, and Arpit Bhardwaj. A long short term memory deep learning network for the classification of negative emotions using eeg signals. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[52] Mimansa Jaiswal, Cristian-Paul Bara, Yuanhang Luo, Mihai Burzo, Rada Mihalcea, and Emily Mower Provost. Muse: a multimodal dataset of stressed emotion. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1499–1510, 2020.

[53] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.

[54] Ömer Türk and Mehmet Siraç Özerdem. The convolutional neural network approach from electroencephalogram signals in emotional detection. *Concurrency and Computation: Practice and Experience*, page e6356, 2021.

[55] Gen Li, Chang Ha Lee, Jason J Jung, Young Chul Youn, and David Camacho. Deep learning for eeg data analytics: A survey. *Concurrency and Computation: Practice and Experience*, 32(18):e5199, 2020.

[56] Ping Wang, Aimin Jiang, Xiaofeng Liu, Jing Shang, and Li Zhang. Lstm-based eeg classification in motor imagery tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11):2086–2095, 2018.

[57] Nina M Ehrhardt, Julia Fietz, Johannes Kopf-Beck, Nils Kappelmann, and Anna-Katharine Brem. Separating eeg correlates of stress: Cognitive effort, time pressure, and social-evaluative threat. *European journal of neuroscience*, 2021.

[58] Houtan Jebelli, Mohammad Mahdi Khalili, Sungjoo Hwang, and S Lee. A supervised learning-based construction workers' stress recognition using a wearable electroencephalography (eeg) device. In *Construction research congress*, volume 2018, pages 43–53, 2018.

[59] Shiu Kumar, Alok Sharma, and Tatsuhiko Tsunoda. Brain wave classification using long short-term memory network based optical predictor. *Scientific reports*, 9(1):1–13, 2019.

[60] P Nagabushanam, S Thomas George, and S Radha. Eeg signal classification using lstm and improved neural network algorithms. *Soft Computing*, pages 1–23, 2019.

[61] Kostas M Tsiouris, Vasileios C Pezoulas, Michalis Zervakis, Spiros Konitsiotis, Dimitrios D Koutsouris, and Dimitrios I Fotiadis. A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals. *Computers in biology and medicine*, 99:24–37, 2018.

[62] Guangyi Zhang, Vandad Davoodnia, Alireza Sepas-Moghaddam, Yaoxue Zhang, and Ali Etemad. Classification of hand movements from eeg using a deep attention-based lstm network. *IEEE Sensors Journal*, 2019.

[63] Mohammad Naim Rastgoo, Bahareh Nakisa, Frederic Maire, Andry Rakotonirainy, and Vinod Chandran. Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications*, 138:112793, 2019.

[64] Pasquale Arpaia, Nicola Moccaldi, Roberto Prevete, Isabella Sannino, and Annarita Tedesco. A wearable eeg instrument for real-time frontal asymmetry monitoring in worker stress analysis. *IEEE Transactions on Instrumentation and Measurement*, 2020.

[65] Arnab Rakshit, Anwesha Banerjee, Ankita Mazumder, Poulami Ghosh, Anilesh Dey, and DN Tibarewala. Fractal analysis of eeg signals for studying the effect of cognitive stress on brain. *International Journal of Biomedical Engineering and Technology*, 25(2-4):336–369, 2017.

[66] Ali H Al-Nuaimi, Emmanuel Jammeh, Lingfen Sun, and Emmanuel Ifeachor. Higuchi fractal dimension of the electroencephalogram as a biomarker for early detection of alzheimer's disease. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2320–2324. IEEE, 2017.

[67] Maie Bachmann, Jaanus Lass, Anna Suhhova, and Hiie Hinrikus. Spectral asymmetry and higuchi's fractal dimension measures of depression electroencephalogram. *Computational and mathematical methods in medicine*, 2013, 2013.

[68] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[69] Seyedmajid Hosseini, Satya Katragadda, Ravi Teja Bhupatiraju, Ziad Ashkar, Christoph W Borst, Kenneth Cochran, and Raju Gottumukkala. A multi-modal sensor dataset for continuous stress detection of nurses in a hospital. *arXiv preprint arXiv:2108.07689*, 2021.

[70] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pages 400–408. ACM, 2018.

[71] Wei Chen, Shixin Zheng, and Xiao Sun. Introducing mdpsd, a multimodal dataset for psychological stress detection. In *CCF Conference on Big Data*, pages 59–82. Springer, 2020.

[72] Elsbeth Turcan and Kathleen McKeown. Dreaddit: A reddit dataset for stress analysis in social media. *arXiv preprint arXiv:1911.00133*, 2019.

[73] Pasquale Arpaia, Nicola Moccaldi, Roberto Prevete, Isabella Sannino, and Annarita Tedesco. A wearable eeg instrument for real-time frontal asymmetry monitoring in worker stress analysis. *IEEE Transactions on Instrumentation and Measurement*, 69(10):8335–8343, 2020.

[74] Rui Wang, Weichen Wang, Alex DaSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–26, 2018.

[75] Varun Mishra, Gunnar Pope, Sarah Lord, Stephanie Lewia, Byron Lowens, Kelly Caine, Sougata Sen, Ryan Halter, and David Kotz. The case for a commodity hardware solution for stress detection. In *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers*, pages 1717–1728, 2018.

[76] Hande Hong, Chengwen Luo, and Mun Choon Chan. Socialprobe: Understanding social interaction through passive wifi monitoring. In *Proceedings of the 13th international conference on mobile and Ubiquitous systems: Computing, networking and services*, pages 94–103, 2016.

[77] Manpreet Kaur, Flora D Salim, Yongli Ren, Jeffrey Chan, Martin Tomko, and Mark Sanderson. Shopping intent recognition and location prediction from cyber-physical activities via wi-fi logs. In *Proceedings of the 5th Conference on Systems for Built Environments*, pages 130–139, 2018.

[78] Philip Scanlon and Alan F Smeaton. Using wifi technology to identify student activities within a bounded environment. In *European Conference on Technology Enhanced Learning*, pages 473–478. Springer, 2017.

[79] Shweta Ware, Chaoqun Yue, Reynaldo Morillo, Jin Lu, Chao Shang, Jayesh Kamath, Athanasios Bamis, Jinbo Bi, Alexander Russell, and Bing Wang. Large-scale automatic depression screening using meta-data from wifi infrastructure. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–27, 2018.

[80] Enrique Garcia-Ceja, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J Oedegaard, and Jim Tørresen. Mental health monitoring with

multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, 51:1–26, 2018.

[81] Jeremy F Huckins, Alex W DaSilva, Rui Wang, Weichen Wang, Elin L Hedlund, Eilis I Murphy, Richard B Lopez, Courtney Rogers, Paul E Holtzheimer, William M Kelley, et al. Fusing mobile phone sensing and brain imaging to assess depression in college students. *Frontiers in neuroscience*, 13:248, 2019.

[82] Kasthuri Jayarajah, Zaman Lantra, and Archan Misra. Fusing wifi and video sensing for accurate group detection in indoor spaces. In *Proceedings of the 3rd International on Workshop on Physical Analytics*, pages 49–54, 2016.

[83] Jin Lu, Chao Shang, Chaoqun Yue, Reynaldo Morillo, Shweta Ware, Jayesh Kamath, Athanasios Bamis, Alexander Russell, Bing Wang, and Jinbo Bi. Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–21, 2018.

[84] John Rooksby, Alistair Morrison, and Dave Murray-Rust. Student perspectives on digital phenotyping: The acceptability of using smartphone data to assess mental health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[85] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 3–14, 2014.

[86] Rui Wang, Andrew T Campbell, and Xia Zhou. Using opportunistic face logging from smartphone to infer mental health: challenges and future directions. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of*

the 2015 ACM International Symposium on Wearable Computers, pages 683–692, 2015.

[87] Fifth Edition et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, 21(21):591–643, 2013.

[88] Soud Nassir and Tuck Wah Leong. Traversing boundaries: Understanding the experiences of ageing saudis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6386–6397, 2017.

[89] Madhura Mestry, Jigesh Mehta, Ajit Mishra, and Kiran Gawande. Identifying associations between smartphone usage and mental health during depression, anxiety and stress. In *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*, pages 1–5. IEEE, 2015.

[90] Luca Canzian and Mirco Musolesi. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 1293–1304, 2015.

[91] Asif Salekin, Jeremy W Eberle, Jeffrey J Glenn, Bethany A Teachman, and John A Stankovic. A weakly supervised learning framework for detecting social anxiety and depression. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2(2):1–26, 2018.

[92] Yoshihiko Suhara, Yinzhan Xu, and Alex'Sandy' Pentland. Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In *Proceedings of the 26th International Conference on World Wide Web*, pages 715–724, 2017.

[93] Davide Carneiro, Paulo Novais, Juan Carlos Augusto, and Nicola Payne. New methods for stress assessment and monitoring at the workplace. *IEEE Transactions on Affective Computing*, 10(2):237–254, 2019.

[94] Damon Daylamani-Zad, Fotios Spyridonis, and Kamal Al-Khafaaji. A framework and serious game for decision making in stressful situations; a fire

evacuation scenario. *International Journal of Human-Computer Studies*, 162:102790, 2022.

[95] Shrey Bagroy, Ponnurangam Kumaraguru, and Munmun De Choudhury. A social media based index of mental well-being in college campuses. In *Proceedings of the 2017 CHI Conference on Human factors in Computing Systems*, pages 1634–1646, 2017.

[96] Andrew G Reece, Andrew J Reagan, Katharina LM Lix, Peter Sheridan Dodds, Christopher M Danforth, and Ellen J Langer. Forecasting the onset and course of mental illness with twitter data. *Scientific reports*, 7(1):1–11, 2017.

[97] Mike Conway and Daniel O'Connor. Social media, big data, and mental health: current advances and ethical implications. *Current opinion in psychology*, 9:77–82, 2016.

[98] Gabriella M Harari, Samuel D Gosling, Rui Wang, Fanglin Chen, Zhenyu Chen, and Andrew T Campbell. Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior*, 67:129–138, 2017.

[99] Hande Hong, Chengwen Luo, and Mun Choon Chan. Socialprobe: Understanding social interaction through passive wifi monitoring. In *Proceedings of the 13th international conference on mobile and Ubiquitous systems: Computing, networking and services*, pages 94–103, 2016.

[100] Mengyu Zhou, Minghua Ma, Yangkun Zhang, Kaixin SuiA, Dan Pei, and Thomas Moscibroda. Edum: classroom education measurements via large-scale wifi networks. In *Proceedings of the 2016 acm international joint conference on pervasive and ubiquitous computing*, pages 316–327, 2016.

[101] Chaoqun Yue, Shweta Ware, Reynaldo Morillo, Jin Lu, Chao Shang, Jinbo Bi, Jayesh Kamath, Alexander Russell, Athanasios Bamis, and Bing Wang. Fusing location data for depression prediction. *IEEE Transactions on Big Data*, 7(2):355–370, 2018.

[102] Kasthuri Jayarajah, Youngki Lee, Archan Misra, and Rajesh Krishna Balan. Need accurate user behaviour? pay attention to groups! In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 855–866, 2015.

[103] Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T Chittaranjan, Andrew T Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 351–360, 2012.

[104] Nicholas D Lane, Mashfiqui Mohammod, Mu Lin, Xiaochao Yang, Hong Lu, Shahid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell. Bewell: A smartphone application to monitor, model and promote wellbeing. In *5th international ICST conference on pervasive computing technologies for healthcare*, volume 10, 2011.

[105] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. Can your smartphone infer your mood. In *PhoneSense workshop*, pages 1–5, 2011.

[106] Gerald Bauer and Paul Lukowicz. Can smartphones detect stress-related changes in the behaviour of individuals? In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 423–426. IEEE, 2012.

[107] Yonatan Vaizman, Katherine Ellis, and Gert Lanckriet. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE pervasive computing*, 16(4):62–74, 2017.

[108] Raneem Alharthi, Rajwa Alharthi, Benjamin Guthier, and Abdulmotaleb El Saddik. Casp: context-aware stress prediction system. *Multimedia Tools and Applications*, 78(7):9011–9031, 2019.

[109] Martin Gjoreski, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. Monitoring stress with a wrist device using context. *Journal of biomedical informatics*, 73:159–170, 2017.

[110] Camellia Zakaria, Rajesh Balan, and Youngki Lee. Stressmon: Scalable detection of perceived stress and depression using passive sensing of changes in work routines and group interactions. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019.

[111] Amee Trivedi, Camellia Zakaria, Rajesh Balan, Ann Becker, George Corey, and Prashant Shenoy. Wifitrace: Network-based contact tracing for infectious diseases using passive wifi sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(1), mar 2021.

[112] Camellia Zakaria, Youngki Lee, and Rajesh Balan. Detection of social identification in workgroups from a passively-sensed wifi infrastructure. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), apr 2021.

[113] Kamran Ali, Mohammed Alloulah, Fahim Kawsar, and Alex X Liu. On goodness of wifi based monitoring of vital signs in the wild. *arXiv preprint arXiv:2003.09386*, 2020.

[114] Weixi Gu, Longfei Shangguan, Zheng Yang, and Yunhao Liu. Sleep hunter: Towards fine grained sleep stage tracking with smartphones. *IEEE Transactions on Mobile Computing*, 15(6):1514–1527, 2015.

[115] Paula Goolkasian Anne Marie Porter. Video games and stress: How stress appraisals and game content affect cardiovascular and emotion outcomes. *Frontiers in Psychology*, 10, 2019.

[116] Johannes Steinrücke, Bernard P Veldkamp, and Ton De Jong. Determining the effect of stress on analytical skills performance in digital decision games towards an unobtrusive measure of experienced stress in gameplay scenarios. *Computers in human behavior*, 99:144–155, 2019.

[117] Regan Lee Mandryk and Max Valentin Birk. The potential of game-based digital biomarkers for modeling mental health. *JMIR mental health*, 6(4):e13485, 2019.

[118] Martin Johannes Dechant, Julian Frommel, and Regan Mandryk. Assessing social anxiety through digital biomarkers embedded in a gaming task. In

Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–15, 2021.

[119] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. *Game analytics*. Springer, 2016.

[120] Abdinasir Ahmed Hassen. Game balancing. *Game Design & Development 2021*, 2021.

[121] Elizabeth Boyle, Thomas M. Connolly, and Thomas Hainey. The role of psychology in understanding the impact of computer games. *Entertainment Computing*, 2(2), 2011. Serious Games Development and Applications.

[122] Ann DeSmet, Dimitri Van Ryckeghem, Sofie Compernolle, Tom Baranowski, Debbe Thompson, Geert Crombez, Karolien Poels, Wendy Van Lippevelde, Sara Bastiaensens, Katrien Van Cleemput, Heidi Vandebosch, and Ilse De Bourdeaudhuij. A meta-analysis of serious digital games for healthy lifestyle promotion. *Preventive Medicine*, 69, 2014.

[123] Elizabeth A. Boyle, Thomas Hainey, Thomas M. Connolly, Grant Gray, Jeffrey Earp, Michela Ott, Theodore Lim, Manuel Ninaus, Claudia Ribeiro, and João Pereira. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 2016.

[124] Krista E. DeLeeuw and Richard E. Mayer. Cognitive consequences of making computer-based learning activities more game-like. *Computers in Human Behavior*, 27(5), 2011. 2009 Fifth International Conference on Intelligent Computing.

[125] John Dewey. Experience and education. In *The educational forum*, volume 50, pages 241–252. Taylor & Francis, 1986.

[126] Chen Gang, Baghaei Nilufar, Sarrafzadeh Abdolhossein, Manford Chris, Marshall Steve, and Court Gudrun. Designing games to educate diabetic children. In *Proceedings of the 23rd Australian Computer-Human*

Interaction Conference, OzCHI '11, New York, NY, USA, 2011. Association for Computing Machinery.

[127] Peter M Bingham, Thomas Lahiri, and Taka Ashikaga. Pilot trial of spirometer games for airway clearance practice in cystic fibrosis. *Respiratory Care*, 57(8), 2012.

[128] Ralf Dörner and Stefan Göbel and Wolfgang Effelsberg and Josef Wiemeyer. *Serious Games Foundations, Concepts and Practice.* Springer International Publishing, 2016.

[129] Beverley Cook and Philip Twidle. Increasing Awareness of Alzheimer's Disease through a Mobile Game. In *2016 International Conference on Interactive Technologies and Games (ITAG)*, 2016.

[130] Mohammad Syahmi Rosli, Nur Nabilah Abu Mangshor, Nurbaity Sabri, and Zaidah Ibrahim. Educational game as interactive learning for hurricane safety. In *2017 7th IEEE International Conference on System Engineering and Technology (ICSET)*, 2017.

[131] Saturnino Luz, Masood Masoodian, Raquel Rangel Cesario, and Manuel Cesario. Using a serious game to promote community-based awareness and prevention of neglected tropical diseases. *Entertainment Computing*, 15, 2016.

[132] Rosemary Garris, Robert Ahlers, and James E. Driskell. Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4), 2002.

[133] Sara de Freitas. Learning in immersive worlds: a review of game-based learning. Workingpaper, 2006. This is an Open Access report distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License (http://creativecommons.org/licenses/by-nc-nd/3.0/), which permits users to copy, distribute and transmit the work for non-commercial purposes providing it is properly cited.

[134] Esther Oprins, Gillian Visschedijk, Maartje Bakhuys Roozeboom, Mary Dankbaar, Wim Trooster, and Stephanie C. E. Schuit. The game-based learning evaluation model gem: Measuring the effectiveness of serious games using a standardised method. *Int. J. Technol. Enhanc. Learn.*, 7(4), 2015.

[135] Petri Giani, Gresse von Wangenheim Christiane, and Borgatto Adriano Ferreti. Meega+: An evolution of a model for the evaluation of educational games, 2016.

[136] Francisco Jonathan de Oliveira Araújo, Ligia Samara Abrantes de Lima, Pedro Ivo Martins Cidade, Camila Bezerra Nobre, and Modesto Leite Rolim Neto. Impact of sars-cov-2 and its reverberation in global higher education and mental health. *Psychiatry research*, 288:112977, 2020.

[137] Meltem Huri Baturay and Sacip Toker. Internet addiction among college students: Some causes and effects. *Education and Information Technologies*, 24(5):2863–2885, 2019.

[138] David N Greenfield. Treatment considerations in internet and video game addiction: A qualitative discussion. *Child and Adolescent Psychiatric Clinics*, 27(2):327–344, 2018.

[139] Min Chen, Ke Shen, Rui Wang, Yiming Miao, Yingying Jiang, Kai Hwang, Yixue Hao, Guangming Tao, Long Hu, and Zhongchun Liu. Negative information measurement at ai edge: A new perspective for mental health monitoring. *ACM Transactions on Internet Technology (TOIT)*, 22(3):1–16, 2022.

[140] W Dimpfel, F Schober, and M Spüler. The influence of caffeine on human eeg under resting condition and during mental loads. *The clinical investigator*, 71(3):197–207, 1993.

[141] Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.

[142] Omar AlZoubi, Rafael A Calvo, and Ronald H Stevens. Classification of eeg for affect recognition: an adaptive approach. In *Australasian Joint Conference on Artificial Intelligence*, pages 52–61. Springer, 2009.

[143] Aditi Sakalle, Pradeep Tomar, Harshit Bhardwaj, Divya Acharya, and Arpit Bhardwaj. A lstm based deep learning network for recognizing emotions using wireless brainwave driven system. *Expert Systems with Applications*, 173:114516, 2021.

[144] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. Learning representations from eeg with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*, 2015.

[145] Yuchae Jung and Yong Ik Yoon. Multi-level assessment model for wellness service based on human mental stress level. *Multimedia Tools and Applications*, 76(9):11305–11317, 2017.

[146] BS Raghavendra, D Narayana Dutt, Harsha N Halahalli, and John P John. Complexity analysis of eeg in patients with schizophrenia using fractal dimension. *Physiological measurement*, 30(8):795, 2009.

[147] Hamidreza Namazi and Ondrej Krejcar. Analysis of pregnancy development by complexity and information-based analysis of fetal phonocardiogram (pcg) signals. *Fluctuation and Noise Letters*, 20(04):2150028, 2021.

[148] Mirra Soundirarajan, Erfan Aghasian, Ondrej Krejcar, and Hamidreza Namazi. Complexity-based analysis of the coupling between facial muscle and brain activities. *Biomedical Signal Processing and Control*, 67:102511, 2021.

[149] Saeid Yazdi-Ravandi, Dorsa Mohammadi Arezooji, Nasrin Matinnia, Farshid Shamsaei, Mohammad Ahmadpanah, Ali Ghaleiha, and Reza Khosrowabadi. Complexity of information processing in obsessive-compulsive disorder based on fractal analysis of eeg signal. *EXCLI journal*, 20:462, 2021.

[150] Santo Banerjee, D Easwaramoorthy, and A Gowrisankar. Fuzzy multifractal analysis in ecg signal classification. In *Fractal Functions, Dimensions and Signal Analysis*, pages 119–128. Springer, 2021.

[151] Tomoyuki Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283, 1988.

[152] Milena Cukic, David Pokrajac, Miodrag Stokic, Vlada Radivojevic, Milos Ljubisavljevic, et al. Eeg machine learning with higuchi fractal dimension and sample entropy as features for successful detection of depression. *arXiv preprint arXiv:1803.05985*, 2018.

[153] Yisi Liu and Olga Sourina. Eeg-based subject-dependent emotion recognition algorithm using fractal dimension. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3166–3171. IEEE, 2014.

[154] Turker Tuncer, Sengul Dogan, and Abdulhamit Subasi. A new fractal pattern feature generation function based emotion recognition method using eeg. *Chaos, Solitons & Fractals*, 144:110671, 2021.

[155] Md AM Joadder, Siuly Siuly, E Kabir, Hua Wang, and Yanchun Zhang. A new design of mental state classification for subject independent bci systems. *IRBM*, 40(5):297–305, 2019.

[156] Qiang Wang and Olga Sourina. Real-time mental arithmetic task recognition from eeg signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2):225–232, 2013.

[157] Hamidreza Namazi, Erfan Aghasian, and Tirdad Seifi Ala. Fractal-based classification of electroencephalography (eeg) signals in healthy adolescents and adolescents with symptoms of schizophrenia. *Technology and Health Care*, 27(3):233–241, 2019.

[158] Reza Khosrowabadi, Chai Quek, Kai Keng Ang, Sau Wai Tung, and Michel Heijnen. A brain-computer interface for classifying eeg correlates

of chronic mental stress. In *The 2011 international joint conference on neural networks*, pages 757–762. IEEE, 2011.

[159] Esther Beraha, Jonathan Eggers, Catherine Hindi Attar, Stefan Gutwinski, Florian Schlagenhauf, Meline Stoy, Philipp Sterzer, Thorsten Kienast, Andreas Heinz, and Felix Bermpohl. Hemispheric asymmetry for affective stimulus processing in healthy subjects–a fmri study. 2012.

[160] Federica Gioia, Alberto Greco, Alejandro Luis Callara, and Enzo Pasquale Scilingo. Towards a contactless stress classification using thermal imaging. *Sensors*, 22(3):976, 2022.

[161] Linlin Yan, Yiqun Gan, Xu Ding, Jianhui Wu, and Hongxia Duan. The relationship between perceived stress and emotional distress during the covid-19 outbreak: Effects of boredom proneness and coping style. *Journal of anxiety disorders*, 77:102328, 2021.

[162] Arthur R Jensen and William D Rohwer Jr. The stroop color-word test: a review. *Acta psychologica*, 25:36–93, 1966.

[163] Emil Šiška. The stroop colour-word test in psychology and biomedicine. *Acta Universitatis Palackianae Olomucensis. Gymnica*, 32(1), 2002.

[164] Shari R Waldstein, Elizabeth A Bachen, and Stephen B Manuck. Active coping and cardiovascular reactivity: a multiplicity of influences. *Psychosomatic Medicine*, 59(6):620–625, 1997.

[165] Rajdeep Ghosh, Nabamita Deb, Kaushik Sengupta, Anurag Phukan, Nitin Choudhury, Sreshtha Kashyap, Souvik Phadikar, Ramesh Saha, Pranesh Das, Nidul Sinha, et al. Sam 40: Dataset of 40 subject eeg recordings to monitor the induced-stress while performing stroop color-word test, arithmetic task, and mirror image recognition task. *Data in Brief*, page 107772, 2022.

[166] Katie Seaborn and Deborah I Fels. Gamification in theory and action: A survey. *International Journal of human-computer studies*, 74:14–31, 2015.

[167] Luis Cabañero-Gomez, Ramon Hervas, Ivan Gonzalez, and Luis Rodriguez-Benitez. eeglib: A python module for eeg feature extraction. *SoftwareX*, 15:100745, 2021.

[168] Luis Cabañero, Ramón Hervás, Iván González, Jesús Fontecha, Tania Mondéjar, and José Bravo. Analysis of cognitive load using eeg when interacting with mobile devices. *Multidisciplinary Digital Publishing Institute Proceedings*, 31(1):70, 2019.

[169] Luis Cabañero, Ramón Hervás, Iván González, Jesús Fontecha, Tania Mondéjar, and José Bravo. Characterisation of mobile-device tasks by their associated cognitive load through eeg data processing. *Future Generation Computer Systems*, 113:380–390, 2020.

[170] Luis Cabañero, Ramón Hervás, José Bravo, Luis Rodríguez-Benitez, and Chris Nugent. Eeglib: computational analysis of cognitive performance during the use of video games. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–12, 2019.

[171] Luis Cabañero Gómez, Ramón Hervás, Iván González, and Vladimir Villarreal. Studying the generalisability of cognitive load measured with eeg. *Biomedical Signal Processing and Control*, 70:103032, 2021.

[172] Stefan Gradl, Markus Wirth, Nico Mächtlinger, Romina Poguntke, Andrea Wonner, Nicolas Rohleder, and Bjoern M. Eskofier. The stroop room: A virtual reality-enhanced stroop test. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, New York, NY, USA, 2019. Association for Computing Machinery.

[173] Santos B., Romão T., Dias A.E, and Centieiro P. eVision: A Mobile Game to Improve Environmental Awareness. In *Lecture Notes in Computer Science*, volume 8253. Springer, 2013.

[174] Akhila Sri Manasa Venigalla, Dheeraj Vagavolu, and Sridhar Chimalakonda. SurviveCovid-19 – A Game for Improving Awareness of Social Distancing and Health Measures for Covid-19 Pandemic, 2020.

[175] Yáñez-Gómez R., Cascado-Caballero, D., and Sevillano. Academic methods for usability evaluation of serious games: a systematic review. *Multimedia Tools and Applications*, 76, 2016.

[176] Anders Drachen, Magy Seif El-Nasr, and Alessandro Canossa. *Game Analytics: Maximizing the Value of Player Data.* Springer, 2013.

[177] Daniel Johnson, M John Gardner, and Ryan Perry. Validation of two game experience scales: the player experience of need satisfaction (pens) and game experience questionnaire (geq). *International Journal of Human-Computer Studies*, 118:38–46, 2018.

[178] Rogers Yvonne, Sharp Helen, and Preece Jenny. *Interaction design: beyond human-computer interaction.* John Wiley & Sons, 2011.

[179] Mehdi Khosrow-Pour. *Gamification: Concepts, Methodologies, Tools, and Applications (4 Volumes).* IGI Global, 2015.

[180] WHO. Coronavirus disease (COVID-19) advice for the public: Mythbusters. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters, 2020. Accessed: 2021-01-01.

[181] Robert Morris Stern, William J Ray, and Karen S Quigley. *Psychophysiological recording.* Oxford University Press, USA, 2001.

[182] Charles D. Spielberger. State-Trait Anxiety Inventory for Adults. https://www.mindgarden.com/145-state-trait-anxiety-inventory-for-adults, 1983. Accessed: 2021-02-07.

[183] Gilang Ridzwanda Esthian Gitarana, Chalimi Fithratu, Alfika Nursyahrinanto Muis, and Reza Darmakusuma. Analysis and evaluation of player engagement in serious education game using game refinement theory case study: Arithmatopia game. In *2020 6th International Conference on Interactive Digital Media (ICIDM)*, pages 1–5, 2020.

[184] Kristian Kiili and Harri Ketamo. Evaluating cognitive and affective outcomes of a digital game-based math test. *IEEE Transactions on Learning Technologies*, 11(2):255–263, 2018.

[185] Heather Desurvire and Magy Seif El-Nasr. Methods for game user research: studying player behavior to enhance game design. *IEEE computer graphics and applications*, 33(4):82–87, 2013.

[186] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. A global measure of perceived stress. *Journal of health and social behavior*, pages 385–396, 1983.

[187] Sally S Dickerson and Margaret E Kemeny. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130(3):355, 2004.

[188] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2009.

[189] Bert Arnrich, Cornelia Setz, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. What does your chair know about your stress level? *IEEE Transactions on Information Technology in Biomedicine*, 14(2):207–214, 2009.

[190] Avi Sadeh, Giora Keinan, and Keren Daon. Effects of stress on sleep: the moderating role of coping style. *Health Psychology*, 23(5):542, 2004.

[191] Shelley D Hershner and Ronald D Chervin. Causes and consequences of sleepiness among college students. *Nature and science of sleep*, 6:73, 2014.

[192] Amir Muaremi, Agon Bexheti, Franz Gravenhorst, Bert Arnrich, and Gerhard Tröster. Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors. In *IEEE-EMBS international conference on biomedical and health informatics (BHI)*, pages 185–188. IEEE, 2014.

[193] David Marcusson-Clavertz, Martin J Sliwinski, Orfeu M Buxton, Jinhyuk Kim, David M Almeida, and Joshua M Smyth. Relationships between daily stress responses in everyday life and nightly sleep. *Journal of Behavioral Medicine*, pages 1–15, 2022.

[194] Kathryn M Orzech, Michael A Grandner, Brandy M Roane, and Mary A Carskadon. Digital media use in the 2 h before bedtime is associated with sleep variables in university students. *Computers in human behavior*, 55:43–50, 2016.

[195] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. Wi-sleep: Contactless sleep monitoring via wifi signals. In *2014 IEEE Real-Time Systems Symposium*, pages 346–355, 2014.

[196] Pedro Ferreira, Pedro Sanches, Kristina Höök, and Tove Jaensson. License to chill! how to empower users to cope with stress. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pages 123–132, 2008.

[197] Jieming Yang, Yanming Liu, Zhiying Liu, Yun Wu, Tianyang Li, and Yuehua Yang. A Framework for Human Activity Recognition Based on WiFi CSI Signal Enhancement. *International Journal of Antennas and Propagation*, 2021, 2021.

[198] Cisco Catalyst 9130AX Series Access Point Getting Started Guide. https://www.cisco.com/c/en/us/td/docs/wireless/access_point/9130ax/quick/guide/ap9130ax-getstart.html, 2019. [Online; accessed 23-November-2021].

[199] ALFA Network AWUS036NHA. https://deviwiki.com/wiki/ALFA_Network_AWUS036NHA, 2012. [Online; accessed 23-November-2021].

[200] Android. tPacketCapture. https://play.google.com/store/apps/details?id=jp.co.taosoftware.android.packetcapture, 2015. [Online; accessed 23-November-2021].

[201] IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2020*.

Appendix A

Publications

The work conducted during the duration of this thesis has resulted in the following publications:

1. Phutela, N., et al. (2021, November). EEG Based Stress Classification in Response to Stress Stimulus. In International Conference on Artificial Intelligence and Speech Technology (pp. 354-362). Springer, Cham [28].

2. Phutela, N. (2021, October). Measuring Stress Appraisal Through Game Based Digital Biomarkers. In Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play (pp. 403-404) [45].

3. Phutela, N., et al. (2022). Stress Classification Using Brain Signals Based on LSTM Network. Computational Intelligence and Neuroscience, 2022 [47].

4. Phutela, N., et al. (2022). Unlock Me: A Real-World Driven Smartphone Game to Stimulate COVID-19 Awareness. International journal of human-computer studies, 164, 102818 [48].

5. Grover, H., Jaisinghani, D., Phutela, N., et al. (2022, January). ML-Based Device-Agnostic Human Activity Detection with WiFi Sniffer

Traffic. In 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS) (pp. 72-77). IEEE [46].

6. Phutela, N., et al., "Effectiveness of Higuchi Fractal Dimension in differentiating subgroups of stressed and non-stressed individuals". (Under review)

7. Phutela, N., et al., , "Adaptive cognitive load and stress detection system" (accepted for patent)