EVALUATION OF USER’S EMOTIONAL EXPERIENCE THROUGH NEUROLOGICAL AND PHYSIOLOGICAL MEASURES IN PLAYING SERIOUS GAMES

TARANNUM ZAKI

MILITARY INSTITUTE OF SCIENCE AND TECHNOLOGY

2021

EVALUATION OF USER'S EMOTIONAL EXPERIENCE THROUGH NEUROLOGICAL AND PHYSIOLOGICAL MEASURES IN PLAYING SERIOUS GAMES

TARANNUM ZAKI

(BSc Engg., MIST)

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF SCIENCE IN ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MILITARY INSTITUTE OF SCIENCE AND TECHNOLOGY

2021

APPROVAL OF BOARD OF EXAMINERS

A thesis titled "EVALUATION OF USER'S EMOTIONAL EXPERIENCE THROUGH NEUROLOGICAL AND PHYSIOLOGICAL MEASURES IN PLAYING SERIOUS GAMES" submitted by Tarannum Zaki, ID: 1017140008, Session: 2016-2017 has been accepted as satisfactory in partial fulfillment of the requirement of the degree of Master of Science in Computer Science and Engineering on 17 January 2021.

BOARD OF EXAMINERS

1. ______________________________ Chairman
Lt Col Muhammad Nazrul Islam, PhD
Associate Professor, Department of Computer Science and Engineering
Military Institute of Science and Technology, Dhaka

2. ______________________________ Ex-Officio
Brig Gen A B M Humayun Kabir
Senior Instructor and Head of the Department
Department of Computer Science and Engineering
Military Institute of Science and Technology, Dhaka

3. ______________________________ Member
Dr. Md. Mahbubur Rahman (Internal)
Professor, Department of Computer Science and Engineering
Military Institute of Science and Technology, Dhaka

4. ______________________________ Member
Dr. Md Forhad Rabbi (External)
Associate Professor, Department of Computer Science and Engineering
Shahjalal University of Science and Technology, Sylhet

<u>DECLARATION</u>

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis (fully or partially) has not been submitted for any degree or diploma in any university or institute previously.

____

Tarannum Zaki
Department of Computer Science and Engineering
Military Institute of Science and Technology
17 January 2021

ABSTRACT

The importance of evaluating user experience (UX) for different types of application is increasing gradually and so are the varieties of methods of UX evaluation. Among various application context, interactive and entertaining computing systems are one of the dominant application context these days. In order to keep the interactive computing and entertaining systems sustainable in business, satisfying the user needs is a must, while UX evaluation plays a vital role in this respect. Again, user experience of gaming context as a part of interactive entertaining system is largely impacted by users' emotions. Among various types of games, serious game is a particular type which provides some purpose along with common gaming entertainment; for example, a simulation game for healthcare training program. The existing literature shows that UX evaluation of serious games are mostly carried out using subjective methods (questionnaires, survey). Thus, the objectives of this thesis are *firstly,* to show how objective methods (neurological and physiological measures) can be used to infer users' emotional experience. *Secondly,* whether there exists any correlation between neurological and physiological measures. To attain these objectives, a machine learning-based approach considering both the neurological and physiological measures is proposed to evaluate the users' emotions (UX) while playing serious game. The proposed approach is simulated through an experimental study with 25 participants to evaluate UX of an educational serious game –'Programming Hero'. The finding of the study indicates that neurological and physiological measures can infer users' emotions as well as a correlation exists to some extent between neurological and physiological data while evaluating UX of serious games.

ACKNOWLEDGEMENTS

I am extremely grateful to the Almighty Creator for successfully completing the MSc Engg. thesis. My heartiest gratitude and deep respect will go to my supervisor *Lt Col Muhammad Nazrul Islam, PhD, Associate Professor, Department of Computer Science and Engineering (CSE), Military Institute of Science and Technology (MIST)* for his meticulous supervision and affectionate guidance throughout the entire period which has made this thesis a success. I am grateful to *Maj Gen Md Wahid-Uz-Zaman, ndc, aowe, psc, te, Commandant, MIST* and *Brig Gen A B M Humayun Kabir*, Head of the CSE Department, MIST for their great encouragement and motivation.

I am wholeheartedly grateful to the Department of Computer Science and Engineering (CSE) of Military Institute of Science and Technology (MIST) for providing their constant support during the thesis work. I am especially grateful to the subjects for their voluntary participation in the experimental study of this thesis. I would also like to acknowledge the contribution of the Information and Communication Technology division, Ministry of Posts, Telecommunications and Information Technology of People's Republic of Bangladesh for providing me the ICT fellowship to pursue this research.

Finally, I am thankful to my family for their relentless support and blessings. I am also grateful to all the members of my thesis committee for their valuable suggestions.

TABLE OF CONTENTS

LIST OF SYMBOLS

β	: Beta Symbol
$\sum$	: Summation Notation

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

HCI	: Human-Computer Interaction
UX	: User-Experience
DT	: Decision Tree
ANN	: Artificial Neural Network
SVM	: Support Vector Machine
KNN	: K-Nearest Neighbour
ML	: Machine Learning
RF	: Random Forest
LDA	: Linear Discriminant Analysis
PPG	: Photoplethysmography
EEG	: Electroencephalography
ECG	: Electrocardiography
MRI	: Magnetic Resonance Imaging
fMRI	: functional Magnetic Resonance Imaging
BPM	: Beat-Per-Minute

CHAPTER ONE

INTRODUCTION

This chapter comprises of the thesis background, motivation and problem statement, thesis objectives, methodological overview, thesis scope and organization of the thesis. Firstly, it provides a brief discussion on the thesis background to introduce the thesis topic. Next, it highlights the motivation and problem statement of the thesis. Then, the thesis objectives are presented followed by methodological overview and scope of the thesis. Finally, the organization of the remaining chapters are described.

1.1 Thesis Background

IT (information technology) systems have become a part and parcel of our day to day life in keeping pace with the technological progress that has been on the rise in this 21st century. Among these diverse range of IT systems, the quantity of interactive computing systems are increasing gradually to make human life comfortable in every aspect possible. These variety of IT systems have become closely bounded with our daily life starting from our regular social communication through online communication media to online financial transaction, online shopping, transportation using riding service apps, online food delivery service apps and many more that have made our life much easier.

Because of such grave importance of interactive computing systems in our daily life, the importance of studying how various systems can be developed to ensure successful interaction between human and computing systems are also on the rise. This study that concerns to what extent computing systems can be interactive with humans in order to full-fill user needs is known as Human-Computer Interaction (HCI) [1]. The primary goal of HCI is to improve interactive experience between user and system in such a way so that the system satisfies the required user needs completely as well as ensures user-friendly and comforting experience. The importance of interaction between user and computing systems is immense because without integrating the viewpoints of users a system can never persist to serve its purpose successfully among its intended users.

The utilization of an interactive computing system by the intended consumers not only depends on the users' desired requirements being full-filled thoroughly but is also highly dependent on how much the users value the system, their attitude and feelings towards using the system and such. In short, users' interaction or experience with a system bears great significance on whether the IT system would persist in the long run of technological advancement. Thus the means and easiness of user interaction greatly affect user experience (UX) in the field of human-computer interaction (HCI) which is defined as users' perspectives as well as both practical and affective responses resulting from using the system [2]. As defined by ISO 9241-210:2010 standard, user-experience is the overall perception of a user basing on their abilities and constraints towards using a system resulting from their attitude, emotion, behaviour, beliefs and psycho-physiological responses regarding the system's expected use [3]. A user experience of a system that is meaningful and gives positive vibes helps the system eventually to gain users' confidence towards using that system whereas a bad user experience leads a system to become outdated eventually [4]. Moreover, user experience also helps to conduct further improvements of a system so that its productivity increases. As a result, user experience is important for users so that they can use a system conveniently as well as important for the system designers and developers so that the system is able to sustain in the competitive business world.

User experience evaluation has become a very concerning issue these days because of its significance to keep an IT system stay in business. There are different kinds of evaluation measures to assess UX of a system which varies from subjective to objective assessment such as interviews, rating scales, surveys are some of the subjective measures and performance measure, physiological and neurological response etc. are some of the objective measures [5]. These evaluation measures are used to evaluate IT applications developed for various purpose on various software environments. For example, eye tracking, facial movements, electrocardiography and electrodermal activity are some objective measures that are being used for UX evaluation of online advertising websites, PC games and desktop-based advertising respectively [6, 7, 8]. However, among these variety of subjective and objective evaluation measures, objective measures are usually preferred than the latter one for providing better evaluation results [9].

1.2 Motivation and Problem Statement

Gaming experience is a widely evaluated context for UX evaluation in the field of entertainment industry [10]. The gaming context includes different types of games and among them serious game is one of the well-known gaming applications focused on teaching or training. More precisely, serious games are referred to the games which are aimed towards a particular purpose rather than pure entertainment usually used in education or semi-formal educational settings [11]. Again, emotion is one of the attributes that is largely affected in an entertainment-based computing environment like gaming experience [12]. UX evaluation of serious games are usually performed using subjective measures particularly by questionnaires [13]. Again, emotion inferring from objective methods like physiological and neurological response are mostly used to provide interactive experience rather than evaluating UX. It is thus quite obvious that some issues require further investigation regarding the applicability of objective measures particularly neurological and physiological measures for evaluating serious games basing on user's emotional experience. Therefore, the problem statements could be formulated as follows:

a. Emotions are mostly measured using objective methods (neurological and physiological measures) for evaluating the UX while subjective methods are usually disregarded in this aspect as emotion is accompanied by psycho-physiological reaction inherently.

b. Emotion-based UX evaluations are mostly adopted to evaluate the UX of gaming application because user emotion bears a vital impact while experiencing within a gaming environment.

c. In case of serious games, UX is evaluated predominantly using subjective methods; questionnaires to be specific rather than taking users' emotional experience into concern for UX evaluation by using objective methods (neurological and physiological measures).

d. Most of the educational serious games are focused to the computer programming domain and are evaluated using subjective methods whereas neurological and physiological measures (objective methods) are not used to derive users' emotional experience.

1.3 Thesis Objectives

The objectives of this thesis are thus associated with UX evaluation of an interactive computing environment and that is, a serious game application. The intention of this UX evaluation is to explore how users' emotional experience using objective methods such as neurological and physiological measures can be used to evaluate UX of a serious game application. In a broader perspective, this research covers the fields of *Human-Computer Interaction (HCI)* and *Serious Games*. In short, the objectives of this research are stated below:

Firstly, to show how neurological and physiological measures are used to evaluate the user's emotional experience in playing serious games.

Secondly, to explore the correlation between physiological and neurological data in assessing the user's emotional experience for serious games.

These research objectives will result in two expected outcomes. *Firstly,* a practical approach for evaluating user's emotional experience through physiological and neurological measures in playing serious games. *Secondly,* results regarding correlation between physiological and neurological data to evaluate user's emotional experience in playing serious games.

1.4 Methodological Overview

The methodology of this thesis follows an experimental procedure followed by an extensive systematic literature review which has been carried out in a well-structured way. The steps of this thesis can be broadly organized into four groups – systematic literature review, experiment design, experimental study and experimental data analysis. Figure 1.1 shows a process flow of the methodical organization as well as the purpose of each step are briefly discussed subsequently.

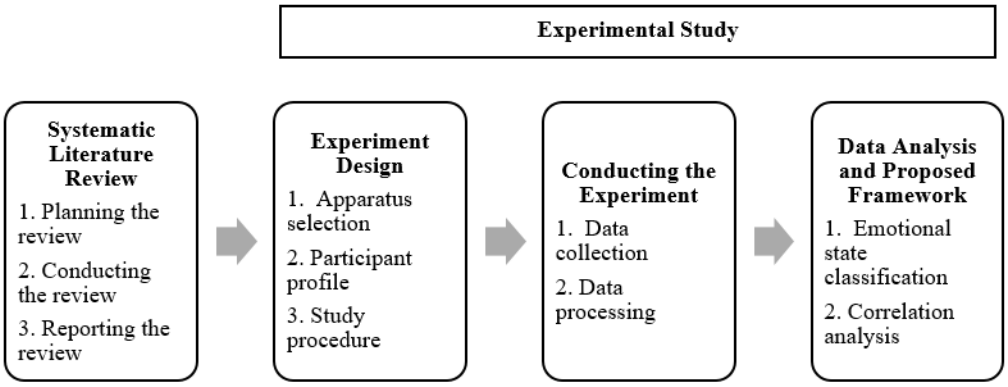


Fig. 1.1. Process flow of the research methodology

Systematic Literature Review: The steps of the systematic literature review consists of three broad phases; namely planning the review, conducting the review and reporting the review. At first, the review is planned following a search strategy, defining some inclusion and exclusion criteria, and finalizing the review materials. Then, the data are extracted to conduct the review. Lastly, the review findings along with the research gaps and future research opportunities are reported.

The experimental study comprises of three sequential phases - experimental design, conducting the experiment and, data analysis and proposed framework which are discussed as follows:

Experimental Design: This phase is concerned with designing the experiment in a systematic way which consists of selecting the required apparatus such as selecting an application context, defining emotional states and selecting evaluation measures; setting up the participant profile and defining the study procedure to conduct the experiment.

Conducting the Experiment: The experiment is conducted in this phase with the intended participants. The required data are collected using the selected evaluation measures for the selected application context and then processed. The processing involves preprocessing the data and selecting certain features for further analysis.

Data Analysis and Proposed Framework: Here, the data are analyzed for emotional state classification where a proposed extended framework using machine learning techniques is presented. Furthermore, a correlation analysis between the selected UX evaluation measure (neurological and physiological data) is performed in this phase.

1.5 Thesis Scope

The scope of this thesis can be defined from a number of perspectives. Assessment of interactive computing system can be performed basing on UX and usability parameters. The scope of this research is limited to only evaluating UX. Regarding UX evaluation methods, which are broadly classified as subjective and objective methods; the UX evaluation methods used for this research is confined to objective methods only. Among different objective methods, neurological and physiological measures are taken into concern which are primarily responsible to have large impact on users' emotional experience. Emotions play a very important role while users play games. Again, from contextual perspective of gaming environment, serious game is narrowed down from a variety of games such as arcade games, adventure games etc. Among different focuses of serious games; educational based, particularly a programming-based mobile application game within the computer science domain is selected to conduct the respective experimental study because of the wide use of such applications which are mostly evaluated using subjective methods.

1.6 Organization of the Chapters

The organization of the thesis for the remaining chapters is as follows:

Chapter 2: This chapter presents the 'Theoretical Background' which discusses the relevant concepts for the concerned research area. These concepts cover human-computer interaction (HCI), user-experience (UX) and usability, UX evaluation measures including neurological and physiological measures as objective methods as well as subjective methods, UX and emotion, gaming context and UX including serious games and, machine learning and emotion classification.

Chapter 3: This chapter presents the 'Related Work'. Here, the methodology of the systematic literature review is discussed elaborately followed by the outcomes of the review study as well as research gaps and future research opportunities. Finally, the focused research opportunity for this thesis is stated along with a critical summary to highlight the issues for which the particular research opportunity has been chosen for further research in this thesis.

Chapter 4: This chapter presents the 'Experimental Study' which is broadly distributed among the sequential phases - experiment design, conducting the experiment and, data analysis and proposed framework. Experiment design combines apparatus selection, participant profile and study procedure. After that, data collection and data processing is covered in conducting the experiment. Lastly, data analysis and proposed framework discusses the process of emotional state classification along with the proposed extended framework and also, a correlation analysis between the selected evaluation measures.

Chapter 5: This chapter presents the 'Discussions and Conclusions'. Finally, the thesis is concluded in this chapter with a summarized discussion of research outcomes and research implications. This chapter also includes certain limitations and possible future work of this research.

CHAPTER TWO

THEORETICAL BACKGROUND

This chapter briefly discusses some of the key concepts to provide basic theoretical knowledge regarding the background of this thesis. At first, a preliminary discussion on Human-Computer Interaction (HCI) is presented. Next, the characteristics of UX and usability are outlined followed by an elaborate discussion on UX evaluation measures. Then, the impact of UX on user emotion is explained. After that, the importance of UX in the context of gaming environment along with the classification of games are discussed. Lastly, machine learning techniques and its application for classifying emotions are described.

2.1 Human-Computer Interaction (HCI)

Human-computer interaction is the study of the design and use of computer technology that is focused on the interactive experiences between human and computer systems [14]. HCI has promoted the importance of cognitive science and human factors while designing a computing system. A system is always developed for its intended users and so satisfying user requirements in the most user-friendly and efficient approach is the primary concern of system design. To understand user requirements and how their acceptability towards a system works is largely dependent on different types of human factors (emotion, attitude, feelings and so on). These human factors help to improve a system's overall aspect including its features, functionality, context of use etc. as well as assures its success and persistence among the intended users. Considering the the effect of human factors along with the context of use of a system, the ISO 9241 has presented a framework regarding the lifecycle of the development of a human computing interactive system [15]. Figure 2.1 adapted from [15] shows that this lifecycle is a repeated process which initiates with identifying the need for user requirements, then understanding the purpose of use within a context, specifying user/organizational requirements, creating design and executing solutions, and finally performing the evaluation of the solutions to understand whether the system satisfies the user requirements. This evaluation of solutions is particularly concerned with UX and usability. As a result, the importance of HCI is on the rise as more and more interactive computing systems are being developed nowadays. HCI has introduced a number of measures to evaluate the level and quality of interaction between human and computing systems [16]. These measures of HCI are broadly classified between the principal parameters of HCI-based systems' evaluation procedure which are UX and usability. In short, these parameters of HCI measure is aimed to provide safe, usable, and efficient systems to everyone with utmost ease, comfort and user-friendliness attributes.

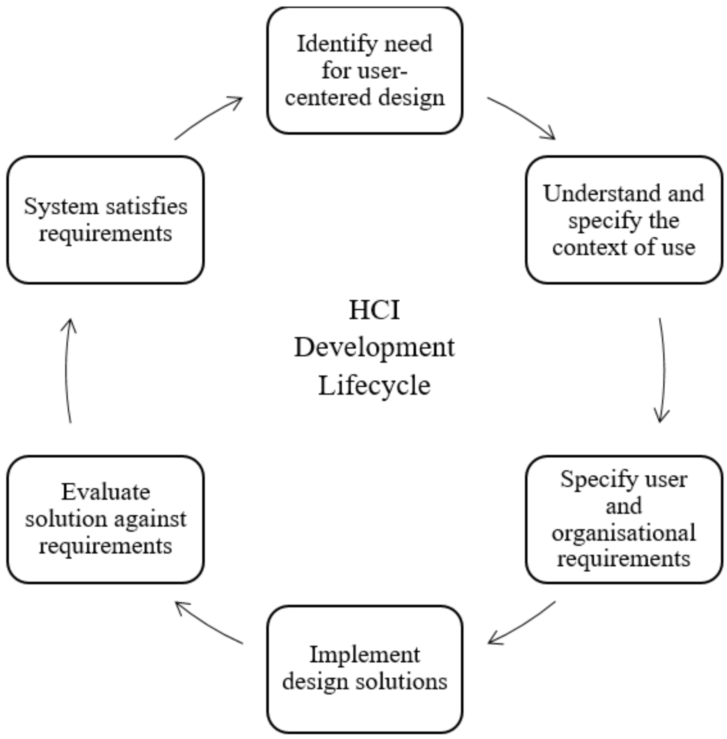


Fig. 2.1. The lifecycle framework of HCI development (ISO 9241)

2.2 User-experience (UX) and Usability

The broad field of HCI is defined by the three core aspects – design, evaluation and implementation [17]. Among these, design is a very crucial aspect in the development lifecycle of a system [18]. The design of a system is influenced by a number of factors where user requirement plays the most vital role [19, 20]. If the design of a system is not devised according to user requirements, then the system loses its acceptability from the users eventually. UX and usability are two of the parameters that are taken into concern very consciously while ensuring a system meets up the user requirements accordingly. The terms UX and usability are interrelated and often used interchangeably to measure whether the user requirements are satisfied but there lies some significant distinction between the two terms [21]. UX is concerned with the thorough experience of a user while they use a system that is more inclined towards their emotional views while usability is assessing the quality of their usage of a system basing on generally effectiveness, efficiency and satisfaction criteria [22]. The components of UX are displayed in Figure 2.2 adapted from [23] which indicates UX is grouped into three types of components broadly - perception that concerns instrumental qualities, users' reactions concerning emotions and perception that concerns qualities that are not instrumental. Instrumental qualities are defined as the attributes which are oriented with differnet types of tasks such as usefulness and utility. On the otherhand, some examples of qualities which are not instrumental are related to aesthetic features, aspects which are symbolic and motivational characteristics. The user reactions related to emotions are the mostly influenced attribute of UX evaluation which involves feelings (subjective), expressions that are motor driven, reactions caused by physiological activities and cognitive judgements. In contrast, the ISO usability framework is shown in Figure 2.3 adapted from [24] which illustrates the components of usability as well as the inter-connectivity between the product, context of use and intended goals.

In summary, UX is measured depending on achieving different hedonic goals such as human behavior, emotion and response that generates from an interactive experience, and usability is measured depending on achieving some sort of performance goals such as learnability, safety, accessibility and the likes [25]. Figure 2.4 specifies some properties of UX

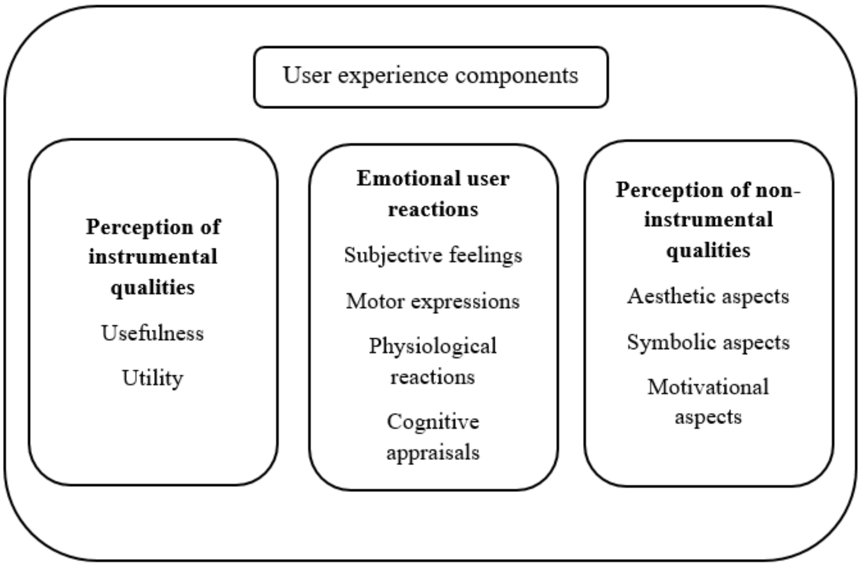


Fig. 2.2. Basic components of user experience (UX)

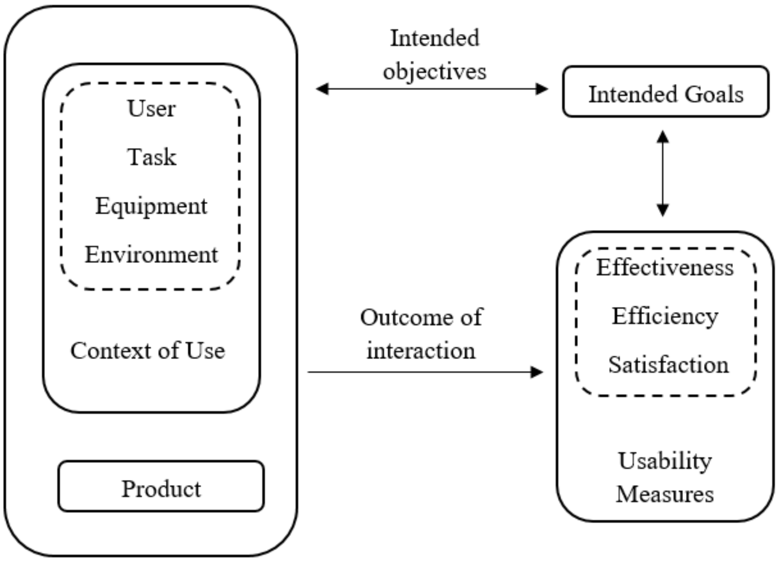


Fig. 2.3. The usability framework by ISO

and usability which are generally used to define these two terms [26]. Both UX and usability bears much importance while developing a system as it is focused to users' concern and desires in the most convenient way [27, 28].

2.3 UX Evaluation Measures

A system gains its acceptability and loyalty from the users when it attains the capability of providing high quality user experience [29, 30]. As such, evaluation of UX is a mainstream

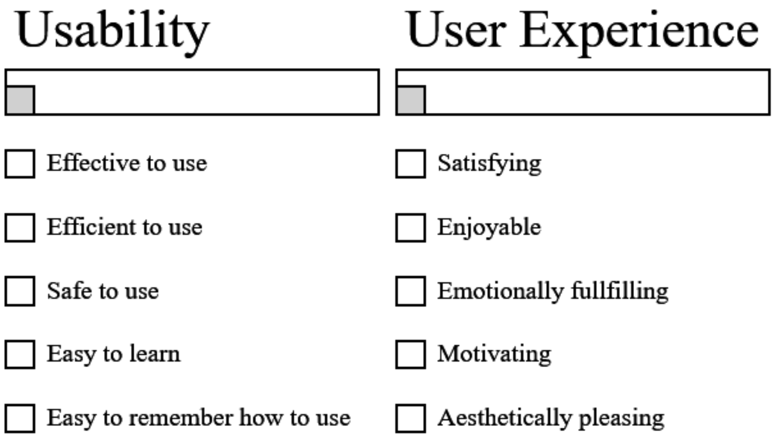


Fig. 2.4. Properties of UX and usability

activity in HCI [31]. Without evaluating UX, it becomes uncertain whether the system satisfies user requirements accurately. To ensure this, there are different types of methods for UX evaluation which are primarily classified as subjective methods and objective methods and are mainly concerned with user participation [32]. Subjective methods are usually defined as the methods which involves data from a user's personal judgements and beliefs like rating a system basing on how effectively it satisfies the user requirements whereas objective methods usually involve data from observations like facial expressions of a user basing on how stressed or calm he/she is feeling [33]. It is important to understand which method (subjective or objective) to use for UX evaluation that will bring out useful evaluation results for a particular context. For example, Biduski et al. [34] evaluated UX of a mobile health application using questionnaire (subjective method). Again, Staiano et al. [35] evaluated UX of some media player applications using both questionnaires (subjective method), facial expression and performance metrics (objective methods). The evaluation of UX plays a substantial role in determining successful utilization of a system.

2.3.1 Subjective and objective methods

Subjective methods are described as those methods which are basically assessment driven by user opinions towards any product or service. It depends on individual point of views and preferences of users and also, the evaluation results are based on facts which are not tactile and cannot be measured or quantified. On the contrary, objective methods of evaluation are measurable nor quantifiable, that is, it makes an effort to quantify values of evaluation data. For instance, how satisfied a user is with a system's features can be assessed by both subjective and objective evaluation methods. If users' satisfaction is measured by asking him/her questions, then the outcomes of this evaluation method would be obviously influenced by his/her personal preference which is subjective evaluation. Again, if the users' satisfaction is measured by using some performance metrics to see whether he/she can perform certain tasks, then the outcomes of this evaluation method would be measurable and is termed as objective evaluation. As a result, the basic difference between subjective and objective evaluation lies in the fact that the outcomes of objective evaluation is measurable by anyone and not only by the observer but the outcomes of subjective evaluation might be different because it depends on the observer [36]. Figure 2.5 shows some of the mostly used subjective and objective methods of evaluation. For example, subjective methods include interviews, surveys, rating scale and questionnaires and objective methods include performance metrics, behaviour observation and neurological and physiological response.

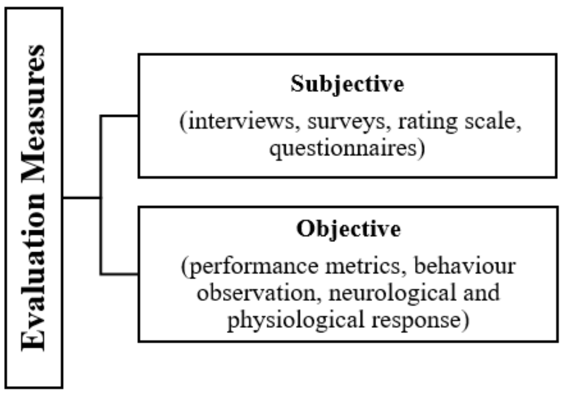


Fig. 2.5. Examples of subjective and objective methods

2.3.2 Neurological and physiological measures

One of the major portion of objective evaluation methods consists of user responses from various kinds of neurological and physiological attributes. Human emotion, behavior, perception and contentment towards using a system are often regulated by their neurological and physiological responses [37]. These variety of neurological and physiological responses are widely used for evaluating human-computing interacting systems because of its notable advantages such as reliability, implicitness, multidimensional, continuous responsiveness and such [38]. Electroencephalography (EEG), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI) are some of the activities that belongs to neurological measures whereas cardiovascular, electrodermal, facial expression, eye tracking, respiration etc. are some of the activities that belongs to physiological measures [39]. Figure 2.6 shows different types of neurological and physiological measures. Psychophysiological measures are usually preferred more for HCI research as it is considered to provide better insights than the measures like questionnaire and surveys [38]. Similarly, like evaluation methods, it is required to identify how neurological and physiological measures individually or jointly can bring about a wide range of valuable findings for different application context.

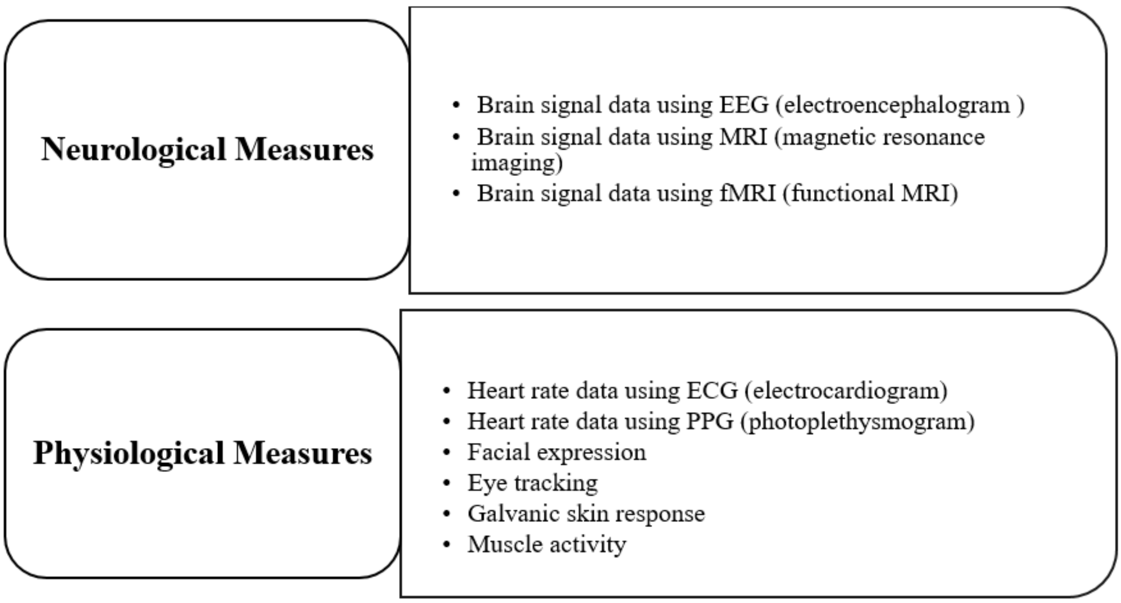


Fig. 2.6. Examples of neurological and physiological measures

2.4 UX and Emotion

UX involves both affective and practical experiences of users towards using a product or system where users' affective experience are usually defined as their emotional experience [40]. Thus, there persists a close relation between user-experience and emotion. These emotional experiences of users are highly influenced by their neurological and physiological process [41]. So, neurological and physiological measures play a very useful role while inferring emotions to evaluate user-experience of a system. Once a user interacts with a system, he/she experiences a number of emotions. Figure 2.7 indicates the relation between Maslow's hierarchy of needs and Aaron Walter's adapted hierarchy of user needs [42]. This relation shows that positive or negative whatever the emotion is, emotional experience helps to create a long-lasting impact on user experience. Thus, emotional experiences are important to evaluate user experience because emotions are inherently natural and they often reflect the true viewpoint of a user while they interacts with a system [43]. The viewpoints acts as a vital factor in the determining success of a system as positive experiences motivate the users to use the system more widely. On the other-hand, negative experiences help the designers to improve the design of a system in a more user intuitive way. Again, positive and negative emotional experiences might vary basing particular context. For example, 'fear' is generally termed as a negative emotional experience but for a context like horror games where various intensity of fear emotion is expected from users and this fear is termed as positive emotional experience in this regard [44]. Consequently, basing on different context, perceiving clear perspective of intended users is a very important matter of concern while evaluating UX of interactive computing systems through emotional experience.

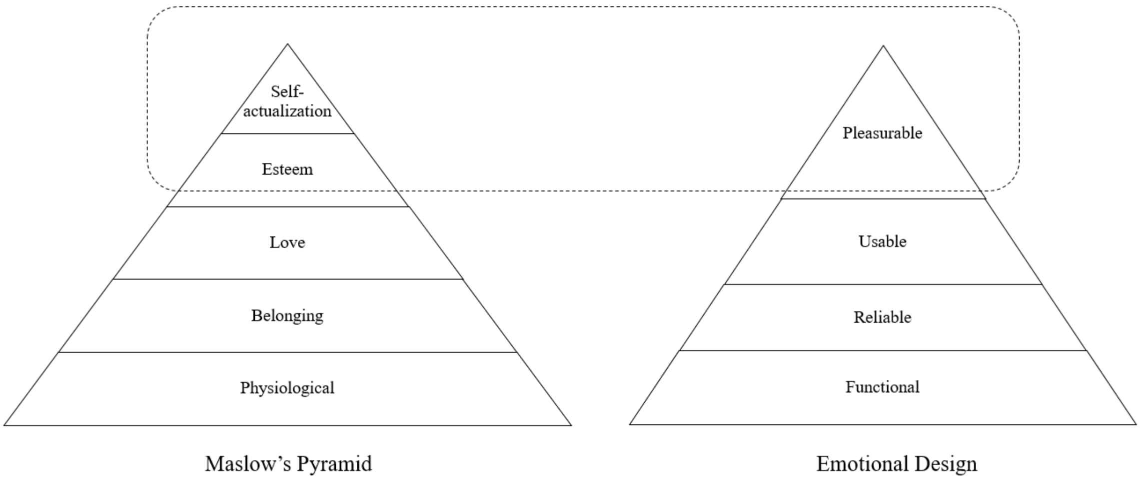


Fig. 2.7. Relation between Maslow's hierarchy of needs and emotional design

2.5 Gaming Context and UX

Interactive applications are being developed for a wide range of context these days where advertising, entertainment industry, healthcare, education and e-commerce are some of the widely prevailing ones [45]. Among these application areas, entertainment industry is a huge sector which largely involves gaming environment and social media. Gaming context is such an application area which is entertains all ages and types of users basing on its purpose. Moreover, the presence of user or user involvement is very intense in case of interactive gaming application areas because there is a one to one relationship between the user and the system and so there exists a constant thirst for seeking entertainment through the games [46]. Users' experience a variety of events while they play any type of game. Both psychological and neurological attributes are affected in a variety of ways during their engagement in a game playing session. Evaluating this intensity of engagement is basically user experience and can bring about a lot of useful information which can further aid to improve a game's design. Hence, in order to design interactive gaming applications in accordance with users' satisfaction and expectations, evaluating UX of gaming context is an essential matter of concern.

2.5.1 Classification of games

There are different types of games which can be classified in many ways. It depends on users' age, interest, user-friendly interface, comforting application platform and on a lot of other characteristics. Among all these different ways of classification, Vlachopoulos et al. has provided a very organized way to classify games basing on a number of characteristics in his study [47]. According to this study, games have been divided into five categories which are purpose, content, technical characteristics, platform and type of a game. A game might belong to a sub-group of a category or to a combination of multiple sub-group of a category. The purpose of the game could be knowledge acquisition, skill acquisition, content understanding and engagement. Next, the content of the game could be any subject social (social science, language, mathematics and such), social attributes (learning some social skill) and intuitive control (controlling car speed in racing games). Then, the technical characteristic of a game could be different strategies, techniques/approaches and modes. The game platform could be console/handheld, online (web-based or mobile-based) and networked. Lastly, the game type could be defined as arcade, adventure, action and serious games. Figure 2.8 shows the summary of the classification of games adapted from [47].

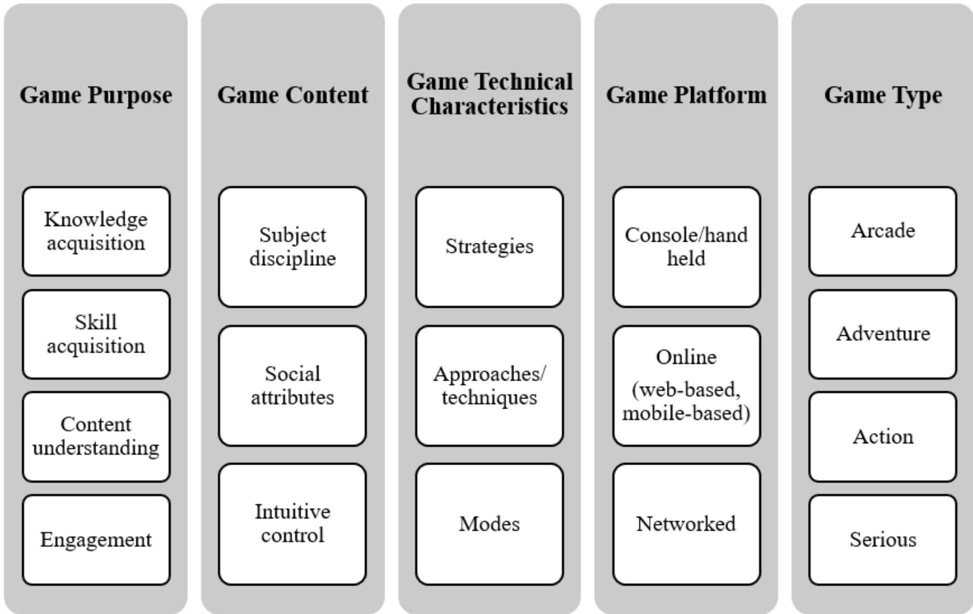


Fig. 2.8. Classification of games

2.5.2 Serious games

Serious games have a particular purpose which is rendered in an entertaining way through gaming experience [11]. The purpose of serious games can be of various types such as education, healthcare, military training programs and the likes where some specific users receive something useful that is more than mere entertainment [48]. A study by Laamarti et al. [49] shows a taxonomy for classifying different types of serious games as displayed in Figure 2.9. The taxonomy is divided into four groups - context (education, training, communication, healthcare, well-being, others), activity (physiological, neurological, physical effort), mode (visual, auditory, haptic, others), means of interaction (tactile interface, brain interface, eye gaze, others) and environment (virtual reality, social presence, mobility, online). Some examples of existing serious games are discussed as follows; a serious game Tactical Language and Culture Training System (TLCTS) is aimed towards teaching communication skills regarding non-native or foreign language [50]. Again, a serious game purposed for medical training program is (JDoc - junior doctor medical simulator) which helps to train junior medical doctors [51]. Another similar medical training based serious game is the Pulse [52]. Mathbreakers is a serious game which is intended to provide basic mathematics teaching facilities for students [53]. Another serious game uses role play strategy to provide educational facilities for learning language online [54]. Figure 4.1 presents snapshots of some serious game applications among the above-discussed ones. So, it is evident that serious games are intended for learning experience and specific goals are defined that allow a player to get engaged or involved in a game and gather learning experience. The expected learning experience is highly dependent on how much users are involved into the game and measuring this learning experience helps to assure the effectiveness of serious games. Therefore, evaluating UX of serious games is one of the best ways to assure how much the intended learning experience of users are achieved.

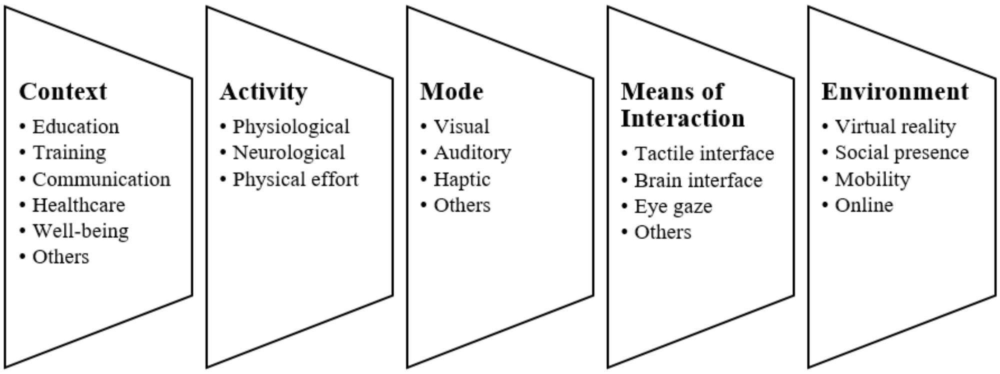


Fig. 2.9. Taxonomy for classification of serious games

2.6 Machine Learning

The study of machine learning (ML) involves those type of computer algorithms which has the ability to improve consistently through a process of experience from a sample dataset [55]. The advantage of using machine learning algorithms is that very minimum or no intervention of human is required and certain patterns and trends could be predicted conveniently from a huge amount of data. Machine learning algorithms can be primarily classified into four types - supervised, unsupervised, semi-supervised and reinforcement learning where supervised learning works on dataset that is labelled, unsupervised learning works on dataset

(a) Medical training based serious game

(b) Mathematics teaching based serious game

(c) Language learning based serious game

Fig. 2.10. Snapshots of serious games (a) medical training program (b) language learning program (c) mathematics teaching program

that is not labelled, semi-supervised learning considers dataset that are labelled and unlabelled, and reinforcement works in a environment that is dynamic by adopting appropriate action for maximizing reward in a definite situation [56]. Figure 2.11 shows the types of machine learning as discussed. The basic procedure of machine learning algorithms is to predict a feature using classification process, known as target feature or dependent feature basing on a number of independent features. The general process of machine learning technique is shown in Figure 2.11 adapted from [57]. It demonstrates how some data of past known as training data is used for building the machine learning model and basing on this trained model later on new data known as testing data is used for predicting the target feature.

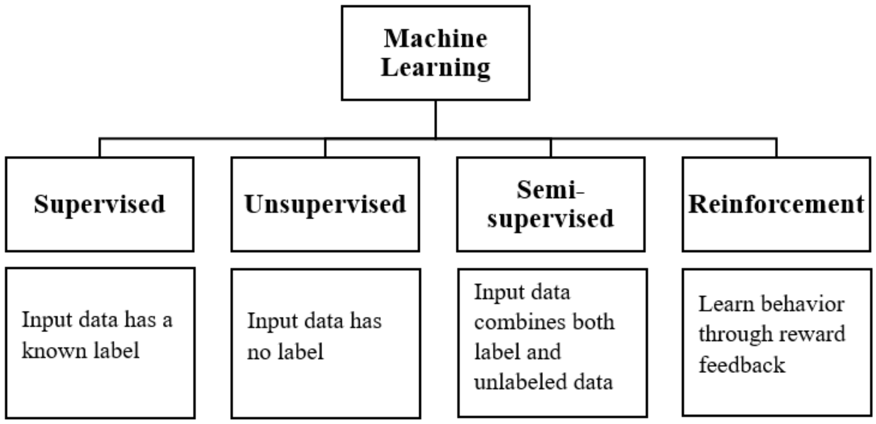


Fig. 2.11. Types of machine learning

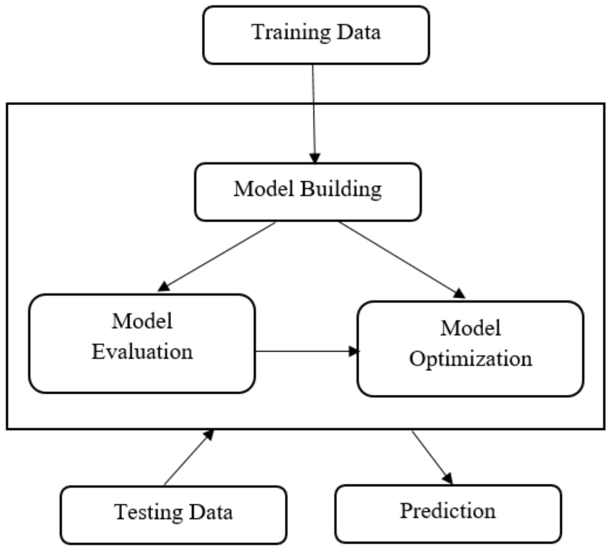


Fig. 2.12. Machine learning process

2.6.1 Application of machine learning for emotion classification

There exists a variety of algorithms among each of the type of machine learning techniques. But in the area of emotional experience prediction using machine learning (ML) techniques, supervised algorithms are mostly used because it is comparatively easier to predict emotions for a labelled dataset [58]. Another category of supervised learning algorithms for more efficient prediction is using multiple learning algorithms for achieving better performance of prediction than could be achieved from any of the algorithms alone is called ensemble learning [59]. The prediction accuracy of these algorithms vary depending on the type of dataset as well as the features used for predicting. It also depends on the size of dataset and the number of features. For example, some algorithm can give better prediction for a particular context but lower prediction accuracy for another context [60]. Moreover, there does not exist any particular criteria for choosing which algorithm is best for which type of application. As a result, the applicability of ML algorithms is highly dependent on a number of variables such as type of context, size and type of dataset and the likes. Anyhow, machine learning is a modern and efficient technique to observe different trends and obtain useful information for a research area.

2.6.2 Machine learning algorithms

Among different types of supervised algorithms some of the traditional and ensemble learning algorithms used for this thesis purpose are Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Xgboost, Catboost and Adaboost. These algorithms have been chosen because of their relevancy to emotional prediction [61, 62]. A brief description regarding these algorithms are prescribed below [63, 64]:

Random Forest (RF): Random forest constructs a number of decision trees and consolidates them to get a more exact and stable expectation. The aim of utilizing a decision tree (DT) is to develop a model as training model which can be used to estimate the target variable's class through learning basic rules of decision construed from earlier data, that is, training data. Rather than looking for the main feature during the splitting process of a node, it looks for the finest feature among different randomly distributed features.

Support Vector Machine (SVM): Every data is plotted as a coordinate in a space which has n dimensions where n equals to the total number of features. Each coordinate represents the value of each feature. The classification is done by searching the hyperplane which separates the two classes quite well. There can be numerous hyperplanes that can perform this classification; however the goal is to find that hyperplane that has a margin which is the highest, that is, which implies greatest distances between two of the classes, so that when a new data is required to be classified, the classification between two classes can be done easily.

K-Nearest Neighbour (KNN): KNN follows a principle where an assumption is made that each data that falls close to one another falls in a similar class. This implies that similar data are close to one another.This algorithm chooses a number ‘k’ which is the closest neighbor to the data that would be classified. The principle is searching the distance between all the points in the dataset and a specified query, choosing a specific number ‘k’ nearest to the query and finally voting for the label which is most frequent.

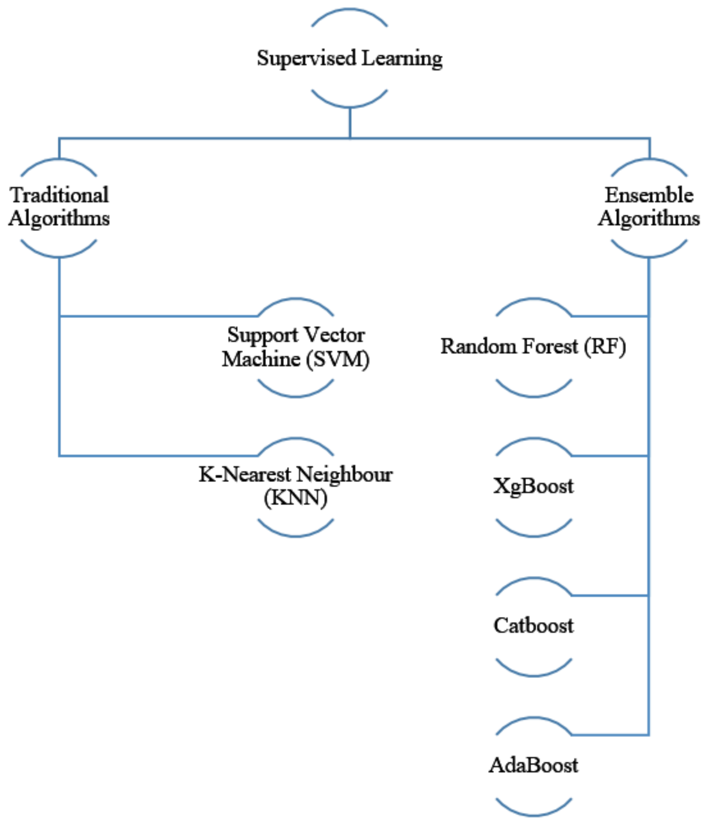


Fig. 2.13. Supervised learning algorithms for emotion prediction purpose

Xgboost: The definition of XGboost is ‘extreme gradient boosting’. In gradient boosting,the learners which are weak are basically regression trees. One of the leaves of each regression tree possess a score that is continuous and a data (input) is mapped to it. XGBoost uses a function termed as loss function and tries to minimize it. The objective of loss function is based on differences between the target and estimated outputs. A continuous interation process is followed by the training where new trees are added that estimate the errors of earlier trees. Then, these residuals gets combined with earlier tress in in order to perform the intended prediction finally.

Catboost: CatBoost follows a principle which follows a greedy manner. In particular, Cat-

Boost consolidates every features which are categorical and all possible combinations effectively utilized for earlier splits in the present tree with all mentioned type of features of the dataset for each split of a tree. In sum, CatBoost depends on ‘gradient boosted decision trees’. While training, a bunch of decision trees is constructed continuously. Each progressive tree is created with diminished loss in comparison with the earlier trees.

Adaboost: Adaboost is characterized as ‘adaptive boosting’. The algorithm establishes an arrangement of weak learners on training data which are of various weights. It begins by anticipating dataset which is original and provides equivalent load to every observation. If using the first learner provides incorrect prediction, then a greater load is given to the observation which has provided the incorrect prediction. During this iterative cycle, it keeps on adding learner(s) until a breaking point is reached in the quantity of models or precision. The discussed algorithms are summarized in a hierarchical chart shown in Figure 2.13.

CHAPTER THREE

RELATED WORK

This chapter discusses the methodology of the systematic literature review following four steps which are - search strategy, inclusion and exclusion criteria, finalizing the review materials, and data extraction. Then, the outcomes of the review study are discussed in an organized manner. Basing on the review outcomes, a number of research gaps and future research opportunities are prescribed afterwards. Then, the focused research opportunity for this thesis and an overview of some of the existing studies that are related to the concerned experimental study are discussed. Lastly, a critical summary is presented to emphasize the issues that motivated to conduct further study regarding the focused research opportunity:

3.1 Methodology for Literature Review

The systematic literature review (SLR) approach suggested by Kitchenham [65, 66] is followed to review the research articles. The aims of this literature review are to explore the significance and effectiveness of neurological and physiological measures in evaluating usability and UX of information systems; existing neurological and physiological measures used in different context and software platforms; comparative views among the neurological and physiological measures for evaluating UX and usability as well as investigating the possible future research opportunities in such aspects. The steps for this literature review are discussed in a chronological manner as follows:

3.1.1 Search strategy

Apart from Google search engine, a number of scholarly digital databases were used to find out the related literature for this review study that includes Google Scholar, SpringerLink, ScienceDirect, ACM Digital Library, Scopus, IEEE Xplore, PubMed and Wiley Online Library. A number of search strings were used for searching research articles from different digital databases. Boolean operators have been used for search syntax which are AND and OR. Also, a wild card character * is used in some search queries which indicates matching results with one or more characters. The keywords were searched in different set of combinations in title, abstract and keywords of a certain article as can be seen in the search syntax format of the different sources. A summary regarding the finally selected total number of articles with respect to source (digital database/library) is portrayed in Table 3.2 which shows most articles have been gathered from Google Scholar and SpringerLink Digital Library.

3.1.2 Inclusion and exclusion criteria

A number of inclusion and exclusion criteria were defined to select the relevant research articles. These inclusion and exclusion criteria to finalize the related review materials are presented in Table 3.1.

Table 3.1: Inclusion and exclusion criteria to select review articles

Inclusion/Exclusion	Criteria
Inclusion Criteria	**IC1:** Research articles that are written in English language. **IC2:** Research works that have been published since 2003 to 2019. **IC3:** Articles published in conference/workshop proceedings, in academic journals and as thesis dissertations. **IC4:** Complete (full-text) research articles.
Exclusion Criteria	**EC1:** Research articles having same title but different source (i.e. duplicate articles). **EC2:** Articles conflicting with the theme of the systematic literature review. **EC3:** Full-text articles lacking sufficient evidence to justify UX and usability assessment. **EC4:** Articles containing physiological and neurological measures that are not being used for the purpose of UX and usability assessment. **EC5:** Articles concerning physiological and neurological measures purposed for UX and usability evaluation other than the context of information systems (e.g. physical products in supermarket).

Table 3.2: Search syntax/strings and the number of selected articles along with the respective scholarly database

Source	Search Syntax/String	No. of Articles
Google Scholar	"UX evaluation AND physiological measure"; "UX evaluation and neurological measure"; "usability evaluation AND physiological measure"; "usability evaluation and neurological measure"; "UX evaluation OR usability evaluation AND neuromarketing"; "HCI and information systems"; "UX evaluation and information systems"; "usability evaluation and information systems".	5
SpringerLink Digital Library	(UX evaluation OR usability evaluation) AND (information systems*), (UX evaluation OR usability evaluation) AND (neurological measure*), (UX evaluation OR usability evaluation) AND (physiological measure*), (UX evaluation OR usability evaluation) AND (neuromarketing*), (HCI AND information systems).	6
ScienceDirect Digital Library	"UX evaluation" OR "usability evaluation") AND "information systems"*), "UX evaluation" OR "usability evaluation") AND "neurological measure"*), "UX evaluation" OR "usability evaluation") AND ("physiological measure"*), ("UX evaluation" OR "usability evaluation") AND ("neuromarketing"*).	4
ACM Digital Library	"UX evaluation and physiological measure", "UX evaluation and neurological measure", "usability evaluation and physiological measure", "usability evaluation and neurological measure", (UX evaluation OR usability evaluation) AND neuromarketing, "HCI and information systems".	3
Scopus	(TITLE-ABS-KEY("UX evaluation") OR TITLE-ABS-KEY ("usability evaluation") AND (TITLE-ABS-KEY ("physiological measure") OR(TITLE-ABS-KEY("UX evaluation") OR TITLE-ABS-KEY ("usability evaluation") AND (TITLE-ABS-KEY ("neurological measure") OR(TITLE-ABS-KEY("UX evaluation") OR TITLE-ABS-KEY ("usability evaluation") AND (TITLE-ABS-KEY ("information systems") OR(TITLE-ABS-KEY("UX evaluation") OR TITLE-ABS-KEY ("usability evaluation") AND (TITLE-ABS-KEY ("neuromarketing") OR TITLE-ABS-KEY ("HCI") AND (TITLE-ABS-KEY ("neuromarketing") AND (LIMIT-TO(LANGUAGE: "English").	3
IEEE Xplore Digital Library	("UX evaluation" OR "usability evaluation") AND ("neurological measure"*), ("UX evaluation" OR "usability evaluation") AND ("physiological measure"*), ("UX evaluation" OR "usability evaluation") AND ("neuromarketing"*), ("HCI" AND "information systems").	2
PubMed Central Online Library	(UX evaluation [Title/Abstract] OR usability evaluation [Title/Abstract]) AND (information systems* [Title/Abstract]), (UX evaluation [Title/Abstract] OR usability evaluation [Title/Abstract]) AND (neurological measure* [Title/Abstract]), (UX evaluation[Title/Abstract] OR usability evaluation [Title/Abstract]) AND (physiological measure* [Title/Abstract]), (UX evaluation [Title/Abstract] OR usability evaluation [Title/Abstract]) AND (neuromarketing* [Title/Abstract]), (HCI [Title/Abstract] AND information systems [Title/Abstract]).	2
Wiley Online Library	("UX evaluation" OR "usability evaluation") AND ("information systems"*), ("UX evaluation" OR "usability evaluation") AND ("neurological measure"*), ("UX evaluation" OR "usability evaluation") AND ("physiological measure"*), ("HCI" AND "information systems").	1
Google Search	"UX evaluation and physiological measure", "UX evaluation and neurological measure", "usability evaluation and physiological measure", "usability evaluation and neurological measure", "UX evaluation or usability evaluation" and "neuromarketing", "HCI and information systems", "UX evaluation and information systems".	1

3.1.3 Finalizing the review materials

Initially, an in-depth and exhaustive search was performed through the stated scholarly databases and Google search engine. From a total of 957 articles, after removing duplicates, 432 remained and after that 180 articles were screened. Next, after applying the inclusion and exclusion criteria, 52 articles were primarily selected, which were further assessed based on the focuses of these articles and finally 27 articles were selected for the systematic literature review. A PRISMA flow diagram is given in Fig. 3.1 that shows the detailed process of searching and selecting the final set of review materials in a number of subsequent phases.

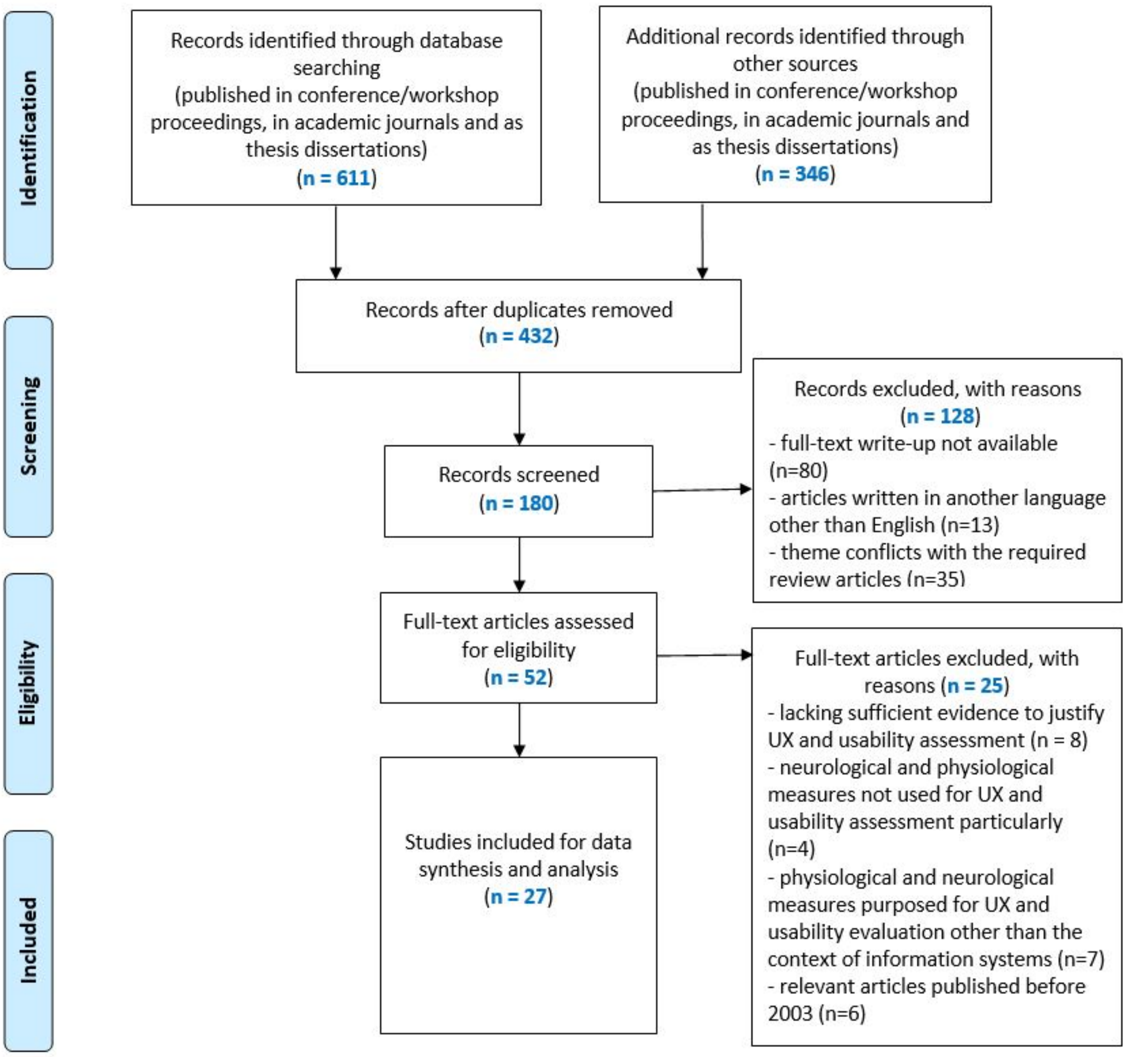


Fig. 3.1. PRISMA flow diagram

3.1.4 Data extraction strategy

In order to extract necessary data and perform the literature review in a structured way, six themes have been taken into consideration as shown in Fig. 3.2. Some of the themes have been further organized into a set of attributes that would help to extract data in a well-structured way to attain the review objectives. The selected themes are discussed briefly below to state what specific type of data was extracted from the reviewed research articles and Table B1 shows a sample of the data extraction process:

Topical Relationship: This theme depicts how closely associated the articles are that have been selected for the review basing on keywords and titles.

Aims and Outcomes: This theme states the research objectives, research outcomes, implications and outcome evaluation of an article.

UX and Usability Evaluation Measures: This theme points out which type of measure (neurological, physiological or other) is being used for evaluating UX and usability.

Research Context: This theme classifies the application type and platform the article is focused to.

Research Type: This theme identifies whether type of the article is an experimental study, literature review, conceptual study, case study or anything else.

Publication Year and Publication Type: This theme specifies the year in which an article was published and the type of publication (conference/workshop proceedings, academic journal or thesis dissertation).

3.2 Outcomes of the Review Study

A detailed discussion on the review findings inferred from the data synthesis and analysis is provided in this section. The findings are described as follows basing on the six themes and the overview of the review findings is presented in Fig. 3.3:

3.2.1 Topical relationship

The world cloud presented for the keywords and title reveals the intensity of the relationship among the topics of the reviewed articles (see Figure A1 and Figure A2). A wide range

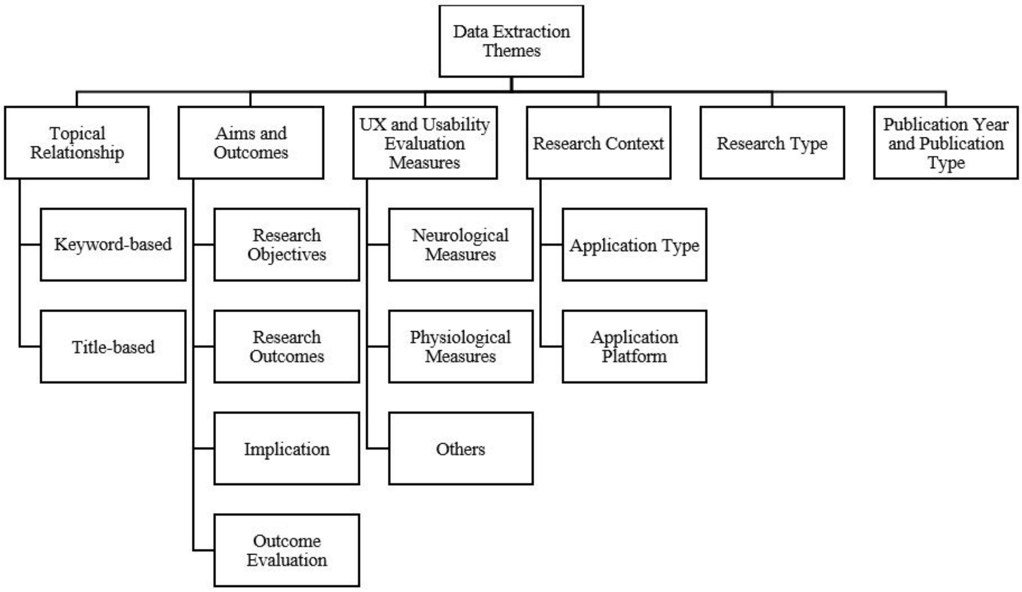


Fig. 3.2. Data extraction themes

of words relevant to the gist of this systematic literature review such as user, experience, usability, evaluation, emotion, research, human-computer interaction, information, systems, psychophysiological measures, neuromarketing have been highlighted broadly in both the keywords and titles of the articles. But the most significantly featured words include user, experience, usability, emotion and neuromarketing. These indicate that the chosen articles for this systematic literature review are closely associated.

3.2.2 Aims and outcomes

The studied research articles are targeted towards assessing different types of UX issues more than usability issues whereas a very few are concerned with both UX and usability issues. Among the four categories of research outcomes, evaluation-oriented recommendations/guidelines is more weighted than evaluation method/approach followed by equally weighted theoretical model/framework and other discussions (see Figure A3). The outcomes of the articles imply almost double for research than practice. Also, the validity of the final outcomes of majority of the articles is evaluated by the researchers (see Figure A3).

3.2.3 UX and usability evaluation measures

As the theme of this article focuses on neurological and physiological measures, so a wide range of such measures have been found in this regard among which EEG/brain wave analysis as neurological measure plays a predominant role for particularly UX evaluation in the area of e-commerce applications. In case of physiological measures, ECG/heart rate variability bears a huge contribution followed by eye tracking, facial expression and galvanic skin response whereas electrodermal activity imparts the least in this respect. On a whole, physiological measures are more likely in contributing to gaming experience, embedded systems and advertising campaigns for usability evaluation. But when combined, both types of measure hold the capability to evaluate UX and usability issues mainly for gaming experience, advertising campaigns and e-commerce applications. On the contrary, others criteria mentioned earlier consist of a wide variety of qualitative and quantitative measures but they are used in combination with neurological and/or physiological measures for comparison purpose in the reviewed articles. Figure A4 shows the summary of the data analysis concerning the UX and usability evaluation measures. For both neurological and physiological measures, some devices have been also identified as means of data collection which are commonly used (see Table B2).

3.2.4 Research context

The findings of this theme are classified into two parts. Firstly, for application type (see Figure A5), advertisement campaigns are the most assessed area followed by gaming experience and e-commerce applications with social media being the least assessed one in the studied research articles. Besides, others criteria mentioned previously also bear some notable contribution in this respect. Secondly, for application platform (see Figure A6), web-based application platforms have been evaluated more than mobile app and desktop-based platforms whereas other platforms like embedded system, client-server system and workstation being the least evaluated one. Besides, Table B3 provides a mapping between evaluation measures with type and platform of applications.

3.2.5 Research type

The data analysis reveals that the gathered research articles are largely experimental study-based that comprises of application of different neurological and physiological measures for UX and usability evaluation. Additionally, an impressive number of literature review and conceptual paper also goes with the theme of this literature review. A case study also contributes to this review. A summary of the data analysis regarding this theme is shown in (see Table B4)

3.2.6 Publication year and publication type

The articles have been selected for the systematic literature review within the time range of 2003 to 2019. It has been observed from the data analysis that (see Figure A7), from 2003 to 2008 and 2013 to 2015 the number of articles aligned with the theme of this literature review was quite similar. During 2009 and 2017, the count of such articles was at its peak and aimed at UX evaluation dominantly than usability (see Figure A8). It is also worthy to mention that throughout this peak period, the research outcomes were also on the rise with recommendation/guidelines for evaluation being the highest followed by theoretical model/framework and evaluation method /approach respectively (see Figure A9). Furthermore, as can be seen from Figure A10, majority of the articles that has been considered for study within this timeframe covers conference proceedings/workshop and academic journal publications mostly where the subject area of journals are principally founded upon human-computer interaction.

3.3 Research Gaps and Future Research Opportunities

The findings of this systematic literature review reveal some constructive research opportunities in the broad area of Human-Computer Interaction (HCI). Some of the potential future research opportunities are recommended below which can be considered in order to meet up the research gaps related to UX and usability issues of information systems to some extent. Fig. 3.4 shows a diagram that briefly outlines the recommendations.

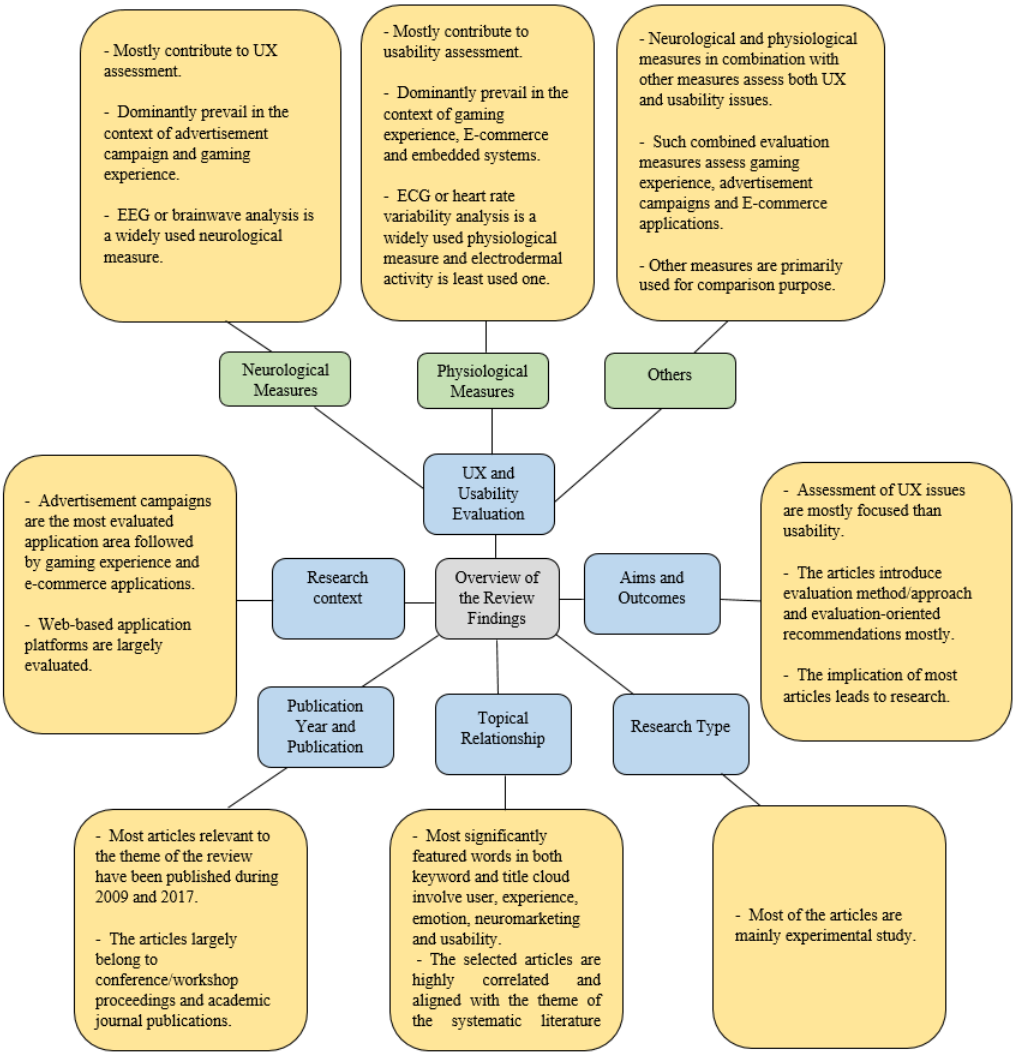


Fig. 3.3. Overview of the review findings

3.3.1 Exploring the significance of using a single or both evaluation measures

It can be observed from the review that EEG/brainwave analysis alone has been used as neurological measure for UX evaluation in [67]. Besides, different kinds of physiological measures have been used individually for both UX and usability evaluation, such as heart rate variability/ECG in [68], eye tracking in [6] and so on. On the contrary, different combination of physiological and/or neurological measures have been also used for UX and/or usability evaluation. For instance, heart rate variability/ ECG, galvanic skin response/electrodermal activity, and electromyography/muscle activity have been used unitedly in [69]. Again, [70] shows the combined use of brainwave analysis/ EEG, heart rate variability/ ECG, galvanic skin response/electrodermal activity, and emotion. Existing researches focus on evaluating an application using neurological and/or physiological measures for UX and/or usability evaluation. However, none of the studies have explored the significance of using a single measure (physiological or neurological) or both measures in this aspect. In other words, the potential future research may explore the significance or impact of using physiological and/or neurological measures for evaluating UX and/or usability of a particular application. Therefore, there arises an opportunity regarding whether a single evaluation measure (neurological/physiological) or both can be reliable enough to validate UX and/or usability of an application.

3.3.2 Investigating priority consideration in multimodal approach for evaluation

This review study has particularly focused on evaluating such interactive information systems that have considered mainly neurological and physiological measures for UX and/or usability evaluation. Many of the existing researches have suggested on adopting multimodal approach in this regard. A triangulation multimodal approach involving neurological, physiological and traditional measures has been introduced in [12] for UX evaluation. Another experimental study in [71] recommends a multimodal approach involving a variety of physiological and subjective measures provide useful information that can improve usability of an application. Now, multimodal approach indicates that a variety of measures (neurological/physiological/traditional subjective and objective) are used in a variety of combination for evaluating an application. Regardless, it is still not revealed how applicable this multimodal approach is for evaluating UX and/or usability. As a result, it is unknown whether any type of priority within such combinations can result in an effective outcome. Thereby, a research concern exists to investigate any type of priority consideration in multimodal approach for UX and/or usability evaluation.

3.3.3 A paradigm shift in technology for selecting evaluation measures

In case of neurological measures, EEG for brain signal analysis has been found as a widely used measure for evaluation in the studied articles. But some more advanced optical brain imaging technology (e.g. functional near-infrared spectroscopy (fNIRS) [72]) can be introduced for such purpose and a comparative analysis can be done between these two similar technologies for a particular application. Similarly, physiological measures that provide similar purpose using different technology can be compared to find out the most convenient technology for certain areas of application. For example, [68] introduces a novel technology called transdermal optical imaging to measure heart rate variability which is much more efficient than traditional technology that involves measuring heart rate variability using electrodes. So, a shift in technology can bring about a revolutionary change in evaluating UX and/or usability of an application and reveal valuable insights regarding timing constraints, cost-effectiveness, performance efficiency, user-comfort and such parameters.

3.3.4 Change in conventional comparative approach among evaluation measures

The existing researches show that most of the common comparative approaches are to compare the results obtained from neurological and/or physiological measures with traditional subjective measures (e.g. interviews, questionnaires, surveys etc.) [12], [6] and with typical objective measures like various performance metrics (e.g. number of clicks, number of errors etc.) [73]. However, there exist other approaches such as the use of heuristic methods [74] which plays a very important role in evaluating UX and usability and is also a most commonly used cost-effective approach in the field of HCI. Again, only a few studies have been conducted focused on comparing neurological and/or physiological measures with heuristic method such as [73] discusses usability evaluation of an embedded system using subjective, objective and heuristic methods. Therefore, further research can be conducted to explore the differences and gain more insights regarding UX and/or usability while comparing neurological and/or physiological evaluation measures with heuristic evaluation methods.

3.3.5 Expanding research for less evaluated application context

The review study revealed that among various application areas, advertising campaign has been found to be the leading one for evaluation and then, gaming experience and e-commerce applications take the place subsequently. A few studies have been found where usability issues are evaluated in embedded systems [73] and also in distributed systems [75]. While now-a-days, due to the technological advancement, the application of distributed and embedded systems are becoming more important in every aspect of our life. Again, such systems are being developed for different areas and for different kinds of hardware integrated platforms which should be taken into consideration for UX and usability evaluation. The review shows that the few articles which exist regarding these systems [75], [73] mainly discusses traditional measures (subjective/objective) for evaluating their usability. But these type of systems are more prone to affect users' neurological and physiological attributes where only traditional measures for evaluation may not be enough to establish a conclusive result, such as automated diagnosis and health assessment system in the field of smart healthcare technology [76]. As a result, some promising opportunities lie in evaluating and comparing the evaluation results of UX and/or usability of these systems as well as proposing some model/framework for evaluating such type of systems using neurological and physiological measures.

3.3.6 Enhancing evaluation for different variety of a particular application context

It is evident that the most prevailing application context are a matter of interest while evaluating UX and usability issues. Existing researches show that advertising campaigns, gaming experience and e-commerce applications are mostly evaluated. But these are some broad context of application and there exists a huge variety for such areas. For example, in the field of advertising context, some review articles concerned tourism [77] and most of them concerned branding campaigns [78] and commercial advertisement [70]. But advertising campaign can be for different types of affairs such as tourism, different kinds of business, healthcare etc. Similarly, gaming experience can aim at purely entertainment or education or special purpose that is serious games. Again, e-commerce applications can focus on a variety of products. Now, among these broad context of application, there prevails a wide range of area where users' neurological and physiological traits have a great contribution. As an example, in the context of advertising campaign, higher education branding [79] is a sector where students' neurological qualities are heavily affected that influences them to pursue the education. Again, for gaming experience, users' neurological and physiological traits play a very important role to understand their behaviour while playing different types of game [80]. And, for e-commerce applications, consumers' neurological and physiological features have substantial influence on their purchase behaviour [81]. Hence, there definitely lies an important research concern to achieve a deeper understanding of UX and usability issues from the diversity of a particular application context using neurological and physiological measures.

3.3.7 Incorporating emotional experience as evaluation measure for serious games

One of the widely evaluated application areas among the review articles is gaming experience. Again, emotion is one of the attributes that is largely affected in an entertainment-based computing environment like virtual reality, companion robots and gaming experience [82]. Now, some studies have considered emotion as UX and usability evaluation measure. For instance, emotion is used to evaluate UX of a gaming context (an action-puzzle game) [71]. Again, emotion is also used to evaluate usability of another gaming environment (a digital car game) [7]. So, it is clear that there exists a close relation between users' emotional experience and gaming applications. But the existing studies have focused on particularly entertainment purposed gaming application and not on serious games while serious games are played for a particular purpose rather than pure entertainment such as a simulation game for healthcare training programs which has effect on users' emotion [83]. Thus, measuring emotion for UX and usability evaluation for serious games is vital and one of the key concerns in the context of gaming experience as it offers a gaming environment that incorporates pedagogical aspects which is highly regulated by user emotion where traditional measures maybe questionable [84]. Besides, how the pedagogical elements of different types of serious games are affected by emotion, how the evaluation results of emotion for general gaming experience and serious games differ, how effective the evaluation results are for a particular type of serious game are some of the open issues which require further investigation for evaluating users' emotional experience for serious games.

3.3.8 Context-based mapping with respect to evaluation measures

A number of neurological and physiological measures have been used either individually or combined for evaluating UX and/or usability of similar context of applications in the reviewed articles. For example, brainwave signal is used as neurological measure [12] whereas, heart rate is used as physiological measure [69] for evaluating UX of gaming applications. Again, similar neurological and/or physiological measures have been user for different kinds of context. For example, EEG/brainwave analysis is used as neurological measure for evaluating UX of e-commerce applications [67], advertising campaigns [70] and multimedia applications [85]. But no studied research shows a proper mapping regarding which particular measure (neurological/physiological) can be termed as an appropriate evaluation measure for a particular context. Moreover, exploring how the appropriateness and applicability of a context-based mapping is ensured, how the context-based mapping can be done (e.g. methodical approach) are some important research issues regarding the context-based mapping with respect to evaluation measures.

3.3.9 Defining a complete set of metrics for neurological and physiological evaluation measures

A number of metrics exist for evaluating usability and UX of an information system. For example, for gaming experience, usability metrics like learnability, memorability [86] while UX metrics like pleasure, boring, exciting [86], [87] matters for gaming context. On the other hand, for business context, usability metrics like flexibility, safety [86] and UX metrics like trust, comfort [86] matters. So, it is quite obvious that, the UX and usability metrics bear great importance for evaluating an application. The review study found that different articles used different metrics for evaluating UX and/or usability for similar purpose like for evaluating UX of an advertising context, user perception has been used in [77]; on the contrary, sad, angry, joy etc. has been used in [8]. Thereby, none of the studies have provided a definite set of metrics for neurological and physiological measures that can be used for UX and usability evaluation for a particular application. So, possible set of metrics for neurological measures, possible set of metrics for physiological measures and also possible set of metrics when both are combined should be investigated for UX and/or usability evaluation. Along with that, the viability of such metrics for different type of context requires to be investigated in order to outline a generalized set of metrics for a certain application context. These are some of the important concerns that paves the way for possible future research scope in this regard.

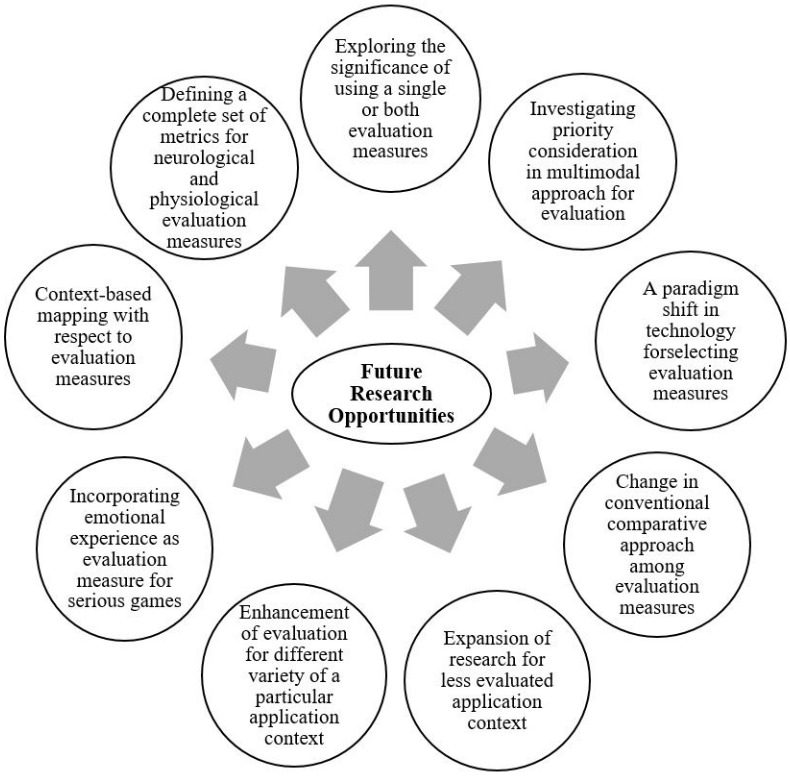


Fig. 3.4. Overview of the recommendations

3.4 Key Focus of this Research

Among these wide variety of research opportunities, *'Incorporating emotional experience as evaluation measure for serious games'* has been focused for further investigation in this thesis. This research is primarily concerned about gaming context, particularly serious game and the application of emotional experience as a means of assessing UX of such gaming environment. Some existing studies are discussed below in this regard.

As of typical games which is purposed for only entertainment, Yu et al. [88] has performed an experimental study to assess positive and negative UX from subjective measure (questionnaire) and objective measure (brain signal and heart rate data) using an improved RIPPER algorithm of a mobile based adventure game 'Temple Run'. Again, Diya et al. [89] has conducted an experimental study to recognize a number of emotions for UX evaluation using frequency bands of brain signals for an arcade game 'DX-Ball'. A different type of UX evaluation study has been discussed by Nacke et al. [90] where impact of sonic stimuli of an action game 'Half Life 2' was observed using electrodermal activity (EDA) and facial muscle activity (EMG) as objective measure and Game Experience Questionnaire (GEQ) as subjective measure. Another experimental study by Nacke et al. [90] has showed the use of electroencephalography (EEG) as objective measure and GEQ as subjective measure to infer emotions while playing an action game 'Resident Evil 4' using two types of game controllers.

The effect of emotions for UX evaluation in the field of HCI bears great importance. Dirican et al. [91] has provided a survey which shows that different psychophysiological measures can help to obtain multi-dimensional information regarding users' cognitive states to evaluate UX of HCI-based applications. Nagalingam et al. [92] has also discussed that emotional and cognitive elements are considered for evaluating UX of particularly educational games in this review study. For serious games context, Nacke et al. [93] discussed UX evaluation methods for such context where it was emphasized that evaluating individual player experience using objective or physiological methods is important for serious game context as it impacts a player's emotional or cognitive state greatly. Again, Anolli et al. [94] has explored how emotions can be introduced for evaluating UX of affective serious game applications and further suggested multimodal measures to assess emotions might bring out useful outcomes regarding UX evaluation.

Regarding serious games categories, a systematic literature review by Calderón et al. [13] reveals that among different criteria of serious games, around 53% educational games are evaluated for UX where 60% of these games fall under the category of university education and majority of them involves computer programming topics. It further reveals that questionnaires are mostly used to evaluate UX of such games. Moreover, other educational games include a computer game providing social education which has been evaluated using questionnaires and interviews [95], a multichannel game providing undergraduate healthcare education which has been evaluated using survey and questionnaires [96], a computer game providing middle school microbiology education which has been evaluated using dynamic knowledge tracing model [97] and a computer game providing soft skill education which has been evaluated using quantitative survey and qualitative observation [98].

As a huge percentage covers computer programming topics, a number of experimental study has been conducted on the serious games that concerns undergraduate computer programming mainly. Donald et al. [99] has evaluated a computer game 'Glitchspace' focusing on basic coding principles using cognitive walkthrough, quantitative questionnaires and qualitative observation. Besides, other two serious games focusing on basic programming skills have been evaluated using questionnaires which are 'Macro Run' – a mobile game for learning basics of programming) [100] and crosswords and shooting-based game – an online mini game for introductory programming) [101] respectively. Another computer game 'Prog&Play' dedicated to programming practice has been evaluated by Muratet et al. [102] using questionnaire and behavior observation.

3.5 Chapter Summary

The open issues that has emerged from the systematic literature review and some related works afterwards are manifold. Firstly, the systematic literature review paved the way towards this research opportunity basing on widely used context of UX evaluation among which gaming experience is a leading one but lacks its extent when it comes to serious games. Secondly, it is clear from the review of related works regarding UX evaluation of gaming context that emotions that are inferred from objective measures are widely used for UX evaluation. Thirdly, UX evaluation basing on users' emotional experience using physiological and/or neurological measures has been performed for different types of games mostly which provides purely entertainment such as different action, adventure and arcade games. Fourthly, for serious games category particularly educational serious games, UX is evaluated predominantly using subjective measures; questionnaires to be specific. Fifthly, this is equally true for computer programming-based serious games which covers a significant portion in the context of educational serious games.

Basing on these number of open issues, the research is focused towards UX evaluation of serious games. The aim is to investigate how emotional experience from psychophysiological measures can be used to evaluate UX of a computer programming-based serious game.

CHAPTER FOUR

EXPERIMENTAL STUDY

This chapter discusses the experimental study which is divided into three broad parts - experiment design, conducting the experiment and, data analysis and proposing the framework. The experiment design involves three primary activities of this experiment which are apparatus selection, participant profile and study procedure. After that, conducting the experiment is related to collecting and processing the data. The last part involves analysing the data and proposing the framework for emotional state classification.

4.1 Experiment Design

This step is concerned with the setup of the experiment. That is, it consists of the apparatus selection, participant profile and the study procedure. Apparatus selection gathering the required resources to conduct the experiment. The resources comprises of firstly, selecting the context of application which will be evaluated for identifying the emotional experience of individuals. Next, a particular set of emotional states is to be selected that would be predicted as there exists a variety of emotions. Then, certain evaluation measures related to UX are to be selected and which will be used to perceive the emotional experience of users while they are exploring the selected application. Participant profile is related to the total number of participants, their demographic information who will be taking part in the experimental procedure, their consent of voluntary participation, their basic knowledge regarding the type of application and software platform they are going to deal with and such. On the other hand, the study procedure depicts a well-defined structure of how the overall experiment is going to be conducted with the participants using the selected apparatus.

4.1.1 Apparatus selection

This section discusses the details of how the different apparatuses have been selected for the experiment as follows:

Serious Game Selection: Among a variety of application areas, this experiment has been conducted for serious games under the broad category of gaming context. As mentioned earlier serious games are focused towards particular purpose which are founded upon education, healthcare, medical training, physical training, military aspect and many more [11]. There are few characteristics which are used to classify different types of serious games which are – application area, activity, modality, interaction style and environment [49]. Basing on this taxonomy, the serious game for this experiment has been selected basing on the following criteria: application area – education, activity – physiological and mental, modality – visual, interaction style – tangible interfaces and environment – social presence. There are different levels of education depending on different subject matter and target users. A survey shows that 60% of the educational games of university level have been evaluated and about 10% of the educational games belong to the domain of computer science where basic programming knowledge based games cover a huge portion [13]. Moreover, it also shows 60% of computer games are evaluated whereas mobile games are only 6% in this respect although mobile devices tend to be very handy, user-friendly and interactive application platform. As a result, programming based mobile applications of computer science education have been taken into concern targeting undergraduate students. An exhaustive search has been performed in order to find a suitable application for the intended purpose where the criteria for searching have been defined as extensive programming curriculum from basic to advanced, highest downloads, good rating and availability. Depending in this criteria the selected application is ‘Programming Hero’ [103]. Figure 4.1 shows some snaps of the ‘Programming Hero’ mobile application.

Fig. 4.1. Snaps of the 'Programming Hero' mobile app

Defining Emotional State: Emotions have a wide range of variation which can be positive, negative or neutral and the intensity of these emotions can also vary from a high scale to low scale. So, it is important to follow a criteria in order to categorize the emotions into different classes. According to Russell's circumplex model of affect [104], emotions can be classified basing on two dimensions which are – arousal (ranges from activation to deactivation) and valence (ranges from pleasant to unpleasant). Each emotion can be defined depending on the varying extent of these two dimensions. Figure 4.2 shows some of the basic emotions classified into four quadrants according to Russell's circumplex model. For example, 'happy' can be defined as positive arousal-positive valence which can be interpreted as an emotion of highly activation and positive experience. Again, 'stressed' can be defined as positive arousal-negative valence which can be interpreted as an emotion of highly activation but negative experience. Now, this experiment is concerned with emotional experience as UX evaluation of the gaming context, particularly serious games. Existing works reveal different types of emotions have been observed in case of gaming experience. Such as for arcade [105], action [106] and adventure games [107] a number of positive and negative valence have been experienced. For serious games, such as social education [95], microbiology [97], undergraduate programming [99] themed games and the likes; some specific emotional experience have been observed among which engaging, motivating, attention, fun, tension, lacking concentration are very common which basically ranges from low arousal to high arousal. Emotional experience of serious games tends to vary basing on user's level of involvement or presence while they play the game which is measured within the scale of arousal dimension [108]. Hence, as for the intent of this experiment, 'activated' and 'deactivated' have been selected broadly within the scale of arousal dimension to evaluate emotional experience of serious games focused to education.

Evaluation Measure Selection: Evaluation measures for user-experience (UX) can be primarily classified into two categories – subjective and objective where subjective measures include interviews, rating scales, surveys etc. and objective measures include performance measure, physiological and neurological response etc. [5]. Among the objective measures, cardiovascular signal, brain wave signal, eye tracking metrics, facial expression and such are some of the common measures ranging from physiological to neurological [109]. These objective measures derive a variety of emotional levels for HCI-based applications' UX assessment and certain correlation exists between emotion and such physiological and neurological measures [110]. In the context of gaming experience, a huge amount of emotional and decision processing is required which consequently affects an individual's neurological and physiological attributes [90]. A review in [111, 112] shows that brain signal data and heart rate data are widely used to infer emotional experience as neurological and physiological measures to evaluate UX of different kinds of software applications. Hence, brain signal data from EEG (electroencephalography) technology and heart rate data from PPG (photo-

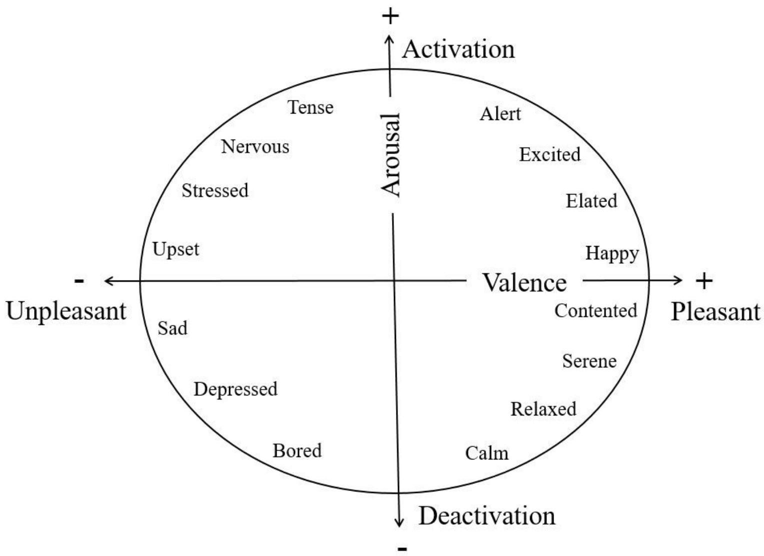


Fig. 4.2. Russell's circumplex model of affect

plethysmography) technology have been chosen for this experiment as evaluation measures for UX evaluation.

4.1.2 Participant profile

A total of 25 subjects participated in the experiment where 13 of them were male and the rest of the 12 participants were female. All the participants were undergraduate students aging from 20 to 25 studying in the department of computer science and engineering. The participants were familiar with having gaming experience using mobile applications particularly for various adventure and arcade games. They had basic knowledge of computer programming but most of them were not familiar with having learning experience through gaming in the field of computer programming education. Besides, none of the participants have played the selected serious game before.

4.1.3 Study procedure

The experiment was conducted in a laboratory environment through individual sessions. A consent form of the subjects' voluntary participation along with some biographical data were collected initially. Next, a brief instruction was provided to them regarding the experiment's aim and their responsibilities throughout the experimental session. They were also requested to download the particular mobile app during this session and a short overview was provided regarding the selected serious game. At first, all participants were instructed to sit comfortably and the wearable devices were equipped upon them for the required data collection. Then, each participant was told to start playing the game. All of them were instructed to play the first level of the game which was concerned with tutorials on basic programming fundamentals initially as well as some relevant quiz and tests at the end. The required duration for each participant to play the first level of the game was documented which was an average of 10 to 12 minutes. Figure 4.3 shows a participant playing the serious game application while being equipped with the data collection devices.

Fig. 4.3. Conducting the study procedure with a participant

4.2 Conducting the Experiment

This is the most extensive step of the experimental procedure. This includes data collection and data processing. The raw data shall be collected from relevant devices corresponding to the selected evaluation measures while the participants are using the application. After the raw data has been received, it would require some processing such as selecting features to be analyzed, filtering out undesired data, structuring the dataset and the likes for further analysis to infer potential and useful findings.

4.2.1 Data collection

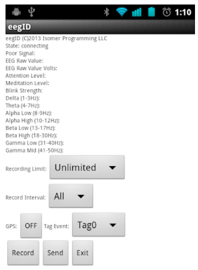
The devices chosen for this experiment for collecting data were wearable devices rather than using traditional and complex arrangement like external electrodes taking their comfort into consideration. For brain signal data, the Mindwave Mobile 2 EEG headset from Neurosky [113] and for heart rate data, Mi Smart Band 4 fitness tracker from Xiaomi [114] were used. As the participants went on playing the game, corresponding raw data were being collected through the devices in the following way:

Brain Signal Data Acquisition: The eegID [115] mobile application was used to collect the brain signal data. The chosen EEG headset connects to this app through Bluetooth connectivity and captures data continuously after a defined interval of time. As human emotion lasts for a very short period of time typically for seconds to minutes [116], the data was captured every 15 seconds. After each individual participant's session was completed, the collected data was exported as CSV file from the eegID app for further processing.

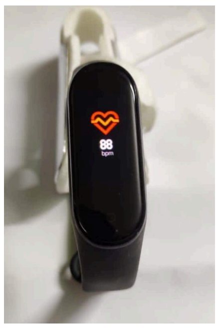
Heart Rate Data Acquisition: The MiFit [117] mobile application was used to collect the heart rate data. The chosen fitness tracker captures heart rate data every 1 minute. After each individual participant's session was completed, the collected data was exported as CSV file from the MiFit app through USB connectivity for further processing.

The data collection devices and associated applications are shown in Figure 4.4.

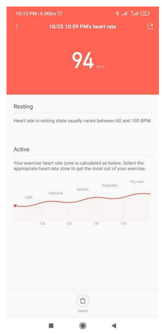
(a) Neurosky Mindwave Mobile 2 EEG headset



(b) eegID app



(c) Xiaomi Mi Smart Band 4



(d) MiFit app

Fig. 4.4. Devices and associated applications for collecting neurological (a), (b) and physiological (c), (d) data

4.2.2 Data processing

The collected data were next gathered for necessary processing according to the following order –

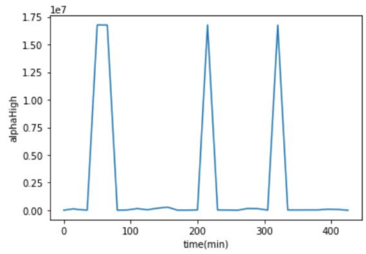
Data Preprocessing: It is evident that the collected data might contain different types of inconsistent data such as noise, outliers and other artifacts. Firstly, data cleaning was performed followed by data transformation. In case of brain signal data, the device has an integrated algorithm for filtering the signals through high pass and low pass filters known as Neurosky's proprietary algorithm [118]. Similarly, in case of heart rate data, the signals are filtered within the range of 40 to 170 bpm using a high pass and low pass filter as a human's heart function safely within this specified range of bpm [119]. Next, for data transformation, in order to infer emotional experience within the same timeframe with respect to brain signal data collection, the collected heart rate data of 1 minute was converted for 15 seconds. That is, each beat-per-minute value was transformed into beat-per-15 seconds.

Feature Selection: The collected data from the EEG headset contains relative amplitude value of five primary frequency bands of EEG – Alpha, Beta, Gamma, Delta and Theta [120]. Among these frequency bands, the activity of alpha low (8-9 Hz), alpha high (10-12 Hz), beta low (13-17 Hz) and beta high (18-30 Hz) brain waves mainly reflects an individual's emotion experience [121]. So, alpha low, alpha high, beta low and beta high features were considered to infer user's emotional experience from brain signal data as neurological measure of UX evaluation. The retrieved signals are represented as waveforms in Figure 4.5.

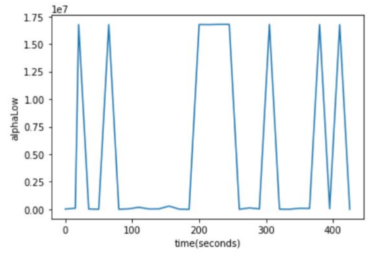
On the other hand, the collected data from the fitness tracker contains beat-per-minute (bpm) as heart rate data. Beat-per-minute (bpm) which is basically change in value of amplitude with respect to time basing on emotional reactivity of a user is a widely used feature found in smart wearable devices to measure heart rate data [122]. So, bpm, more specifically beat-per-15 seconds was considered to infer user's emotional experience from heart rate data as physiological measure of UX evaluation. The retrieved signal is represented as waveforms in Figure 4.6.

4.3 Data Analysis and Proposing the Framework

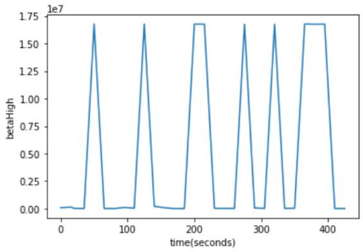
This section presents the extended proposed approach and how machine learning has been used to classify emotional states from the collected neurological and physiological data. Also, further analysis has been performed that provides some insightful findings to full-fill the research objectives.



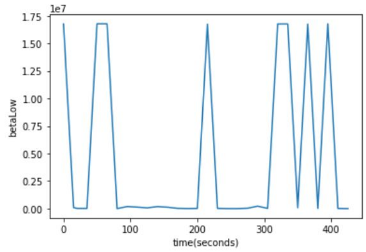
(a) Alpha high



(b) Alpha low



(c) Beta high



(d) Beta low

Fig. 4.5. Waveform of processed brain signals (a) Alpha high (b) Alpha low (c) Beta high and (d) Beta low

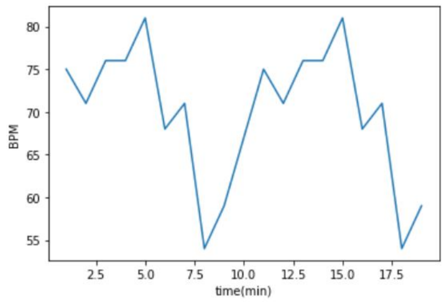


Fig. 4.6. Waveform of processed heart rate signal (bpm)

4.3.1 Emotional state classification

Basing on the collected data the two types of selected emotional experience – ‘activated’ and ‘deactivated’ for the selected serious gaming context has been classified from neurological features firstly and then, from machine learning techniques from both neurological and physiological features which is discussed as follows:

4.3.1.1 Classification of emotional experience from neurological features

This classification was performed using a defined procedure concerning frontal brain asymmetry activity based frequency band oscillations for emotion classification where the value of the selected neurological features corresponds to some emotions while playing games

Table 4.1: Conditions for the existing algorithm

Brainwave	State	Emotion
Alpha	Low	Stressed
	High	Over Relaxed
	Optimum	Relaxed
Beta	Low	Lack of Attention
	High	Anxious
	Optimum	Focused

[89]. The emotions classified here were – stressed, anxious, focused, lack of attention, relaxed and over-relaxed which can be plotted in Russell's circumplex model [104] as shown in Figure 4.7. Basing on this procedure, the conditions for stressed, anxious and focused were considered as activated emotion and the conditions for lack of attention, relaxed and over-relaxed were considered as deactivated emotion for this research. The conditions used in the algorithm [89] for mapping neurological features – alpha high, alpha low, beta high and beta low to different set of emotions in the defined procedure as well as the application of it in this research are displayed in Table 4.1. Finally, the emotional experience inferred from the neurological features using this procedure for the first 1 minute of 25 participants while they were playing the serious game application is illustrated in Table 4.2.

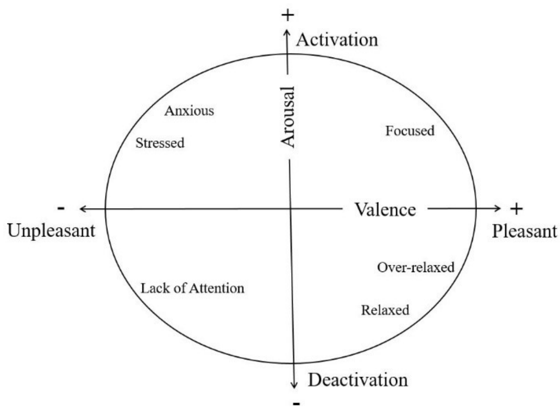


Fig. 4.7. Russell's circumplex model of affect mapped according to the existing algorithm

4.3.1.2 Classification of emotional experience from physiological and neurological features combined

This classification was performed using machine learning models with an intent to apply supervised learning for emotion classification using both neurological and physiological

Table 4.2: Emotional experience inference from neurological features for 25 participants (1 min)

Participants	**15 sec**	**30 sec**	**45 sec**	**60 sec**
P1	deactivated	activated	activated	activated
P2	activated	activated	activated	activated
P3	deactivated	activated	activated	activated
P4	deactivated	deactivated	deactivated	deactivated
P5	activated	activated	deactivated	activated
P6	deactivated	deactivated	deactivated	deactivated
P7	deactivated	activated	deactivated	deactivated
P8	activated	activated	activated	activated
P9	deactivated	deactivated	deactivated	activated
P10	activated	activated	activated	deactivated
P11	activated	deactivated	activated	activated
P12	activated	deactivated	deactivated	deactivated
P13	deactivated	activated	deactivated	deactivated
P14	activated	activated	deactivated	activated
P15	activated	deactivated	deactivated	activated
P16	activated	deactivated	deactivated	deactivated
P17	deactivated	activated	deactivated	activated
P18	deactivated	deactivated	deactivated	deactivated
P19	deactivated	activated	deactivated	deactivated
P20	activated	activated	deactivated	activated
P21	activated	deactivated	activated	activated
P22	activated	activated	activated	deactivated
P23	activated	deactivated	activated	activated
P24	activated	activated	deactivated	activated
P25	activated	activated	deactivated	deactivated

features. Supervised learning was chosen as it is preferably used more for emotion recognition [58]. For classifying emotions various supervised machine learning algorithms are used. Some recent reviews reveal that among traditional algorithms support vector machine (SVM), k-nearest neighbors (KNN) are mostly used; apart from these random forest (RF), linear discriminant analysis (LDA) and artificial neural network (ANN) are some other algorithms used in this regard [61, 62] Moreover, a review in [123] reveals that ensemble learning algorithms are rarely used for emotion classification from bio-signals whereas ensemble learning algorithms in machine learning are considered to improve performance of prediction [124]. As there are no specific criteria to select which machine learning algorithm is appropriate to elicit emotion for which type of dataset or context [60], both traditional and ensemble learning algorithms were taken into consideration to observe the classification accuracy. The algorithms considered were – SVM, KNN, RF, XgBoost, AdaBoost and CatBoost were implemented.

Proposed framework for emotional state classification:

Now, a 2-step machine learning approach was adopted to infer emotions from a combination of neurological and physiological features. In the first step, the aim is to find machine predicted beat-per-15 seconds value – the physiological feature from the neurological features (alpha high, alpha low, beta high, beta low) and corresponding labeled emotion as found from using the existing algorithm. In the second step, the aim is to infer emotion from the machine predicted beat-per-15 seconds value and neurological features.

1. *First step model learning:* The dataset was prepared as follows – the value of neurological (beat-per-15 seconds) and physiological (alpha high, alpha low, beta high, beta low) features of the participants with respect to time (15 seconds) were tabulated which was around 1500 rows of data and the emotions were labeled with respect to neurological features as mentioned previously (Table C1 shows a sample of the dataset). As a part of data preprocessing phase, the features were encoded and normalized. Also, the class labels were almost balanced, so no balancing techniques were required. Next, the dataset was divided into training and testing dataset as 80%

and 20% of the data respectively. The performance of eight different regression algorithms, such as RF, SVM, KNN, DT, AdaBoost, CatBoost, XgBoost and Naive Bayes, were compared in terms of accuracy. Based on higher accuracy, decision tree was selected for developing the first model. The output of this model was machine predicted beat-per-15 seconds values or the heart rate data. A graph representing machine predicted beat-per-15 seconds with respect to actual beat-per-15 seconds is shown in Figure 4.10.

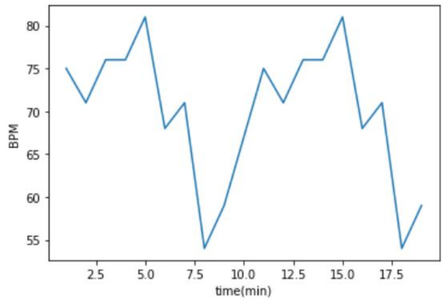


Fig. 4.8. Waveform of processed heart rate signal (bpm)

2. *Second step model learning:* Next, emotions were classified basing on the machine predicted physiological feature (beat-per-15 seconds) and neurological features (alpha high, alpha low, beta high, beta low). The six selected algorithms (SVM, KNN, RF, XgBoost, AdaBoost, CatBoost) were run on the dataset to build the model for this step and it showed that ensemble learning, particularly XgBoost provides comparatively better classification accuracy as well as better precision and recall (precision 0.916, recall 0.920) than traditional learning (See Figure 4.9)

Figure 4.11 shows the architecture of the presented 2-step machine learning approach.

The entire experimental procedure is summarized in Figure 4.12 where the three main phases - preparing experimental apparatus, conducting experimental study and analyzing experimental data are outlined along with the related sub-phases.

4.3.2 Correlation Analysis

This section discusses a brief analysis on the data that has been performed applying the 2-step procedure to classify emotional experience using machine learning techniques and

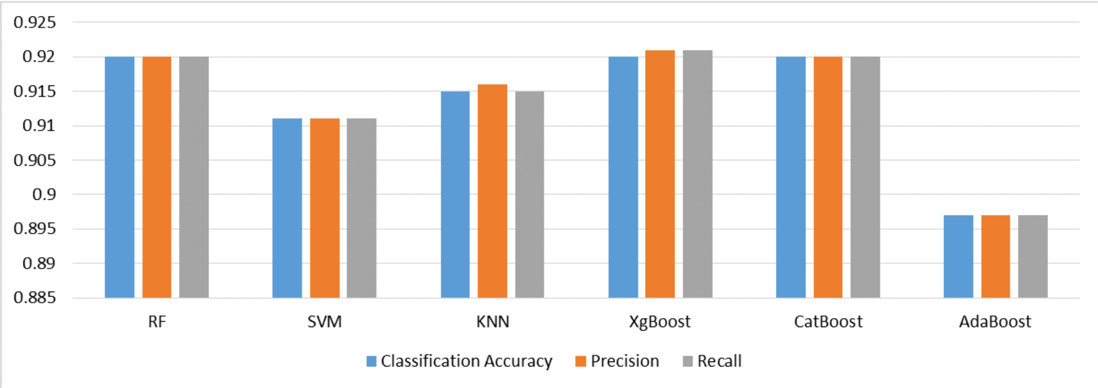


Fig. 4.9. The emotion classification accuracy, precision and recall metrics for the selected machine learning techniques

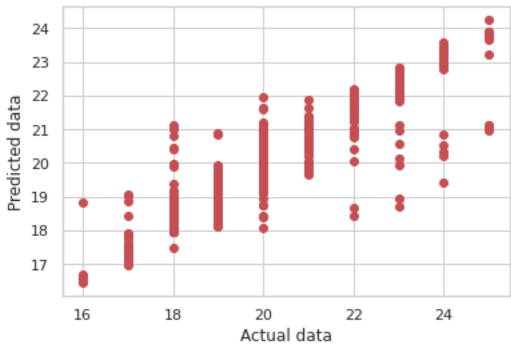


Fig. 4.10. Machine predicted beat-per-15 seconds with respect to actual beat-per-15 seconds

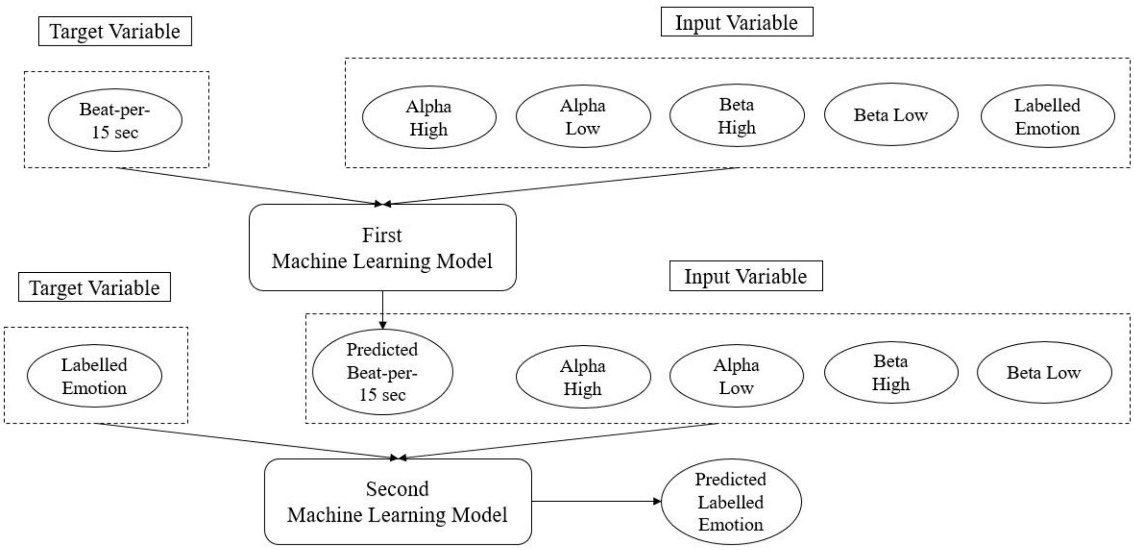


Fig. 4.11. Framework of the 2-step machine learning approach to predict emotions

some relevant findings afterwards.

The output from the second step model learning using XgBoost technique provides a score

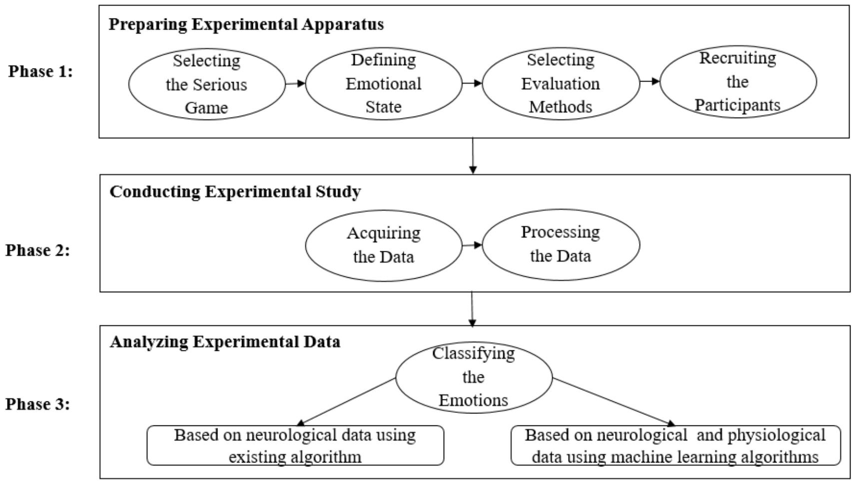


Fig. 4.12. Overview of the experimental procedure

metric called feature importance [125] which indicates how useful a feature is to predict the target variable that is, for this context, the emotions – activated and deactivated. The graph in Figure 4.13 shows the score of feature importance of the selected physiological and neurological measures. It denotes that beat-per-15 seconds has greater feature importance followed by alpha high, beta low, alpha low and beta high respectively to infer the corresponding emotions.

Next, the heart rate data (beat-per-15 seconds) of both actual dataset and machine predicted

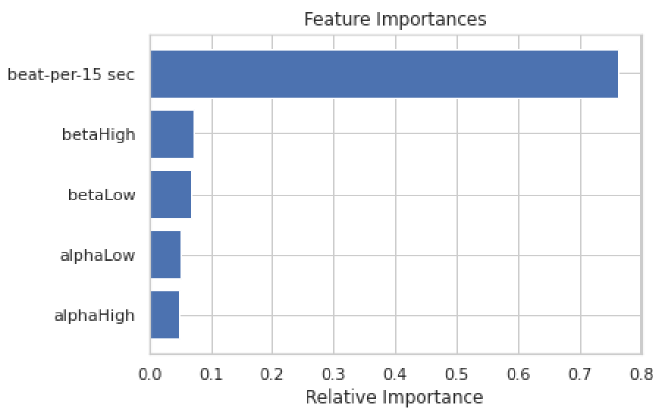


Fig. 4.13. Feature importance of the selected physiological and neurological measures

dataset was changed within the range of 18-25 of 25 participants keeping the neurological features (alpha high, alpha low, beta high, beta low) constant. This variation in beat-per-15 seconds with respect to constant neurological features depicts that emotional experience also varies in accordance with the variation of heart rate data as shown in the graphical representation of Figure 4.14.

Another graph (see Figure 4.15) regarding the machine predicted emotional experience

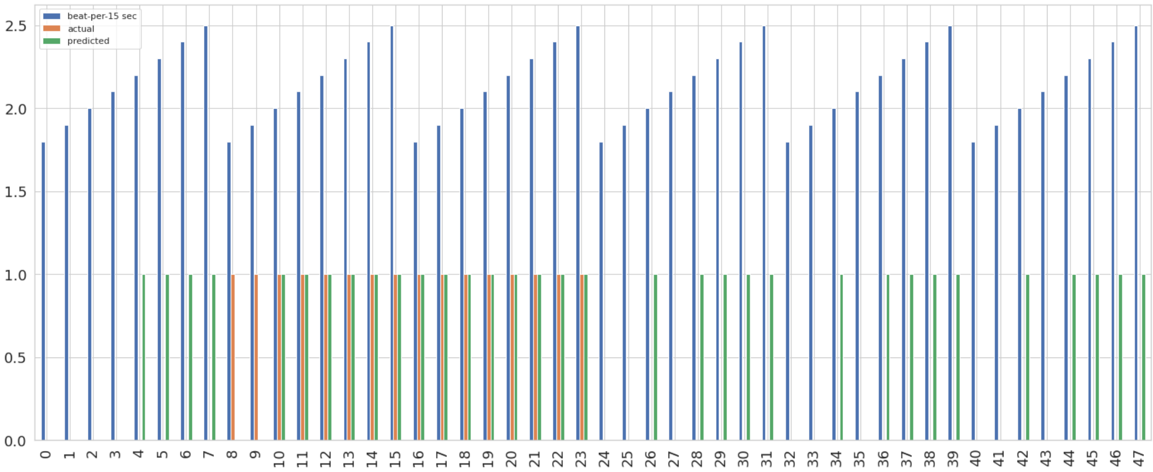


Fig. 4.14. Varying beat-per-15 seconds with respect to constant neurological features depicts varying emotional experience

prescribes that for this serious game application in the context of education, beat-per-15 seconds value below 17 often represents deactivated emotion and beat-per-15 seconds value above 22 often represents activated emotion. Besides, a beat-per-15 seconds value ranging from above 17 to 22 might infer activated or deactivated emotion.

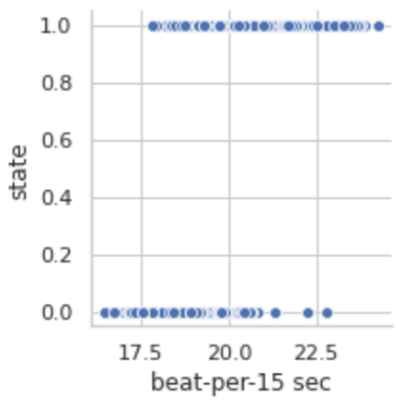


Fig. 4.15. Estimation of beat-per-15 seconds with respect to emotional state

Moreover, the algorithm of linear regression returns an equation for predicting the target feature while predicting the beat-per-15 seconds value. Fundamentally, linear regression aims for modeling an association between variables (generally two but can be more than two). The association is made by constructing a linear equation for the data observed where a variable is termed as an independent variable, and the other one is termed as a dependent variable [126]. The basic format of the equation that a linear regression line possess is as the equation given in equation 4.1, where X is the variable known as independent variable and Y is the variable known as the dependent variable. Moreover, β_1 is known as the slope of the line and the intercept is β_0 which is basically the value of Y when the value of X is zero. In mathematical terms; X, Y are variables and β_0, β_1 are constant.

$$\begin{aligned}
&Y = \beta_0 + \beta_1 * X_1 \\
&X = independent\ variable \\
&Y = dependent\ variable \\
&\beta_1 = slope \\
&\beta_0 = intercept
\end{aligned} \tag{4.1}$$

For multiple independent variables, the equation for the linear regression would be as follows:

$$\begin{aligned}
Y &= \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + ...\beta_n * X_n \\
X_1, X_2....X_n &= multiple\ independent\ variables \\
Y &= dependent\ variable \\
\beta_1, \beta_2....\beta_n &= slope \\
\beta_0 &= intercept
\end{aligned} \tag{4.2}$$

Equation 4.2 can be written in a more compact form as follows using the summation notation:

$$Y = \sum_{i=0}^{n} \beta_i * X_i \tag{4.3}$$

The equation generated from the proposed linear regression model for the context of this research is as follows:

$$beat - per - 15\ seconds = 18.636 + 0.015 * alphaLow + 0.108 * alphaHigh$$
$$-0.128 * betaLow - 0.211 * betaHigh + 2.378 * labelled\ emotional\ state \tag{4.4}$$

Similarly, as equation 4.3, equation 4.4 can be written as follows in the form of summation notation:

$$beat - per - 15\ seconds = \sum_{i=0}^{5} \beta_i * X_i \tag{4.5}$$

Here, β_0 = 18.636, β_1 = 0.015, β_2 = 0.108, β_3 = -0.128, β_4 = -0.211, β_5 = 2.378
X_1 = alphaLow, X_2 = alphaHigh, X_3 = betaLow, X_4 = betaHigh, X_5 = labelled emotional state

In the equation, each feature is associated with their respective weights. A fixed constant value is also associated with the equation. Apart from betaHigh and betaLow, other features have positive weights whereas, the constant value in the equation having a high value of 18.636. This equation contributes to determine machine predicted beat-per-15 seconds value from neurological features (alpha low, alpha high, beta low and beta high) and labelled emotional states.

CHAPTER FIVE

CONCLUSION

This chapter provides the concluding remarks which is discussed into a number of modules. These are thesis outcomes, thesis implications, thesis limitations and future work. The thesis outcomes are outlined in a structured way firstly. Then, implications of this thesis are described. Next, some limitations of this thesis work are stated followed by a number of potential opportunities for future work.

5.1 Thesis Outcomes

A number of findings have been found from this experimental study which are outlined and discussed below as the outcomes of this thesis:

a. **A practical approach to infer emotion from psychophysiological measures for serious games:** A machine learning based 2-step predictive modeling approach has been presented in this research. It follows a procedure where the output of the first model is fed to the input of the second model. The first model predicts heart rate basing on the neurological features and associated (labeled) emotions found using BCI technology. This machine predicted heart rate is fed to the second model which predicts emotions – activated and deactivated along with the neurological features. In summary, the neurological feature based emotion labeling is used as a baseline to predict emotional experience using both neurological and physiological features through the presented approach. Past studies showed that, particularly for serious games emotions are generally classified using subjective measures (questionnaires, surveys etc.) mostly [13, 127, 96], while, this research presents a practical approach to classify emotional states from neurological and physiological measures for serious gaming context.

b. **Correlation between neurological and physiological measures to infer emotions:** In this research, firstly, emotions were inferred from neurological features using BCI technology. Then, heart rate data as physiological feature was incorporated along with the neurological features to infer emotions. That is, a multimodal evaluation measure-based approach was adopted to observe whether any sort of dependency prevails between these two measures while inferring emotional experience. It is found that, when both neurological and physiological measures are combined to infer emotions, the feature importance of heart rate data (physiological feature) is higher than the neurological features. Again, if heart rate data is changed within a certain range keeping the neurological features constant, the emotions also vary basing on the heart rate data. Therefore, there definitely exists a correlation between neurological and physiological measures to some extent when they are used to infer emotional experience of users in the context of educational serious games. This finding is aligned with outcomes of few existing studies that multi-modal evaluation measures are often preferred over using a single evaluation measure for assessing UX [12]. Hence, the correlation between neurological and physiological measure in this experimental study persuades towards using multi-modal measures in order to obtain better evaluation results.

c. **Ensemble learning provides better emotion prediction performance for educational serious games:** The emotions have been predicted using both traditional and ensemble learning algorithms. Regarding traditional learning and ensemble learning, XgBoost and RF gives better classification accuracy than all other selected algorithms. But, as shown in Figure 4.9, the precision and recall metric of XgBoost is more than RF. Precision and recall bears much importance in case of classification process because it specifies the relevancy of the classification results and weight of this relevancy [125]. Moreover, ensemble learning is considered to provide better prediction results by some studies [124]. For this study, it appears to be true and consequently it can be said that, ensemble learning predicts emotions with a comparatively higher performance for educational serious games than the traditional algorithms.

d. **Heart rate (bpm) estimation for emotional experience in the context of educational serious games:** Another important revelation from this experimental study lies in estimating the heart rate or beat-per-minute value while the users play the serious game application. It has been observed from the predicted heart rate that, if the predicted bpm value is below 68, then the emotion is classified as deactivated and if the predicted bpm value is over 88, then the emotion is classified as activated. According to sources of health information, lower heart rate usually refers to relaxed state and higher heart rate usually refers to stressed state [128]. Hence, the estimated bpm value for emotional experience in the context of educational serious games is quite reasonable.

e. **An equation for predicting beat-per-15 seconds from the proposed linear regression model:** Linear regression model helps reveal useful information by generating equations from dependent and independent variables [126]. Using linear regression model, an equation has been generated which could predict beat-per-15 seconds value (physiological feature) from alpha low, alpha high, beta low and beta high (neurological feature), and labelled emotional state. It is a multiple linear regression model which uses neurological features and labelled emotional state as independent variables and beat-per-15 seconds as dependent variable. The value of the coefficients that is associated with the independent variables and the constant value provides a prediction regarding what could be the possible value of beat-per-15 seconds for different values of the independent variables. As a consequence, if some values of neurological features and emotional state is known, beat-per-15 seconds and afterwards beat-per-minute or heart rate could be predicted easily while users are interacting with a serious game application.

5.2 Thesis Implications

In this thesis, an extended machine learning based 2-stage emotion classifying approach is proposed to infer emotion from psychophysiological data which contributes to the practical implication of this thesis. The proposed approach follows a multistage procedure where the output of the first model is fed to the input of the second model. Here, the neurological data based emotion labeling is used as a baseline to classify emotional experience using both neurological and physiological data in this entire process of the proposed extended approach. This shows the practical application of machine learning techniques in the field of UX evaluation. Furthermore, this research showed that the feature importance and the variation in heart rate data with respect to constant features of brain signal data depicts a correlation to a certain extent between neurological and physiological data. These findings can be considered together to infer emotional experience of users for evaluating UX in the context of educational serious games. Furthermore, this will eventually aid UX designers to design more intuitive applications to full-fill the intended purpose of serious games.

However, there are some theoretical implications too. It is evident that UX evaluation basing on users' emotional experience is very important in the context of gaming experience. Consequently, this study demonstrates how users' neurological and physiological data have an immense impact on users' emotional experience. This article discusses an experimental study in the context of serious games, particularly for a programming-based mobile app which provides valuable insights in this regard. These insights help to understand the applicability of objective measures, particularly psychophysiological measures in the research area of serious games that is concerned with UX evaluation. Moreover, the study explores how emotional states can be inferred from neurological and physiological data by using a proposed approach based on machine learning techniques which shows how a relation between the broad area of UX evaluation and machine learning techniques is established. The outcomes of this experimental study will greatly help to contribute in the field of HCI for understanding how emotional experience of a user can be used to measure UX of an educational serious game by using user's neurological and physiological data.

5.3 Thesis Limitations

Few limitations exist for this thesis which are described below:

a. **Subjective measure is not considered for comparison:** The emotions as classified by the proposed extended machine learning-based approach have not been compared with users' subjective feedback regarding their emotional experience while playing the serious game. As physiological and neurological data were collected every 15 seconds to infer emotions, so variation of emotion according to users' subjective feedback within that small amount of time-frame was not possible to collect reliably.

b. **Only one application of a particular serious game context is evaluated for UX:** Serious games are intended towards different types of teaching and training programs. This experimental study is concerned with educational based serious games which falls in the domain of computer science. Computer science consists of a wide variety of topics. Here, UX is evaluated for only a particular software application which is aimed towards teaching basic computer programming skills.

c. **Only mobile application platform is evaluated for UX:** There exists a variety of application platforms such as web-based, mobile-based, embedded platforms and such [129]. Also, there exists cross-platform applications which are a basically a single application distributed over multiple platforms for improving accessibility. The selected application of this research that is evaluated for UX is only based on mobile platform.

d. **Number of participants is not adequate:** The total number of participants who took participation in the experimental study was 25. As the evaluated application is relevant to computer programming so the intended participants is required to have some familiarity with computer programming knowledge. Although utmost effort was provided to collect data from maximum participants possible from this particular domain but the experimental study was conducted during corona pandemic situation and so due to safety issues the number of participants were less.

5.4 Future Work

There prevails a number of ways to enhance this research work in future. These are discussed as follows:

a. **Comparing the output of the proposed framework with users' subjective feedback:** The future work would be to find a reliable approach to collect the subjective feedback from users' regarding their emotional experience within the defined short interval of time. Afterwards, comparing the data of the subjective feedback with respect to the inferred emotional states from the proposed extended framework. This will help to investigate any sort of deviation between emotional states basing on the two types of evaluation methods, that is, users' subjective feedback and outcomes from the proposed framework.

b. **Comparing among other similar programming-based software applications:** There are different types of software applications available which is purposed for teaching basic computer programming skills. Such similar programming-based applications can be explored using the proposed extended approach. This will result in a comparative study which will help to identify which UX elements are appreciated by users and which creates unpleasant experience for them. Eventually, this type of findings will lead to encourage UX designers for improving their designs in order to receive better acceptability from users.

c. **Increasing the number of participants with varying knowledge level:** Recruiting more number of participants for the experimental study is another work to be done to enhance the findings of the research. The more the data, the more variety of trends could be observed which will help to establish an argument more strongly. Moreover, participants from different knowledge level of computer programming can be included to find out whether any variation exists in UX assessment for varying knowledge level of a certain area of expertise. Thus, increasing participants with varying knowledge level might bring a number of viable outcomes.

d. **Comparing applications distributed among different software platforms:** UX is evaluated in this study for a mobile-based application platform. As mobile devices are considered to be handy, portable, widely used and a user-friendly computing device, so a mobile-based application was selected for the purpose of UX evaluation. However, applications are developed over a wide variety of software platforms these days. Application developed on different platforms like mobile-based, web-based etc. can be taken into consideration to explore whether the type of software platform creates any impact on user experience while users interact with an application.

e. **Using multiple number of neurological and physiological evaluation measures:** Neurological and physiological measures fall under the broad criteria of objective methods of UX evaluation. Basing on the findings of the literature review, it has been found that heart rate data as neurological measure and brain signal data as physiological measure are widely used for evaluating UX. Therefore, heart rate data and brain signal data were selected for evaluating UX of the selected mobile application. Only one type of measure of both neurological and physiological measure was used for evaluation in this research. But, a lot more useful and variant findings could be achieved from using multiple neurological and physiological measures which could contribute to the field of UX evaluation greatly.

f. **Using improved devices for collecting data:** Two separate wearable devices were used to collect the required data. These devices were easily available as well as comforting for users. An activity tracker based smart watch called Mi Band and a headset called Neurosky Mindwave Mobile 2 was used to collect heart rate data and brain signal data respectively. The systematic literature review showed that these devices are commonly used for this purpose. Hence, these devices were selected for data collection. Anyhow, more improved devices for similar purpose exists nowadays like Emotive EPOC X [130] wireless device for collecting brain signal data, ProComp Infiniti System [131] for collecting both brain signal and heart rate data. Such improved devices could be used to collect the physiological and neurological data more reliably in order to perform a more accurate data analysis.

g. **Exploring other research opportunities:** As from the systematic literature review, it can be found that quite a good number of research opportunities have been outlined; that includes exploring the significance of using a single or both evaluation measures, investigating priority consideration in multimodal approach for evaluation, a paradigm shift in technology for selecting evaluation measures, change in conventional comparative approach among evaluation measures, expanding research for less evaluated application context, enhancing evaluation for different variety of a particular application context, incorporating emotional experience as evaluation measure for serious games, context-based mapping with respect to evaluation measures and defining a complete set of metrics for neurological and physiological evaluation measures. Among these, one research opportunity relevant to the applicability of evaluation measures for the purpose of UX evaluation of serious games has been investigated through an experimental study in this thesis. Accordingly, some promising possibilities of research lie in the remaining variety of opportunities which could be explored in future to bring out the fruitful findings and contribute to the research community of HCI.

REFERENCES

[1] Y. Rogers, "HCI theory: classical, modern, and contemporary," *Synthesis lectures on Human-Centered Informatics*, vol. 5, no. 2, pp. 1–129, 2012.

[2] A. Schmidt, "The user experience," *Library Journal*, vol. 135, no. 1, pp. 28–29, 2010.

[3] "Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems, ISO 9241-210:2010(en)," Accessed on: Dec. 20, 2020. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en.

[4] D. Benyon, *Designing User Experience: a guide to HCI, UX and Interaction design*. Pearson, UK, 2019.

[5] K. Hornbæk, "Current practice in measuring usability: Challenges to usability studies and research," *International Journal of Human-Computer Studies*, vol. 64, no. 2, pp. 79–102, 2006.

[6] J. Jankowski, J. Watróbski, K. Witkowska, and P. Ziemba, "Eye tracking based experimental evaluation of the parameters of online content affecting the web user behaviour," in *Selected Issues in Experimental Economics*. Springer, Cham, 2016, pp. 311–332.

[7] Y. Trindade, F. Rebelo, and P. Noriega, "Potentialities of a face reading tool to a digital game evaluation and development: A preliminary study," in *Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics, Los Angeles, California, USA, 17-21 July 2013*. Springer, 2017, pp. 371–381.

[8] A. Baraybar-Fernández, M. Baños González, O. Barquero-Pérez, R. Goya-Esteban, and A. De-la Morena-Gómez, "Evaluation of emotional responses to television ad-

vertising through neuromarketing," *Comunicar: Media Education Research Journal*, vol. 25(2), pp. 19–28, 2017.

[9] P. Brooks and B. Hestnes, "User measures of quality of experience: Why being objective and quantitative is important," *IEEE Network*, vol. 24(2), pp. 8–13, 2010.

[10] R. Bernhaupt, Ed., *Evaluating user experience in games: Concepts and methods*. Springer, UK, 2010.

[11] J. Breuer and G. Bente, "Why so serious? on the relation of serious games and learning," *Journal for Computer Game Culture*, vol. 4, no. 1, pp. 7–24, 2010.

[12] J. Bardzell, S. Bardzell, T. Pace, and J. Karnell, "Making user engagement visible: A multimodal strategy for interactive media experience research," in *Proceedings of the CHI'08 extended abstracts on Human Factors in Computing Systems, Florence, Italy, 5-10 April 2008*. ACM, 2008, pp. 3663–3668.

[13] A. Calderón and M. Ruiz, "A systematic literature review on serious games evaluation: An application to software project management," *Computers & Education*, vol. 87, pp. 396–422, 2015.

[14] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, *Human-computer interaction*. Addison-Wesley Longman Ltd., UK, 1994.

[15] A. Blandford, J. Gibbs, N. Newhouse, O. Perski, A. Singh, and E. Murray, "Seven lessons for interdisciplinary research on interactive digital health interventions," *Digital Health*, vol. 4, no. 2018, p. 2055207618770325, 2018.

[16] N. S. Shami, J. T. Hancock, C. Peter, M. Muller, and R. Mandryk, "Measuring affect in hci: going beyond the individual," in *CHI'08 extended abstracts on Human Factors in Computing Systems*. ACM, 2008, pp. 3901–3904.

[17] T. T. Hewett, R. Baecker, S. Card, T. Carey, J. Gasen, M. Mantei, G. Perlman, G. Strong, and W. Verplank, *ACM SIGCHI curricula for human-computer interaction*. USA: ACM, 1992.

[18] T. Zaki, Z. Sultana, S. A. Rahman, and M. N. Islam, "Exploring and comparing the performance of design methods used for information intensive websites," *MIST International Journal of Science and Technology*, vol. 8, pp. 49–60, 2020.

[19] N. Hasan and M. N. Islam, "Exploring the design considerations for developing an interactive tabletop learning tool for children with autism-spectrum disorder," in *Proceedings of the International Conference on Computer Networks, Big data and IoT, Madurai, India, 19-20 December 2019*. Springer, 2019.

[20] M. N. Islam, M. A. Ahmed, and A. N. Islam, "Chakuri-bazaar: A mobile application for illiterate and semi-literate people for searching employment," *International Journal of Mobile Human Computer Interaction*, vol. 12(2), pp. 22–39, 2020.

[21] I. Wechsung, A. B. Naumann, and R. Schleicher, "Views on usability and user experience: from theory and practice," in *Proceedings of the NordiCHI Conference, Lund, Sweden, 20-22 October 2008*. ACM, 2008.

[22] C. Rusu, V. Rusu, S. Roncagliolo, and C. González, "Usability and user experience: what should we care about?" *International Journal of Information Technologies and Systems Approach*, vol. 8(2), pp. 1–12, 2015.

[23] D. Sward and G. Macarthur, "Measuring affect in hci: going beyond the individual," in *Workshop on Towards a UX Manifesto*. ACM, 2007, pp. 35–40.

[24] M. Georgsson and N. Staggers, "Quantifying usability: an evaluation of a diabetes mhealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics," *Journal of the American Medical Informatics Association*, vol. 23, no. 1, pp. 5–11, 2016.

[25] N. Bevan, "Classifying and selecting ux and usability measures," in *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement, Reykjavik, Iceland, 13-18 June 2008*. ACM, 2008.

[26] J. Mifsud, "The difference (and relationship) between usability and user experience," Accessed on: Dec. 20, 2020. [Online]. Available: https://usabilitygeek.com/the-difference-between-usability-and-user-experience/.

[27] M. N. Islam, M. M. Karim, T. T. Inan, and A. N. Islam, "Investigating usability of mobile health applications in Bangladesh," *BMC Medical Informatics and Decision Making*, vol. 20(1), p. 19, 2020.

[28] M. N. Islam and H. Bouwman, "Towards user-intuitive web interface sign design and evaluation: A semiotic framework," *International Journal of Human-Computer Studies*, vol. 86, pp. 121–137, 2016.

[29] M. N. Islam, H. Bouwman, and A. N. Islam, "Evaluating web and mobile user interfaces with semiotics: An empirical study," *IEEE Access*, vol. 8, pp. 84 396–84 414, 2020.

[30] K. M. Munim, I. Islam, M. Khatun, M. M. Karim, and M. N. Islam, "Towards developing a tool for UX evaluation using facial expression," in *Proceedings of the 3rd International Conference on Electrical Information and Communication Technology, Khulna, Bangladesh, 7-9 December 2017*. IEEE, 2017.

[31] N. McNamara and J. Kirakowski, "Functionality, usability, and user experience: Three areas of concern," *Interactions*, vol. 13(6), pp. 26–28, 2006.

[32] T. K. Landauer, "Research methods in human-computer interaction," in *Handbook of human-computer interaction*. London: North-Holland, 1988, pp. 905–928.

[33] S. Howard and D. M. Murray, "A taxonomy of evaluation techniques for hci," in *Proceedings of the Second IFIP Conference on Human-Computer Interaction, Federal Republic of Germany, 1-4 September 1987*. North Holland, 1987.

[34] D. Biduski, E. A. Bellei, J. P. M. Rodriguez, L. A. M. Zaina, and A. C. B. De Marchi, "Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire," *Computers in Human Behavior*, vol. 104, p. 106169, 2020.

[35] J. Staiano, M. Menéndez, A. Battocchi, A. De Angeli, and Sebe, "Ux_mate: from facial expressions to ux evaluation," in *Proceedings of the Designing Interactive Systems Conference, Newcastle Upon Tyne, UK, 11-15 June 2012*. ACM, 2012.

[36] "Objective vs. subjective evaluation criteria," Accessed on: Dec. 20, 2020. [Online]. Available: https://classroom.synonym.com/objective-vs-subjective-evaluation-criteria-8265343.html.

[37] B. U. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Helenius, L. Ahonen, O. Barral Mery de Bellegarde, I. J. Kosunen, T. Valtonen, M. J. Huotilainen, and J. N. Ravaja, "The psychophysiology primer: A guide to methods and a broad review with a focus on human–computer interaction," *Foundations and Trends in Human-Computer Interaction*, vol. 9(3-4), pp. 151–308, 2015.

[38] B. Park, "Psychophysiology as a tool for HCI research: promises and pitfalls," in *Proceedings of the 13th International Conference on Human-Computer Interaction, California, USA, 9-24 July 2009*. Springer, 2009.

[39] A. C. Dirican and M. Göktürk, "Psychophysiological measures of human cognitive states applied in human computer interaction," *Procedia Computer Science*, vol. 3, pp. 1361–1367, 2011.

[40] M. Hassenzahl and N. Tractinsky, "User experience-a research agenda," *Behaviour & Information Technology*, vol. 25, no. 2, pp. 91–97, 2006.

[41] L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," *Annual Review of Psychology*, vol. 58, no. 2007, pp. 373–403, 2007.

[42] P. Saraswat, "The design of emotions and emotional intelligence," Accessed on: Dec. 20, 2020. [Online]. Available: https://uxdesign.cc/the-design-of-emotions-and-emotional-intelligence-ba00855107d3?gi=34c7a8b93582.

[43] J. Forlizzi and K. Battarbee, "Understanding experience in interactive systems," in *The 5th Conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 2004, pp. 261–268.

[44] P. M. Desmet and P. Hekkert, "Special issue editorial: Design & emotion," *International Journal of Design*, vol. 3, no. 2, p. 00, 2009.

[45] E. L. C. Law and S. Abrahão, "Interplay between user experience (ux) evaluation and system development," *International Journal of Human-Computer Studies*, vol. 72, no. 6, pp. 523–525, 2014.

[46] L. E. Nacke, P. Mirza-Babaei, and A. Drachen, "User experience (ux) research in games," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, pp. 1–4.

[47] D. Vlachopoulos and A. Makri, "The effect of games and simulations on higher education: a systematic literature review," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, p. 22, 2017.

[48] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*. Muska Lipman/Premier-Trade, 2005.

[49] F. Laamarti, M. Eid, and A. El Saddik, "An overview of serious games," *International Journal of Computer Games Technology*, vol. 2014, 2014.

[50] W. L. Johnson and A. Valente, "A tactical language and culture training systems: Using ai to teach foreign languages and cultures," *AI magazine*, vol. 30, no. 2, p. 72, 2009.

[51] A. Sliney and D. Murphy, "Jdoc: A serious game for medical learning," in *The 1st International Conference on Advances in Computer-Human Interaction*. IEEE, 2008, pp. 131–136.

[52] "Serious games futuring medical training in france," Accessedon: Dec. 20,

2020. [Online]. Available: https://www.seriousgamemarket.com/2009/11/serious-games-futuring-medical-training.html.

[53] “Serious games as a playground for learning math,” Accessed on: Dec. 20, 2020. [Online]. Available: https://www.seriousgamemarket.com/2013/12/serious-games-as-playground-for.html.

[54] “Serious games creating cultural experiences to reinforce language-learning,” Accessedon: Dec. 20, 2020. [Online]. Available: https://www.seriousgamemarket.com/2011/01/serious-games-creating-cultural.html.

[55] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, “Machine learning,” *Neural and Statistical Classification*, vol. 13, no. 1994, pp. 1–298, 1994.

[56] I. Goodfellow, Y. Bengio, and A. Courville, “Machine learning basics,” *Deep learning*, vol. 13, no. 2016, pp. 98–164, 2016.

[57] “Real-world machine learning: model evaluation and optimization,” Accessed on: Dec. 20, 2020. [Online]. Available: https://www.developer.com/mgmt/real-world-machine-learning-model-evaluation-and-optimization.html.

[58] C. A. Devi, D. KarthikaRenuka, and S. Soundarya, “A survey based on human emotion identification using machine learning and deep learning,” *Journal of Computational and Theoretical Nanoscience*, vol. 15, no. 5, pp. 1662–1665, 2018.

[59] T. G. Dietterich, *The handbook of brain theory and neural networks*. UK: MIT Press, 2003.

[60] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, “Physiological signals based human emotion recognition: a review,” in *Proceedings of the 2011 IEEE 7th International Colloquium on Signal Processing and its Applications*. IEEE, 2011, pp. 410–415.

[61] A. F. Bulagang, N. G. Weng, J. Mountstephens, and J. Teo, “A review of recent

approaches for emotion classification using electrocardiography and electrodermography signals," *Informatics in Medicine Unlocked*, vol. 20, p. 100363, 2020.

[62] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: A review," *Electronic Notes in Theoretical Computer Science*, vol. 343, pp. 35–55, 2019.

[63] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," *International Journal of Computer Trends and Technology*, vol. 48(3), pp. 128–138, 2017.

[64] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications*. UK: Springer Science & Business Media, 2012.

[65] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, Tech. Rep., 2007.

[66] B. Kitchenham, "Procedures for performing systematic reviews," Keele University Technical Report TR/SE-0401 and NICTA Technical Report 0400011T.1, Tech. Rep., 2004.

[67] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. P. Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimedia Tools and Applications*, vol. 76(18), pp. 19 087–19 111, 2017.

[68] J. Wei, H. Luo, S. J. Wu, P. P. Zheng, G. Fu, and K. Lee, "Transdermal optical imaging reveal basal stress via heart rate variability analysis: A novel methodology comparable to electrocardiography," *Frontiers in Psychology*, vol. 9, p. 98, 2018.

[69] E. Ganglbauer, J. Schrammel, S. Deutsch, and M. Tscheligi, "Applying psychophysiological methods for measuring user experience: Possibilities, challenges and feasibility," in *Proceedings of the Workshop on user experience evaluation methods in product development, Uppsala, Sweden, 24-28 August 2009*. Springer, 2009.

[70] G. Ural, F. Kaçar, and S. Canan, “Wavelet phase coherence estimation of EEG signals for neuromarketing studies,” *NeuroQuantology*, vol. 17(2), pp. 112–120, 2019.

[71] T. Lin and A. Imamiya, “Evaluating usability based on multimodal information: An empirical study,” in *Proceedings of the 8th International Conference on Multimodal Interfaces Banff, Alberta, Canada, 2-4 November 2006*. ACM, 2006, pp. 364–371.

[72] S. M. Coyle, T. E. Ward, and C. M. Markham, “Brain-computer interface using a simplified functional near-infrared spectroscopy system,” *Journal of Neural Engineering*, vol. 4(3), p. 219, 2007.

[73] C. Harvey, N. A. Stanton, C. A. Pickering, M. McDonald, and P. Zheng, “A usability evaluation toolkit for in-vehicle information systems (iviss),” *Applied Ergonomics*, vol. 42(4), pp. 563–574, 2011.

[74] D. Quiñones, C. Rusu, and V. Rusu, “A methodology to develop usability/user experience heuristics,” *Computer Standards & Interfaces*, vol. 59, pp. 109–129, 2018.

[75] D. T. Toledano, R. Fernández Pozo, A. Hernández Trapote, and L. Hernández Gómez, “Usability evaluation of multi-modal biometric verification systems,” *Interacting with Computers*, vol. 18(5), pp. 1101–1122, 2006.

[76] J. Vong and I. Song, “Automated health care services,” in *Emerging Technologies for Emerging Markets, Singapore, vol.11, pp 89-102*. Springer, 2015.

[77] H. Boz, A. Arslan, and E. Koc, “Neuromarketing aspect of tourısm pricing psychology,” *Tourism Management Perspectives*, vol. 23, pp. 119–128, 2017.

[78] V. Boricean, “Neuromarketing - a technique to go inside the consumers mind,” *Sibiu Alma Mater University Journals, Series A. Economic Sciences*, vol. 2(3), pp. 183–189, 2009.

[79] M. Durkin, S. McKenna, and D. Cummins, “Emotional connections in higher education marketing,” *International Journal of Educational Management*, vol. 12, pp. 153–161, 2012.

[80] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä, and N. Ravaja, "A review of the use of psychophysiological methods in game research," *Journal of Gaming & Virtual Worlds*, vol. 3(3), pp. 181–199, 2011.

[81] Y. J. Jung, H. Y. Kim, T. Y. Kim, Y. Lee, and S. Y. Yoon, "Preliminary evidence for the psychophysiological effects of technologic feature in e-commerce," in *Proceedings of the International Textile and Apparel Association Annual Conference, St. Petersburg, Florida, USA, 14-18 November 2017*. Iowa State University Digital Press, 2017.

[82] S. Mahlke, "User experience: Usability, aesthetics and emotions in human-technology interaction," in *Proceedings of the Workshop on Towards a UX Manifesto, University of Lancaster, UK, 3-7 September 2007*. eWiC, 2007.

[83] T. Susi, M. Johannesson, and P. Backlund, "Serious games: An overview," Department of Communication and Information, Skövde, Tech. Rep., 2007.

[84] L. Anolli, F. Mantovani, L. Confalonieri, A. Ascolese, and L. Peveri, "Emotions in serious games: From experience to assessment," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 5, no. 2010, 2010.

[85] T. Avinash, L. Dikshant, and S. Seema, "Methods of neuromarketing and implication of the frontal theta asymmetry induced due to musical stimulus as choice modeling," in *Proceedings of the 2018 International Conference on Computational Intelligence and Data Science, Gurugram, India, 7-8 April 2018*. Elsevier, 2018, pp. 55–67.

[86] H. Petrie and N. Bevan, "The evaluation of accessibility, usability, and user experience," in *The Universal Access Handbook*, 1st ed., C. Stephanidis, Ed. Florida: CRC Press, 2009, vol. 1, pp. 1–6.

[87] M. N. Islam, T. T. Inan, N. T. Promi, S. Z. Diya, and A. K. M. N. Islam, "Design, implementation, and evaluation of a mobile game for blind people: Toward making mobile fun accessible to everyone," in *Information and Communication Technologies for Humanitarian Services, UK, pp. 291-312*. IET, 2020.

[88] Q. Yu, X. Che, S. Ma, S. Pan, Y. Yang, W. Xing, and X. Wang, "A hybrid user experience evaluation method for mobile games," *IEEE Access*, vol. 6, pp. 49 067–49 079, 2018.

[89] S. Z. Diya, R. A. Proma, I. I. Rahman, A. B. Islam, and M. N. Islam, "Applying brain-computer interface technology for evaluation of user experience in playing games," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, Cox'sBazar, Bangladesh, 2019, pp. 1–6.

[90] L. E. Nacke, "Games user research and physiological game evaluation," in *Game user experience evaluation*. Springer, Cham, 2015, pp. 63–86.

[91] A. C. Dirican and M. Göktürk, "Psychophysiological measures of human cognitive states applied in human computer interaction," *Procedia Computer Science*, vol. 3, pp. 1361–1367, 2011.

[92] V. Nagalingam and R. Ibrahim, "User experience of educational games: a review of the elements," *Procedia Computer Science*, vol. 72, pp. 423–433, 2015.

[93] L. E. Nacke, A. Drachen, and S. Göbel, "Methods for evaluating gameplay experience in a serious gaming context," *International Journal of Computer Science in Sport*, vol. 9, no. 2, pp. 1–12, 2010.

[94] L. Anolli, F. Mantovani, L. Confalonieri, A. Ascolese, and L. Peveri, "Emotions in serious games: From experience to assessment," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 5, no. 2010, 2010.

[95] A. Protopsaltis, L. Pannese, D. Pappa, and S. Hetzner, "Serious games and formal and informal learning," *E-Learning Papers*, pp. 1887–1542, 2011.

[96] R. A. Tubelo, F. F. Portella, M. A. Gelain, M. M. C. de Oliveira, A. E. F. de Oliveira, A. Dahmer, and M. E. B. Pinto, "Serious game is an effective learning method for primary health care education of medical students: A randomized controlled trial," *International Journal of Medical Informatics*, vol. 130, p. 103944, 2019.

[97] J. C. Lester, E. Y. Ha, S. Y. Lee, B. W. Mott, J. P. Rowe, and J. L. Sabourin, "Serious games get smart: Intelligent game-based learning environments," *AI Magazine*, vol. 34, no. 4, pp. 31–45, 2013.

[98] J. Moizer, J. Lean, E. Dell'Aquila, P. Walsh, A. A. Keary, D. O'Byrne, A. Di Ferdinando, O. Miglino, R. Friedrich, R. Asperges *et al.*, "An approach to evaluating the user experience of serious games," *Computers & Education*, vol. 136, pp. 141–151, 2019.

[99] I. Donald and K. MacLeod, "Glitchspace: teaching programming through puzzles in cyberspace," in *European Conference on Games Based Learning*. Academic Conferences International Limited, 2017, pp. 148–154.

[100] C. Meftah, A. Retbi, S. Bennani, and M. K. Idrissi, "Evaluation of user experience in the context of mobile serious game," in *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*. IEEE, Taza, Morocco, 2019, pp. 1–5.

[101] R. Ibrahim, R. C. M. Yusoff, H. Mohamed-Omar, and A. Jaafar, "Students perceptions of using educational games to learn introductory programming," *Computer and Information Science*, vol. 4, no. 1, pp. 205–216, 2011.

[102] M. Muratet, P. Torguet, F. Viallet, and J.-P. Jessel, "Experimental feedback on Prog&Play: a serious game for programming practice," in *Computer Graphics Forum*, vol. 30, no. 1. Wiley Online Library, 2011, pp. 61–73.

[103] "Programming Hero," Accessed on: Dec. 20, 2020. [Online]. Available: https://play.google.com/store/apps/details?id=com.learnprogramming.codecamp.

[104] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.

[105] L. E. Nacke, “Wiimote vs. controller: electroencephalographic measurement of affective gameplay interaction,” in *Proceedings of the international academic conference on the future of game design and technology*, 2010, pp. 159–166.

[106] Q. Yu, X. Che, S. Ma, S. Pan, Y. Yang, W. Xing, and X. Wang, “A hybrid user experience evaluation method for mobile games,” *IEEE Access*, vol. 6, pp. 49 067–49 079, 2018.

[107] L. E. Nacke, M. N. Grimshaw, and C. A. Lindley, “More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game,” *Interacting with computers*, vol. 22, no. 5, pp. 336–343, 2010.

[108] J. Cusveller, C. Gerritsen, and J. de Man, “Evoking and measuring arousal in game settings,” in *International Conference on Serious Games*. Springer, Cham, 2014, pp. 165–174.

[109] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, “Using psychophysiological techniques to measure user experience with entertainment technologies,” *Behaviour & information technology*, vol. 25, no. 2, pp. 141–158, 2006.

[110] R. B. Knapp, J. Kim, and E. André, “Physiological signals and their use in augmenting emotion recognition for human–machine interaction,” in *Emotion-oriented systems*. Springer, Berlin, Heidelberg, 2011, pp. 133–159.

[111] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, “A review of emotion recognition using physiological signals,” *Sensors*, vol. 18, no. 7, p. 2074, 2018.

[112] S. Wioleta, “Using physiological signals for emotion recognition,” in *2013 6th International Conference on Human System Interactions (HSI)*. IEEE, Sopot, Poland, 2013, pp. 556–561.

[113] “Neurosky Mindwave Mobile 2 EEG Headset,” Accessed on: Dec. 20, 2020. [Online]. Available: https://store.neurosky.com/pages/mindwave.

[114] "Mi Smart Band 4," Accessed on: Dec. 20, 2020. [Online]. Available: https://www.mi.com/global/mi-smart-band-4/.

[115] "eegID," Accessedon: Dec. 20, 2020. [Online]. Available: https://play.google.com/store/apps/details?id=com.isomerprogramming.application.eegIDhl=en.

[116] T. Dalgleish and M. Power, Eds., *Handbook of cognition and emotion*. John Wiley & Sons, UK, 2000.

[117] "MiFit," Accessed on: Dec. 20, 2020. [Online]. Available: https://play.google.com/store/apps/details?id=com.xiaomi.hm.health&hl=en&gl=US.

[118] "ThinkGear measurements (MindSet Pro/TGEM)," Accessed on: Dec. 20, 2020. [Online]. Available: http://support.neurosky.com/kb/science/thinkgear-measurements-mindset-protgem.

[119] "Understanding Your Target Heart Rate," Accessedon: Dec. 20, 2020. [Online]. Available: https://www.hopkinsmedicine.org/health/wellness-and-prevention/understanding-your-target-heart-rate.

[120] "EEG band power values: Units, amplitudes, and meaning," Accessed on: Dec. 20, 2020. [Online]. Available: http://support.neurosky.com/kb/development-2/eeg-band-power-values-units-amplitudes-and-meaning.

[121] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, no. 4700, pp. 750–752, 1985.

[122] L. Shu, Y. Yu, W. Chen, H. Hua, Q. Li, J. Jin, and X. Xu, "Wearable emotion recognition using heart rate data from a smart bracelet," *Sensors*, vol. 20, no. 3, p. 718, 2020.

[123] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2017.

[124] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.

[125] "Xgboost documentation," Accessed on: Dec. 20, 2020. [Online]. Available: https://xgboost.readthedocs.io/en/latest/index.html.

[126] M. H. Diskin, "Definition and uses of the linear regression model," *Water Resources Research*, vol. 6, no. 6, pp. 1668–1673, 1970.

[127] A. Protopsaltis, L. Pannese, D. Pappa, and S. Hetzner, "Serious games and formal and informal learning," *E-Learning Papers*, pp. 1887–1542, 2011.

[128] "How does stress impact heart," Accessed on: Dec. 20, 2020. [Online]. Available: https://share.upmc.com/2014/11/how-does-stress-impact-heart/.

[129] U. Hasan, M. F. Islam, M. N. Islam, S. B. Zaman, S. T. Anuva, F. I. Emu, and T. Zaki, "Towards developing an iot based gaming application for improving cognitive skills of autistic kids," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2020, pp. 411–423.

[130] "EMOTIV EPOC X," Accessed on: Dec. 20, 2020. [Online]. Available: https://www.emotiv.com/epoc-x/.

[131] "ProComp Infiniti System w/ BioGraph Infiniti Software - T7500M," Accessed on: Dec. 20, 2020. [Online]. Available: https://thoughttechnology.com/procomp-infiniti-system-w-biograph-infiniti-software-t7500m/#.

[132] R. D. Ward and P. H. Marsden, "Physiological responses to different web page designs," *International Journal of Human-Computer Studies*, vol. 59(1-2), pp. 199–212, 2003.

[133] S. Hong and J. Tanvi, "Effects of web atmospheric cues on users' emotional responses in e-commerce," *AIS Transactions on Human-Computer Interaction*, vol. 4(1), pp. 1–24, 2012.

[134] J. U. Garbas, T. Ruf, M. Unfried, and A. Dieckmann, "Towards robust real-time valence recognition from facial expressions for market research applications," in *Proceedings of the 5th Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2-5 September 2013*. IEEE Explore, 2013, pp. 570–575.

[135] K. Hercegfi, M. Pászti, S. Tóvölgyi, and L. Izsó, "Usability evaluation by monitoring physiological and other data simultaneously with a time-resolution of only a few seconds," in *Proceedings of the 13th International Conference on Human-Computer Interaction, San Diego, California, USA, 19-24 July 2009*. Springer, 2009.

[136] A. N. Alfimtsev, M. A. Basarab, V. V. Devyatkov, and Levanov, "A new methodology of usability testing on the base of the analysis of user's electroencephalogram," *Journal of Computer Sciences and Applications*, vol. 3(5), pp. 105–111, 2015.

[137] Q. X. Qu, L. Zhang, W. Y. Chao, and V. Duffy, "User experience design based on eye-tracking technology: A case study on smartphone apps," in *Advances in Applied Digital Human Modeling and Simulation of the AHFE 2016 International Conference on Digital Human Modeling and Simulation*, 1st ed., G. D. Vincent, Ed. Cham: Springer, 2017, vol. 481, pp. 303–315.

[138] E. De Kock, J. Van Biljon, and M. Pretorius, "Usability evaluation methods: Mind the gaps," in *Proceedings of the 2009 Annual Conference of the South African Institute of Computer Scientists and Information Technologists, Vanderbijlpark, Emfuleni, South Africa, 12-14 October 2009*. ACM, 2009.

[139] J. Frey, C. Mühl, F. Lotte, and M. Hachet, "Review of the use of electroencephalography as an evaluation method for human-computer interaction," in *Proceedings of the International Conference on Physiological Computing Systems, Lisbon, Portugal, 7-9 January 2014*. Springer, 2014.

[140] C. Morin, "Neuromarketing: The new science of consumer behavior," *Society*, vol. 48(2), pp. 131–135, 2011.

[141] M. D. Bercea, “Anatomy of methodologies for measuring consumer behavior in neuromarketing research,” in *Proceedings of the LCBR European Marketing Conference: 9-10 August 2012; Munich, Germany*. LCBR, 2012, pp. 1–14.

[142] C. D. Ahmet and G. Mehmet, “Psychophysiological measures of human cognitive states applied in human computer interaction,” in *Proceedings of the 2nd World Conference on Information Technology, Antalya, Turkey, 22-27 November 2011*. Elsevier, 2011, pp. 1361–1367.

[143] A. Mileti, G. Guido, and M. I. Prete, “Nanomarketing: A new frontier for neuromarketing,” *Psychology & Marketing*, vol. 33(8), pp. 664–674, 2016.

[144] V. A. Roth, “The potential of neuromarketing as a marketing tool.” 2014, bachelor’s Thesis, University of Twente.

APPENDIX A DATA SYNTHESIS AND ANALYSIS GRAPHS OF SLR

Fig. A1. Word cloud for the keywords of the selected articles

Fig. A2. Word cloud for the titles of the selected articles

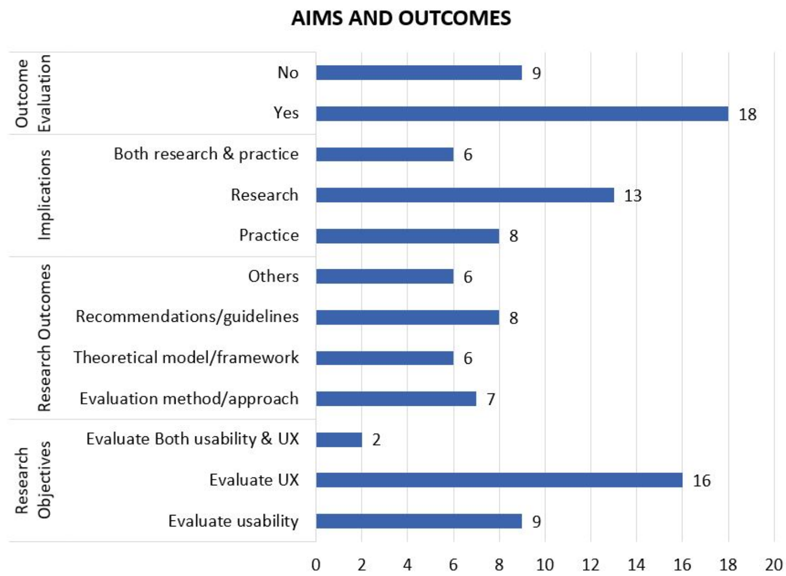


Fig. A3. Aims and outcomes of the reviewed articles

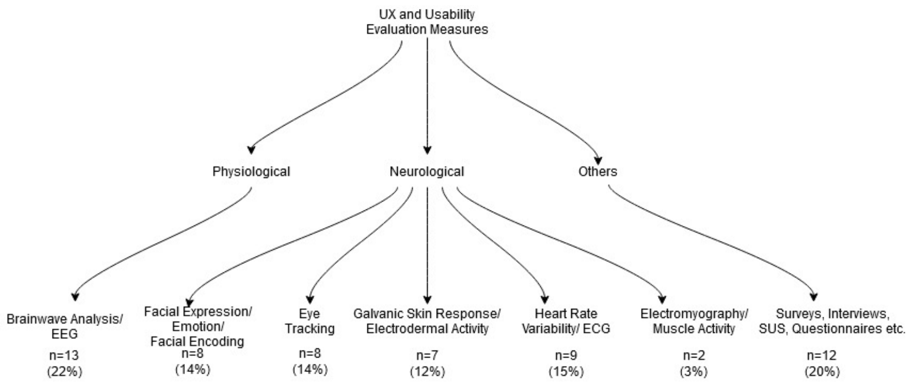


Fig. A4. UX and usability evaluation measures with respect to number of articles

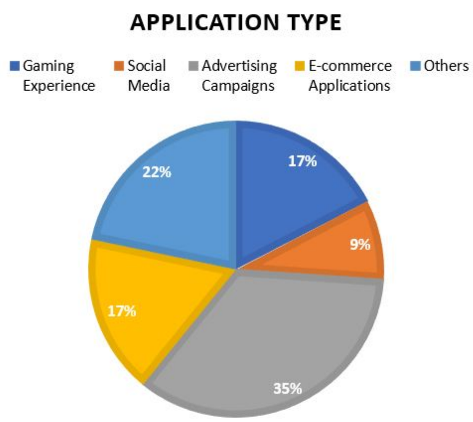


Fig. A5. Application type with respect to number of articles

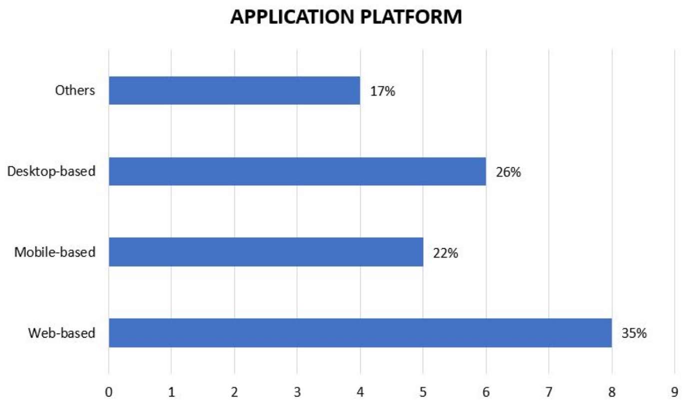


Fig. A6. Application platform with respect to number of articles

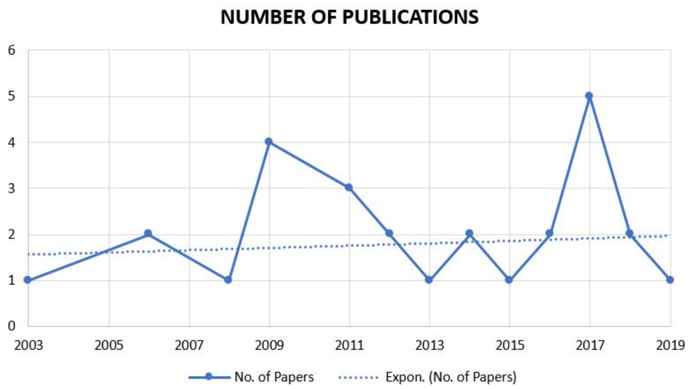


Fig. A7. Publication year with respect to number of articles

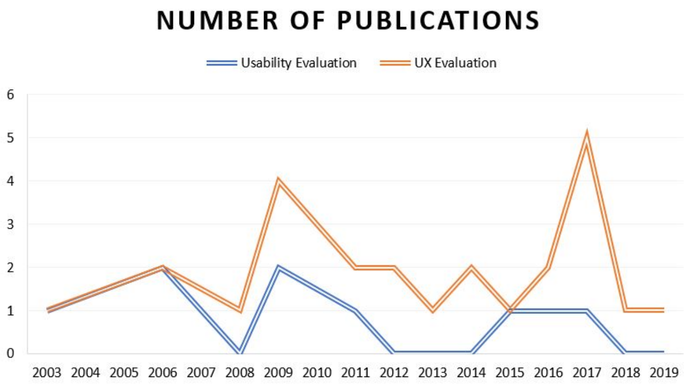


Fig. A8. Year wise publications based on research objectives

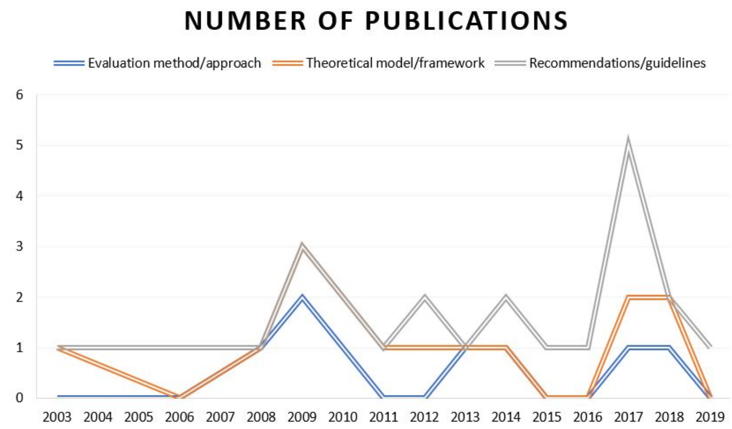


Fig. A9. Year wise publications based on research outcomes

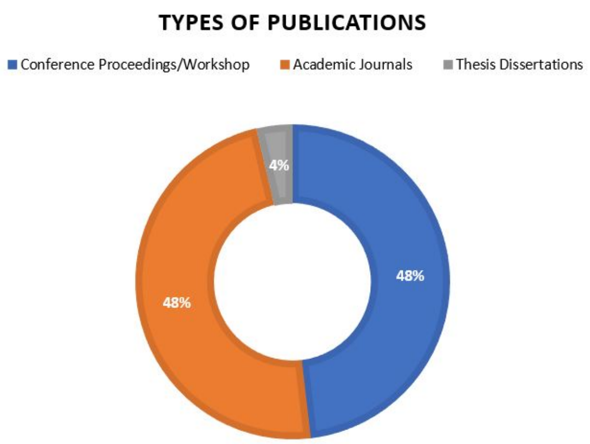


Fig. A10. Publication type with respect to number of articles

APPENDIX B DATA EXTRACTION TABLE OF SLR

Table B1: Summary of extracted data related to the aims and outcomes

Ref.	Research Objectives	Category of Objectives			Research Outcomes	Category of Outcomes				Implications			Outcome Evaluation	
		Evaluate usability	Evaluate UX	Evaluate both usability & UX		Evaluation method/ approach	Theoretical model/ framework	Recommen-dations /guidelines	Others	Practice	Research	Both research & practice	Yes	No
[12]	To evaluate UX of digital games and social media combining neurological, physiological and traditional measures.		✓		A triangulation multimodal approach that would aid designers and developers to improve UX of interactive media products.	✓				✓			✓	
[132]	To evaluate usability of web application using short and long term psychophysio-logical measures.	✓			A proposed model that shows psychological measures have notable effects on usability evaluation of software applications.		✓				✓		✓	
[133]	To evaluate UX of an online e-commerce website using physiological responses and subjective measures.		✓		A framework for evaluating UX where the results of physiological responses show that users' purchase intention is heavily affected by such measures.		✓				✓		✓	

Table B2: Types of data considered in different types of experimental evaluation.

Evaluation Measures	Types of Evaluation Data	Means of Data Collection	Purpose of Evaluation	Ref.
Neurological	Brainwave analysis/EEG	Emotiv EPOC+ wireless device	UX of E-commerce applications	[67]
Physiological	Heart rate variability/ECG, galvanic skin response/electrodermal activity	DataLab 2000 system	Usability of an HTML based web application	[132]
	Heart rate variability/ECG, galvanic skin response/electrodermal activity, electromyography/muscle activity	ProComp Infiniti System	UX of entertainment and gaming context	[69]
	Heart rate variability/ECG, galvanic skin response/electrodermal activity, emotion	BITalino system, couple of electrodes and a band pass filter	UX of advertising context	[8]
	Facial expression/emotion	Webcam	UX of TV commercials	[134]
	Facial encoding/emotion	Affdex SDK	Usability of gaming environment	[7]
	Eye tracking	Eye tracker	Usability of an advertising webpage	[6]
	Heart rate variability/ECG	LED lights, a digital video camera (Canon VIXIA HF R62)	UX and usability of HCI-based applications.	[68]
	Heart rate variability/ECG, galvanic skin response/electrodermal activity	ISAX system	Usability of a call center operating system	[135]
Neurological and Physiological	Brainwave analysis/EEG, Heart rate variability/ECG, galvanic skin response/electrodermal activity, emotion	Emotiv Epoc 14-channel wireless EEG machine, iMotions technology	UX of commercial advertising application	[70]
	Brainwave analysis/EEG, emotion	32-channel EEG machine	UX of a musical stimulus	[85]
Neurological and Others	Brainwave analysis/EEG and subjective data	NeuroSky MindWave device and survey questionnaire	Usability of a web and mobile app of an advertising context	[136]
Physiological and Others	Heart rate variability/ECG, emotion and subjective data	Quantemo's Mindkit platform and survey, interview	UX of digital games and social media	[12]
	Heart rate variability/ECG, galvanic skin response/ electrodermal activity and subjective data	Eye tracker, biofeedback sensors and online survey	UX of an E-commerce web application	[133]
	Eye tracking and subjective data	Tobii TX300 Eye Tracker and questionnaire	UX of instant messaging apps	[137]
	Heart rate variability/ECG, eye tracking and subjective data	ProComp Infiniti System, NAC EMR-HM8 headed mounted eye tracker and questionnaire	Usability of gaming context	[71]
	Eye tracking and subjective data	Tobii 1750 eye tracker and multiple heuristics method	Usability of an web application	[138]
	Eye tracking, facial encoding, subjective and objective data	Biometric sensors, web questionnaire and performance metrics	Usability of a commercial biometric identity verification system	[75]

Table B3: A mapping between evaluation measures with type and platform of applications

Evaluation Measures	Application Type	Application Platform	Evaluation Objective	Ref.
Neurological	E-commerce, Advertising campaigns, HCI-based applications	Desktop-based	UX	[67], [139], [140], [141]
Physiological	A digitized web directory, Gaming experience, Call center operator, Commercial biometric identity verification system	Desktop-based, Web-based, Client-server distribution system, Workstation	UX and Usability	[132], [69], [8], [134], [7], [6], [68], [75], [135]
Neurological and Physiological	Instrumental musical stimulus, Advertising campaign	Desktop-based, Multimedia, Web-based	UX and Usability	[70], [85], [78], [77], [142], [143], [144]
Neurological, Physiological and Others	Social media, E-commerce, Advertising campaign, In-vehicle information system	Desktop-based, Web-based, Mobile-based, Embedded system	UX and Usability	[136], [12], [133], [14], [15], [138], [73]

Table B4: Research Type with respect to number of papers

Research Type	No. of Articles	Percentage	Ref.
Experimental Study	18	66%	[12], [132], [133], [85], [69], [8], [70], [134], [7], [137], [6], [136], [68], [67], [71], [138], [75], [135]
Conceptual Study	4	15%	[139], [78], [140], [73]
Literature Review	4	15%	[142], [141], [143], [144]
Case Study	1	4%	[77]

APPENDIX C SAMPLE DATASET OF THE EXPERIMENTAL STUDY

Table C1: Sample dataset

beat per 15 sec	alphaLow	alphaHigh	betaLow	betaHigh	emotional state
18	29261	7274	16757043	88263	deactivated
18	100438	129658	100862	139764	activated
18	16760423	68667	24708	29698	activated
18	32103	13834	21764	10605	activated
18	15432	16760554	16774031	16770011	activated
18	16754732	16750650	16772870	16431	deactivated
18	10823	7474	3748	8448	activated
18	47808	27010	188865	108259	deactivated
18	186117	147886	138234	41797	activated
18	36440	40654	70870	16763791	deactivated
18	43774	182473	183974	211901	activated
18	298118	282607	132005	84459	deactivated
21	22911	7912	30607	7978	deactivated
21	223374	403035	96143	105460	activated
21	16761277	31103	19910	16759691	activated
21	16750554	16765826	16748972	16752940	activated
22	16772715	26228	24295	28589	activated
22	16771913	20304	4664	23992	activated
22	131427	154380	38573	16745883	deactivated
22	7413	4947	3452	24275	activated
19	40297	140860	223613	59361	deactivated
19	13024	43801	16745483	20741	deactivated
19	5953	13557	4364	8683	deactivated
19	11404	26451	16761888	24640	deactivated
20	96781	25671	77984	38098	activated
20	85186	30001	16751491	16767131	activated
20	16754115	29131	18656	16749955	activated
20	56213	91639	16774959	16757725	activated
20	16746096	68929	18637	23188	activated
20	9737	2868	2314	4084	activated
20	8450	10838	4042	4684	activated
20	6405	16749756	16748987	89173	deactivated
20	28770	16749808	28902	26546	activated
20	49465	16757498	76997	16762880	activated
20	16767045	12620	24447	25041	activated
20	21382	16751445	16760116	16756598	activated