DESIGN AND DEVELOPMENT OF NOVEL EMOTION RECOGNITION FRAMEWORK FOR VIRTUAL LEARNING ASSESSMENT USING INFORMATION FUSION

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

In

Computer Applications

By

Fayaz Ahmad Fayaz
Reg. No. 41800688

Supervised by
DR. ARUN MALIK (17442)
Professor Department of Computer Science & Engineering
Lovely Professional University (LPU) Jalandhar, Punjab, India

Transforming Education Transforming India

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB
February-2024

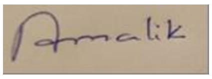
DECLARATION

I, hereby declare that the presented work in the thesis entitled "Design and Development of Novel Emotion Recognition Framework for Virtual Learning Assessment using Information Fusion", in fulfilment of degree of **Doctor of Philosophy (Ph.D)** is outcome of research work carried out by me under the supervision of Dr. Arun Malik, working as Professor, in the Department of Computer Sciences and Engineering of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of another investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

Fayaz Ahmad Fayaz
Reg. No. 41800688
Department of Computer Applications
Lovely Professional University,
Punjab, India

<u>CERTIFICATE</u>

This is to certify that the work reported in the Ph. D. thesis entitled "Design and Development of Novel Emotion Recognition Framework for Virtual Learning Assessment using Information Fusion" submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the Department of Computer Applications, is a research work carried out by Fayaz Ahmad Fayaz, Registration No. 41800688, is bonafide record of his original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.



Dr. Arun Malik
Professor Department of
Computer Science & Engineering
Lovely Professional University Jalandhar, Panjab, India

ABSTRACT

A **Hot-bed** for research in the field of artificial intelligence is multisensory emotion identification, which employs computer vision to cash inflows features based on sensed data inputs. The conventional approach to inter emotion detection merely connects the various forms; however, this approach has a poor interaction usage rate of paradigm knowledge and is unable to accurately depict the true emotions amid paradigm features dispute. With the development of ICT, the idea of virtual learning and its implementations, such as the online education, that is a shared online environment where students and teachers may collaborate at the same time. Covid19 has educated us how to get taught while sitting at any place be at home or in car or garden etc. Students got benefit while attending lectures given by delegates/ teachers that are distributed across regional borders. In this type of engagement, the instructor plays a crucial role as a facilitator, overseeing the student learning and encouraging group activities, such as discussions, exams, and other instructional activities. To use a feedback mechanism, the suggested Hybrid Human Emotion Recognition (HHER) system can help make the system adaptable and error-proof. The active learning benefits greatly from the presence of emotions, for inter - connected learning systems or the virtual classroom mode. According to research, the curriculum should arouse and guide the learner's emotions toward the right condition. However, the system should initially be able to identify the learner's emotions. A s in HER situation Voice recognition (VR), facial expression detection (FED), audio data, and gesture recognition (GR) will be able to identify the feelings by asking the person and mapping underlying attributes. Unlike existing methods, the Focusing on the assessment of human activities, the Multi Factor Hybrid Emotion Detection System blends face, mobility, speech, text, self-reporting features, and wearable electronics. Additionally, it should be mentioned that a learner's ability to identify emotions depends on their generation, demographics, geography, and society. To ensure that the effort of human specialists may subsequently be focused on generating appropriate robust understanding and building genuine datasets for model training, more systems must be influenced by the integrity and dependability of the instructional score. As a

result, the suggested gadget should have several applications in numerous industries. By utilising the proliferation of cordless breakthrough, the use of smart internet powered by artificial intelligence and the expansion of human-machine interactions (HMI), dependence on technical tools is increasing daily. Intelligent Tutoring System (ITS) currently offers a virtual learning system, also known as an e-learning system, in place of the traditional teaching-learning model. Additionally, according to research, pupils who receive individualized instruction comprehend concepts more thoroughly than those who receive training in a traditional classroom setting. Recognizing the emotions of an online student while putting technical resources to use to generate feedback to improve the virtual teaching-learning process is a difficult task for an Artificial Intelligence (AI) and information fusion system. Accordingly, through human-computer interaction (HCI), we anticipate that machines will be extremely interactive and capable of comprehending and perceiving diverse emotional states. In relation to the key elements of a smart tutoring system, the learner's modal consists of motivational, emotive, and cognitive states with significant implications for performance-oriented education. The interplay of human emotions with HCI is a movement. In that specifically, emotion detection enables machines to recognize and comprehend human emotions. There have been various alternative approaches for developing e-learning emotion recognition systems, but so far none have proved effective. This research compares various e-learning emotion identification techniques, including gesture and facial expression recognition. Both its advantages and disadvantages are present. With the sophisticated tools available to strengthen emotion recognition in virtual electronic environments, the on-going effort can be maintained. This dissertation thoroughly examines and condenses the crucial technologies in the area of multi-modal sensor amalgamation for data-driven sentiment detection, using the real-time Human emotion recognition methods as an instance. The open dataset's pre-existing extraction technologies are taken into account, as well as audio, figurative, written, and electrophysiological characteristics, featured fusing layers, decision layer fusion, and categorization. These conversations are intended to give a thorough large-scale depiction of this fascinating and popular scientific discipline.

DEDICATION

Never forget two people in your life

The person who last everything just to make you win

(Your Father your superhero)

&

The person who was with you in every pain

(Your loving Mother)

This thesis is dedicated to my loving parents may Allah be pleased with them, my family (wife and two sweet daughters), my back-in love two sisters and above all my elder brother (Dr. Peerzada Mohammad Tariq) who held my hand when I was raw and showed me the real path of success. His contribution in my life is worth which has no parallels and I owe him a lot.

-Love you Bro.-

ACKNOWLEDGEMENTS

To begin with, by the name of Almighty Allah, who enabled me to use wisdom and knowledge for this research work. Moreover, the patience Allah has bestowed me in all tough times and finally made it to achieve the goal.

I am beyond grateful to my esteemed supervisor Dr. Arun Malik Professor Department of Computer Sciences & Engineering LPU for his endless support, guidance and supervision at each stage of my research. I feel humbled and fortunate enough to have the opportunity to work with such a kind and knowledgeable person.

I am highly thankful to my family (Mrs. Nusrat Gazala my wife) two daughters (Ayisha & Amira), my brothers & my two sisters Mrs. Fiza Jan & Mrs. Rubby Jan, for their endless prayers. High regards and tons of thanks to my fellow friends Prof. Arshad Ahmad Yatoo & Prof. Aadil Hussain for their motivation & timely valid suggestions.

I am pleased to thank Prof. Dr. Syed Immumal Ansarullah for helping me to publish my work in journals of repute and also would like to keep on record my sincere thanks to entire family of Lovely Professional University (LPU) especially the Division of Innovation & Entrepreneurship, for all their valuable support. Further I extend my sincere thanks to my Father-in-law who supported me while handling kids during this whole process. With all prays I wish NIELIT J&K, Govt. of India to prosper by leaps and bounds where I utilized hassle-free all resources in my personal development to complete this research work. Last but not least I would say thanks to Allah for blessing me my Elder Brother Dr. Peerzada Mohammad Tariq whose contribution in my life is priceless. After Allah credit goes to him for bestowing and nurturing me to touch the skies.

Date: February 17th 2024 **Fayaz Ahmad Fayaz**

CONTENTS

LIST OF FIGURES

CHAPTER-1

1. INTRODUCTION

Covid-19 has educated us how to learn from home. Students got benefit while attending lectures given by delegates/ teachers that are distributed across regional borders. In this type of engagement, the instructor plays a crucial role as a facilitator, overseeing the student learning and encouraging group activities, such as discussions, exams, and other instructional activities. There is a function for the projected Hybrid Human Emotion Tracking (HET) solution in making the system flexible and error-proof with a feedback policy (Pirouz et al., 2020). As wireless Internet and cognitive computing technology are in success, more and more communication is now conducted between humans and machines. Additionally, there is a rising requirement for an AI based devices that can identify a user's emotions and provide responses in line with those emotions. People now have higher expectations for human-computer contact because they believe interactive robots should be able to see, interpret, and express a wide range of emotions (Jiang et al., 2020). Moreover, the current human-to-human interaction strategy of many robotic devices is mechanistic and monotonous, focusing mainly on the backstory research and keyword searching, which is insufficiently smart and requires awareness of the contextual framework (M. Chen et al., 2020)(Sarker, 2022). As a result, in order to achieve emotional engagement, we must also include deliberate and emotional elements and apply affective computing technologies. Inside the sophisticated age of information, emotional mapping has taken the lead in human-computer interaction. Additionally, emotive engagement influences the intelligence of human-computer interactions. Specifically, it creates a deep human-machine interaction and awareness by making human-machine interface as a genetic, friendly, vibrant, emotive, and temperature oriented as can be seen in human-human interaction is(Fayaz et al., 2019). The emotional contact between people and computers is also significantly influenced by emotion classification. Machines may now sense human emotional states and develop empathy thanks to emotion recognition technology(Katona, 2021). Emotions have an essential role to play in the teaching and understanding what has been taught procedure. For distance learning mode or virtual electronic learning (e-learning) systems, considering the learner 's emotions is vital(D'Mello & Kory, 2015). Research has suggested that the program should stimulate and lead the emotions of the learner to the correct state. But, at first, the system should first recognize the learner's emotions. Like HER context. By knowing from the individuals, mapping concealed characteristics, facial emotion identification (FEI), voice detection (VD) and gestural identification (GI) will perceive the emotions. Focusing on the assessment of human activities, the Multi Factor Hybrid Emotion Detection System blends face mobility, speech, text, self-reporting features, and wearable electronics. It should also be noted that the identification of emotions by the learner is geared towards the age of the learner, demographic characteristics, geography, and society. More system must be informed by the integrity and reliability of the training mark to ensure that the work of human experts can then be focused on developing acceptable deep learning and building real databases for model skilling(Gomez-Donoso et al., 2019)(Soundararajan & Biswas, 2019). Therefore, the device proposed would have broad implementations in diversified fields(Soundararajan & Biswas, 2019). The term "emotion" refers to the neural system of humans' subjective attitude toward other people. The mind initially transmits commands for the accompanying response, which affects a person's facial gestures, voice frequency, speed, plus body language, as well as the heart, arms, legs, mind, and other human organs. In attempting to account for this synergy between diverse multimodal affective inputs, researchers have started to integrate nonverbal cues, blink, movements, and other psychophysical indications in the emotion detection research. For instance, to identify human moods, the authors combine the three physiological signals EDA, PPG, and EMG (Mehedi et al., 2019). Multifunctional sensor fusion for data-driven sentiments recognition has attracted the attention of researchers in the field of emotion recognition. When contrasted with single-mode emotion detection, holistic information fusion for data-driven emotion detection exhibits high accuracy. Bigun and Duc originally put up the idea of fusing and interpreting multimodal emotion data in 1997. They combined the speech and visual data and proposed an analytical strategy based on the Bayesian theory(Beno Duc, 1997). AI and multi-sensor data fusion techniques have been advancing rapidly over the past decade. As a response, significant progress has been made in the study of the fusion and recognition of multimodal emotion information(Wang et al., 2017). The potential applications for multimodal emotion identification are numerous and diverse. Additionally, it aids in offering some beneficial services to youngsters and senior adults who are empty nesters.

By documenting human emotions, psychologists can relieve their workload and provide psychological support for old and young people who are raising empty nesters. Through conversation, a machine with advanced artificial intelligence takes into account the patient's emotion and aids in the treatment of the disease(Fortino, Galzarano, et al., 2014). One must give a thorough and methodical research in order to assist those who are interested in emotion classification in truly realizing multi - modal emotion recognition (MMER). Although there were a few review publications on multimodal emotion recognition, for instance, the survey paper(D'Mello & Kory, 2015) examined the key trends and system-level variables associated with its impacts. Numerous multi-modal emotion identification findings have been obtained during the past few years as a result of significant advancements in deep learning and other Artificial Intelligent technologies. The multimodal sentiment classification of AI technology is not entirely covered in the aforementioned studies (Yin et al., 2017). Additionally, we believe that the autonomic nervous system and endocrine system are in control of the physiological change and are largely unaffected by subjective notions, making sentiment detection based on the physiological signal accurate. Accordingly, the unique swings in human feelings can be seen and the accompanying emotional shift can be recognized with the change in the human being's subtle physiological condition (such as EEG and electrodermal activities). For instance, the pathetic nerve will trigger the appropriate somatic responses, such as heartbeat accelerating, blood pressure changes, breath acceleration, body temperature rise, and even muscle or skin trembling, when people get anxious under pressure or thrilled due to a wicked motivation (Khezri et al., 2015). The assessment of emotions based on

EEG data or other physiological data has more credibility than emotion recognition based on facial recognition and movement since it is natural and cannot be faked or altered artificially(Ringeval et al., 2015). In addition to attempting to contribute to a clear knowledge of the theoretical challenges and future possibilities of this research in the field, this initiative aims to provide a comprehensive description of the data-driven multimodal emotion data fusion.

1.1 Inspiring Information fusion example

The genuine emotion recognition surveillance system employed in this work is demonstrated in Figure 1.1. The mentioned tasks are carried out in this system by gathering multimodal emotion signals, labelling and choosing unlabelled emotion datasets on edge clouds, recognizing and analysing emotional data fusion using AI algorithms on distant clouds, as well as controlling the intellectual sentimental communication machine's emotional responses for decision-making. Participants may be provided with a meaningful, individualized emotional framework.

1.2 Information gathering module for multidimensional emotions

It is challenging to recognize and engage with emotions while building an elevated sentiment data collection framework. The multimodal emotion data collection includes user-extracted EEG, speech, expression, and socialization data from their smartphones. According to the results of neurobiology and cognition science, the

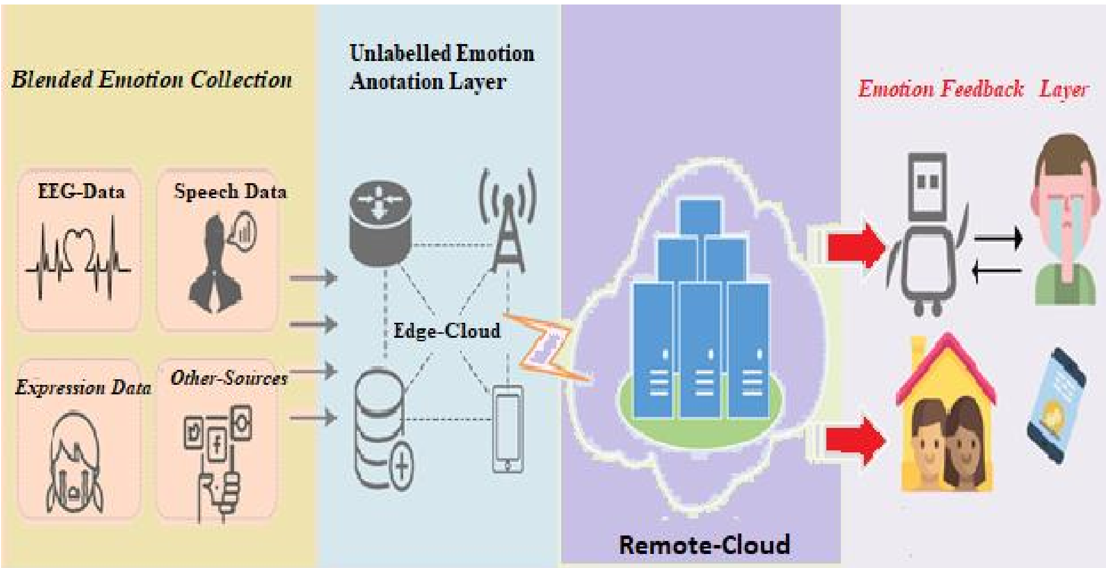


FIGURE 1.1 REAL-TIME HUMAN EMOTION FRAMEWORK

physiological activities of the cerebrum and the formation of emotions are highly correlated. This provides a theoretical framework for identifying a user's emotions through investigation of his cortical activity (the EEG data). Wearable's, which are continually evolving, is used to capture EEG data. The above work makes use of the separately developed 22-channel cognition apparatus.

The EEG signal is captured using a TI firm ADS1299-8 device. Moreover, the master chip is a CH559L. The user's EEG data can be taken for necessary collected in actual time with a cognition device that is pleasant, accessible(Yang & Hu, 2018), and unobtrusive. In an effective emotional interaction robot, a MIC module and camera module are integrated. Users' voice and facial expression data are gathered using them, respectively, in real - time basis on the site. In significant part, cognitive activities can be inferred from a user's speech, inflection, and facial expression. The Dataset is quite helpful for users who are skilled at hiding their emotions because it cannot be altered or hidden. As a result of the advancement of mobile networks with interactive interfaces, many users now frequently post their everyday thoughts and feelings on social media platforms(Muhammad & Alhamid, 2017). Users' textual data that they post on social networks displays changes in their emotions over time. In order to improve expression detection accuracy, the aforementioned four categories of sentiment data are examined and modeled in this study.

1.3 Emotional data categorization layers without labels

The edge cloud's untrained training algorithm can transmit significant data, remove redundant information, and evaluate the reliability of multimodal emotion information. The distant cloud must accurately and quickly discern the user's emotions. A machine that is smart and capable of emotional engagement must also decide as soon as desirable and give the user the appropriate emotional responses(Fortino et al., 2013). The technique gathers a lot of user multimodal emotion data, improving interface impede and feeling comprehension while keeping the system's intelligence in a difficult task(X. Chen et al., 2019). As a result, the work uses the unstructured learning algorithm that was proposed in(M. Chen, Hao, et al., 2018). By doing this, the volume of data posted is reduced while maintaining the remote cloud's functionality.

1.4 Layer of emotive stimuli

The method may precisely determine the person's feelings based on the fusing outcome of multimodal emotional data that got collected from the cloud server. The smart robots may console the user by offering the appropriate emotion treatment through emotional engagement after recognizing the user's regretful and melancholy emotions.

For illustration, the machine might play some music, offer the user some consoling words, give him an embrace, and make them feel the robot's sympathy(Fortino, Parisi, et al., 2014). After recognizing an emotion, the platform can also communicate information about the user to colleagues' and family's cell devices. Among people who are depressed, the symptoms can be identified as soon as feasible, and treatment recommendations should be made. Based on each user's personality and emotional experience, the genuine emotional monitoring system provides personalized, smart, and humanized emotional response.

1.5 Testing datasets

The majority of procedures used to gather sensor information for emotion intelligence include inducing distinct emotions in test subjects through the use of films or other media. The related data is identified and recorded when a desired feeling is produced. The collected information is kept in a dataset known as the influenced or acted dataset. In certain works, the evaluated clients' spontaneous feelings are captured; these emotions were not prompted by any outside cause. This study briefly introduces few multimodal datasets that may have been recently introduced. The comparability of these databases is depicted in Table 1.1.

TABLE 1.1 MULTI-MODALS HER DATASET CLASSIFICATION

Data set	Type of emotion	Subjects	Annotator	Class
RECOLA Dataset	Valence & arousal	44	8	ECG, EDA Audio, Visual.
AFEW Dataset	Surprise, fear, sadness, anger etc.	338	3	Audio visual
EMOEEG	Valence & arousal	10	---	EEG, ECG, EDA etc.
CMU-MOSEI	Surprise, fear, sadness, anger etc.	980	Approx. 3	Audio, Visual & Text
WESAD	Stress, neutral etc.	19	Not decided	ACC, EDA, ECG, EMG etc.
BAUM-1	Surprise, fear, sadness, anger etc.	34	Approx. 6	Audio & Visual

1.6 Identification of features

The same feeling can be expressed in many ways by various individuals. Several people prefer to vocally communicate their sentiments, thus their auditory information includes fewer emotive clues. Other people, conversely, prefer using body language to convey certain emotions. Furthermore, despite the fact that some people are adept at concealing and covering their sentiments, physiological factors cannot be deceived or kept under wraps.(M. Chen, Herrera, et al., 2018). The EEG characteristics are the primary physiological characteristics examined in my study. People frequently publish updates on their everyday lives and emotional status on social networks due to the significant developments in these platforms refer Figure 1.2. As a result, the merger of multimodal emotion data and intelligence has started to play an increasingly important role in emotion detection. The crucial phase in emotion intelligence is the identification of emotional properties from multiple methods. In recent years, several modalities' feature extraction techniques have been tricked.

2. Human Computer Interaction and its frame works

This research work with an initiative of spontaneous sentiment recognition using multimodal data fusion under Human Computer Interaction (HCI). During research review of sentiment affective computing, facial expression recognition and multimodal fusion strategies, a 3D emotion model with fusion framework has been presented, which is a major research contribution in this thesis. We have proposed and implemented 3D emotion model under the umbrella of affective computing, an active research area of HCI. The information fusion approaches are divided in various categories such as rule based, its characterization and estimate based approaches. The proposed fusion framework has dependence on the multi-resolution evaluation of multimodal signals as shown in Figure 1.3 The facial expression recognition technique is also being implemented by us by considering many databases and under different illumination and noise conditions. We have applied the proposed fusion framework to recognize spontaneous emotion from physiological signal i.e. EEG and peripheral signals on the real multimodal database DEAP. Emotion recognition from audio-visual signal is also implemented with remarkable performance. We can summarize the thesis as follows:

As in the introduction we have discussed the motivation and statement of the problem with major research contributions to this study. In the chapter second detailed the related works on multimodal affective computing in context to HCI perspective. This chapter is divided in five major sections. The first section introduces the fundamentals of emotion and affective computing. Second section deals with the basics of facial expression, modeling and representation of facial expression. Theory of emotions, which includes categorical, appraisal and dimensional approach of emotion are explained in third section. Affective information extraction and processing of the information is discussed in section four and finally, we have reviewed the multimodal information fusion approaches and performance evaluation matrices for emotion recognition. We have proposed multimodal fusion framework in chapter-3, to predict emotions in three dimensional spaces with multi-resolution analysis algorithms such as wavelet and curvelet transforms.

The experimental part of the dissertation started with methodologies. That deals with sentiment detection from facial expression. In this chapter, we have performed multi-resolution analysis with different MRA algorithms, namely wavelet and curvelet transform to predict emotion from facial expression under different illumination and noise conditions. A cooperative study is also done to check the performance of the above two algorithms.

A spontaneous affect recognizer based on audio-visual signals, has been also introduced. For multimodal feature extraction, we have explored various aspects of feature extraction, modeling and classification strategy and found MRA better compared to other similar approaches. For classification, SVM shows significantly improved performance compared to other classifiers like MLP, KNN, Meta multiclass. It is shown that the results obtained using fusions are outperformed compare to the system based on a single signal. Further, a multimodal fusion framework has been implemented for sentiment detection and classification from physiological cues.

The hybrid modal signals for assessing EEG, video & physiological signals. Multimodal features are extracted using multi-resolution algorithms. We have evaluated the proposed sentiment detection model with standard multimodal datasets, DEAP. The experimental results obtained from classification using SVM and MLP for single and multimodal cues are very promising. Further we wish to test these features for real time emotion recognition to create human-machine interface.

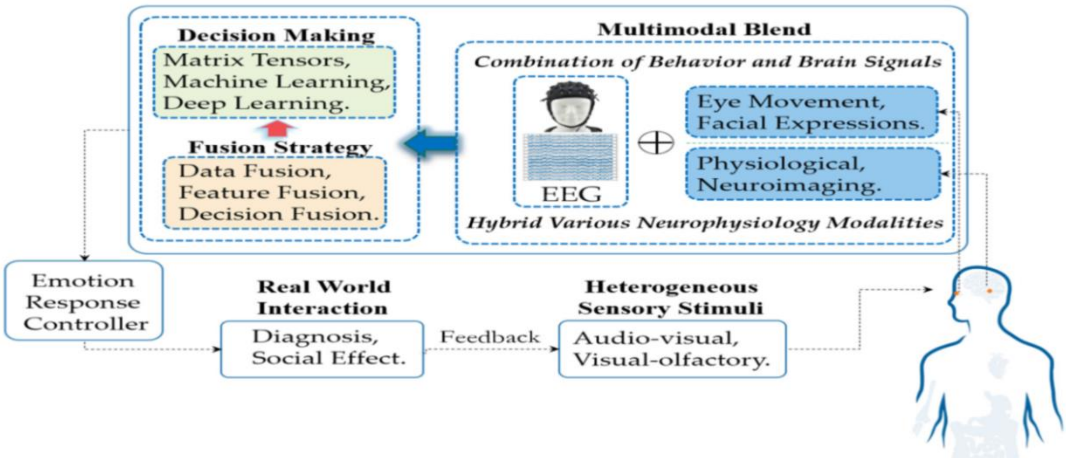


FIGURE 1.2 BASED AFFECTIVE BRAIN–COMPUTER INTERFACES (BCI)

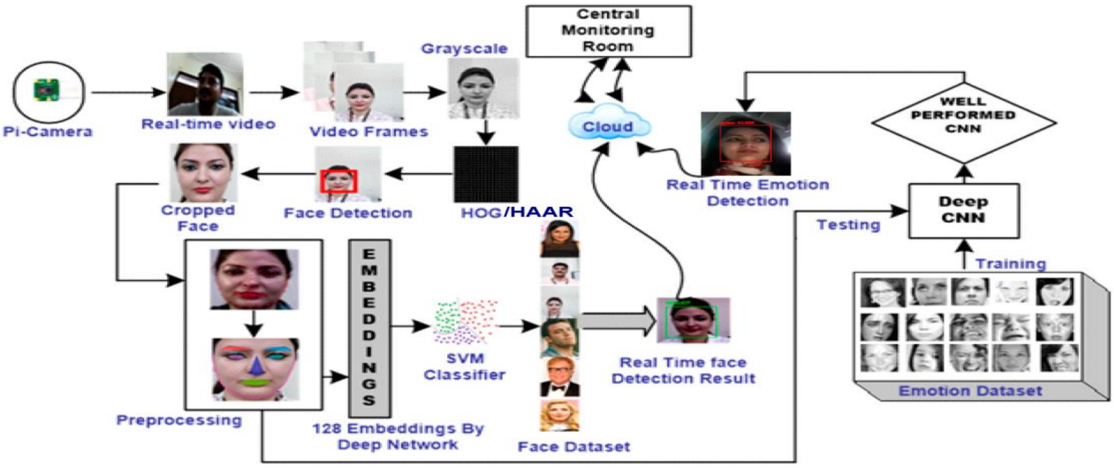


FIGURE 1.3 REAL-TIME FACIAL EMOTION RECOGNITION

3. Novelty in the research

The idea is very unique in a sense that proposed model doesn't only select a set of stake holders like in a class we can have certain students with some craniofacial anomalies, which once captured via camera feed will give different expressions in terms of emotional data. Hence a formulated model i.e., eagerness measuring technique for online teaching learning will be a generalized model for both normal class of stakeholders and the class of people who would be different from normal. So, as proposed a general sensor-based hybrid emotion recognition framework will assist for online teaching assessment to cater the need of era like Covid-19, where education sector has taken a move from offline to online teaching learning process for all sects of people (normal/craniofacial anomalies etc.).

Recognizing data from fusing, real time e-learning emotion surveillance modal to be incorporated to collect emotional signals from multiple sources. Further labeling and selection of unlabeled emotion datasets from sources would be required to prepare feedback for decision making to know the result whether learners have understood the concept or what percentage of improvement is required to overcome the loopholes in teaching via virtual mode of teaching learning process which is the need of hour especially during lockdowns to maintain the standard of education system.

3.1 DESCRIPTION OF THE INVENTION

With the universal adoption of advanced technology in all sections of life and growing market, the demand for advanced technologies that can determine a potential customer's needs and recommend the best solution for them is skyrocketing. Robotics, marketing, education, and the entertainment industry, for example, all benefit from automated emotion assessment.

- ✓ the edge based architecture
- ✓ Hybrid human emotion recognition.
- ✓ Data fusion recognition, real time e-learning emotion surveillance modal to be incorporated.
- ✓ In this model, we need to:
- ✓ Collect emotional signals from multiple sources.
- ✓ Have labeling and selection of un-labeled emotion datasets from various sources on required emotional feedback for decision making control.
- ✓ In education this application will result to know whether Learners have understood the concept or an improvement is required in teaching via virtual mode of teaching learning process.

Varied applications of Novel Human Emotion Recognition are like:

- ✓ Online learning/ e-Learning (Need of the hour with reference to Pandemic COVID19)
- ✓ Medical Science
- ✓ Robotics & Marketing
- ✓ Advertising
- ✓ Virtual Reality
- ✓ E-Education & Gaming
- ✓ Automobile Industry
- ✓ Enterprise level training and implementation
- ✓ Automobile Industry
- ✓ Sweet Home appliances, etc.

Are major stakeholders to use emotion sensor technology to be more user friendly and trustworthy to the public. With its performance, it will be considered almost in every sector of life. As an emotional IOT will make people dependent by its high performance.

The tech tools like computer vision speech recognition, DL of AI and other sister concerned technologies are hand to hand with this developmental process where a Novel Hybrid Emotion recognition framework combines all the result-oriented emotional recognition input sources plus the hardware and software with an accurate real-time emotion recognition solution for multi sects. Figure-1.4 depicts a simple architecture for an online e-learning feedback system.

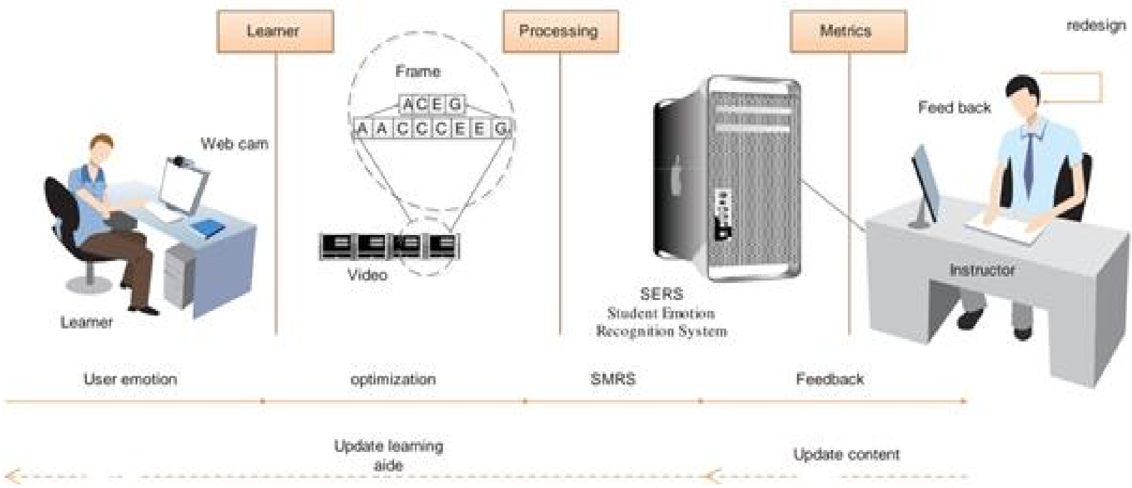


FIGURE 1.4 PROPOSED TENTATIVE ARCHITECTURE OF E-LEARNING SYSTEM

- Information fusion is the technique of integrating a set of data sources to obtain consistent, useful, and accurate information.
- The Internet of Things refers to objects that are individually recognizable and their virtual representations as constructs in the World.
- Deep learning is a powerful AI feature that imitates human brain activity to process data that can be used in speech recognition, language translation, object detection and decision-taking, without human oversight it can learn.

Multiple sources gather huge information about a user's emotions to be passed in a real-time data center for decision-making about a human-computer emotion interaction. So, by having various data sources available a multi-modal approach is to be developed wherein all the sources of information will be consolidated via information fusion. The evidence of development in the internet of things sensors concerning the wearable humans body and brain wearable equipment. Hence, a model wherein multiple emotion recognition methods like voice, text, vital signals, facial expressions and IOT based sensors will become a source of inputs to a Hybrid Emotion Recognition Model.

Smartness by having mobile internet as a communication medium makes strong bonding between humans and machines. The greater role of Artificial Intelligence (AI) regarding deep learning models is in force. Deep learning multi-layer framework can automatically perform high accuracy feature extraction and classification. Making availability of training samples for supervised learning of DL, they are both supervised and unsupervised to be implemented for better results. The IoT & progression in sensor technology, large data sets retrieved from multiple sources connected to emotion extraction from humans.

4. Description of Chapters

This thesis is organized into five chapters whose brief introduction is as follows:

Chapter 1

This chapter introduces the concept of information fusion, emphasizing the aggregation of emotional data from diverse sources. It explores the characterization and testing of datasets within the realm of HCI. The inclusion of novelty in the research is highlighted, accompanied by a detailed description of the proposed invention, showcasing a sensor-based hybrid emotion recognition framework with applications in various domains.

Chapter 2

This chapter introduces role of emotions in e-learning and decision making. Heart Rate monitoring using wearable devices, feedback based on information fusion. Negative emotions, patter and pattern recognition, affective Human-Computer

Interaction (AHCI). Emotion modeling with limitations of facial emotions. Levels of information fusion with research gap, problem definition and research objectives.

Chapter 3

In Chapter 3, the research methodology unfolds, elucidating the pre-processing and organization of EEG and video data. Various algorithms are applied for feature extraction. The chapter delves into data integration and standardization, aligning with the objectives of the proposed work, while engaging in thematic discussions.

Chapter 4

Chapter 4, explores spontaneous affect recognition, integration of eNTERFACE and RML databases, and the development of a system incorporating audio features and visual cues are highlighted. The chapter unfolds with experimental insights within the proposed fusion framework, offering valuable comparisons with existing research.

Chapter 5

In this concluding Chapter 5, critical discussions on emotional assessment tools involving EEG, ECG, and EDS are presented. The chapter outlines future research directions and offers a comparative analysis with related studies, providing valuable insights into the evolving landscape of emotional assessment methodologies.

5. Chapter Conclusion

This research advocates for a Hybrid Human Emotion Tracking (HET) solution to enhance online learning experiences, emphasizing the importance of affective computing technologies. The proposed emotion recognition system integrates facial, vocal, and gestural cues, considering diverse factors. The study underscores the credibility of physiological signals like EEG for accurate emotion assessment. The novelty lies in its sensor-based hybrid framework, addressing the needs of online education, healthcare, and various industries. The envisioned future involves widespread integration of emotion recognition technology for improved human-machine interactions, supported by AI, deep learning, and the Internet of Things. Further this chapter includes outlined description of the succeeding chapters.

CHAPTER-2

2.1 Review of Literature Introduction

This chapter presents a review on multimodal human emotion information extraction and processing. It is divided into five major portions. The first portion deals with the emotion under the umbrella of affective computing. Affective computing can be seen as a problem of automatic recognition of human emotion for better communication between man and machine. It involves the perception and understanding of human sentiment and prediction of the mental state of the user. It also involves the analysis of emotional information of a human being in order to know its mental state of the human being. In the second section, facial emotion recognition is discussed with fundamentals of facial expression followed by emotion modeling. Facial expression representation is also present with the limitations of the system. The third section deals with the theory of emotion. This section reviews the major emotion theories such as categorical, appraisal and dimensional approach of emotion reported in literature. Various techniques for affective information extraction and processing are discussed under the section four and finally, section five covers the performance evaluation matrices for emotion recognition.

The research on emotion has been started from late 19th century with the Darwin's work on emotion. In 1970"s FACS was proposed by the [Ekman p. & Friesen, 1978]. FACS system created a milestone in the field of emotion research. Thereafter many researches are going on in the area of facial expression and emotion.

2.2 Role of emotions in e-learning

Emotions has an important role in the cycle of learning(Imani & Montazer, 2019). With electronic learning (e-learning) methodologies it has earned importance to understand the emotions of the learner. Some researchers have suggested the program should activate and lead the emotions of the learner to the correct state. But, at least, the program must understand the learner's emotions. There are various approaches for understanding the humanized gestures. The emotions can be acknowledged by enquiring participants, tracking instrumental parameters, authentication of voices, recognition of facial expressions, vital signals and recognition of gestures.

For e-learning programs the trainee is the operator of the system. For some reasons presented in this report, some of the methods for recognizing user emotional reactions has a relevance in the e-education methods and few of them are unacceptable.

2.2.1 Content emotional identification

An overview of multisensory information fusion study and use for content emotional identification. (Santos et al., 2020)(Jiang et al., 2020). This study thoroughly examines as well as summarizes the key technologies in the area of multisensory information fusion for research-based emotion detection, using real-time mental wellbeing surveillance systems as an illustration. Throughout discussions held regarding the various extraction techniques commonly in use for the open datasets, EEG, audio, visual, and writing features, feature layer fusion, judgment layer fusion, and categorization. Those conversations are intended to give a comprehensive review and a wide perspective of such a intriguing and trending field of study(Benoît Duc et al., 1997).

2.2.2 Emotions and decision making

The IoT encourages more individualized and rigorous treatment regimens. Its numerous applications have caught the interest of a sizeable portion of the science establishment. Those who want to use this innovation to improve their quality of life have also used this strategy. The study of successful techniques for automatically or semi-automatically transforming information from various sources and points in time into a representation that effectively supports either humans or computerized decision-making is commonly referred to as fusion. Smart, wearable, cloud-connected spectacles and a wristband for discrete and continuous (HR) heart rate measurements are available(Santos et al., 2020).

2.2.3 HR Monitoring using wearable's

The Pulse-Glasses wearable heart rate (HR) monitoring is a pair of wearable technology with electronic components which include a photoplethysmography (PPG) reader, a Bluetooth low energy (BLE) chip, as well as a replaceable battery. Clients of Pulse-Glasses can track their HR continually and covertly thanks to the (IoT) device links with a cell phone and cloud databases. The hardware for Pulse-Glasses has been developed, and preliminary trial findings demonstrate how efficiently the gadget performs in a variety of settings, covering every day routines(D'Mello & Kory, 2015).

2.2.4 Feedback based on information fusion

A composite sensory fusion-based recommendation system that take emotions into account will increase the effectiveness of recommender services. To thoroughly examine user features, three typical types of data are combined, greater rating accuracy for predictions and recommendations compared with traditional methods. This article's shortcoming is that it doesn't use internet recommendations in this case. Only unsupervised learning data is relevant to this approach. However, visible feedback data and collaborative filtering data may both be present at the same time in reality, therefore further research is needed to determine how to combine these two sources to improve the accuracy of predictions(Qian et al., 2019).

2.2.5 Pattern Recognition Letters

Here, a combined deep CNN and RNN model is suggested. Additionally, the suggested model to be tested using various conditions and hyper - parameters to ensure adequate tuning (Jain et al., 2018).

2.2.6 Frequency domain information

An approach using generalized sequence features to identify postural movements. Incredible stress and inertial data are processed concurrently using sensing and feature-level fusion. These emotional actions serve as a fresh stream of data to aid in emotion detection. Combining time- and frequency-domain information sets from each of the different deployed sensor nodes can identify the targeted activity with high precision(Gravina & Li, 2019).

2.2.7 Negative Emotions

A mechanism that detects negative sentiment and plays a clip to elicit +ve ones was put to the test. Anger, Grief, and Anxiety were the unfavorable feelings taken into account. The four participants in the tests were used. The Arousal and Valence scores for each emotion either increased or decreased, according to thorough EEG data. Researchers were able to come to the conclusion that the system that converts negative emotions into positive emotions can be constructed utilizing Brain-IoT after analyzing EEG data combined with HRV and GSR sensor information (Shirke et al., 2020).

2.2.8 Patter Recognition

A supervisory methodology is used in the individual trial of a person authentication method that uses the face and voice as different modalities. According to experimental data, the suggested fusion approach achieves success rates of 99.5%, which enhances the quality of particular specialist judgments(Benoît Duc et al., 1997).

2.2.9 Reliable emotion recognition system

The total recognition rate in recognizing the emotions was found to be 84.7% using SVM and 80% by using KNN algorithms. Using the physiological or frontal information in the suggested scheme also shows that developing a robust emotional identification system is possible even without requirement for increased psychological modality(Benoît Duc et al., 1997).

2.3 Emotion and Affective Computing

In HCI, the cognitive, affective and emotive information is more important to make better communication between a user and computers. It effectively improves the learning environment [Schaaff, K., 2008]. Human emotion detection has gained popularity in finding extensive applications in the fields like HCI and Human-Robot Interaction (HRI) [Cowie, R. et al., 2001] and other emerging fields as well. Affective Computing is an active research area of HCI. Affective Computing can be illustrated as follows:

2.3.1 Affective Human-computer Interaction (AHCI)

Researcher reported two approaches to examine sentiments. The first approach is based on categorization of emotion into distinct classifications such as joy, fun, sad etc. Alternative approach is to present feelings on multiple dimensions or continuous scale. The three most common parameters are valence, arousal and dominance. Valence scale measures how much happy or sad an individual is. The stimulation scale measures the calm, bore or excited etc. The dominance scale represents obedient (in control) or dominant (authorized) [Koelstra, S. et al., 2010].

Affective HCI includes human emotion detection from ventral expression and speech cue. Therefore, we will focus on above two modalities, particularly, in context to emotion recognition. One of the major requirements of MMHCI is that multi-sensory information should be taken individually then joined at the end. Its essential requirement for a hybrid modal should be able to cater with imperfect or noisy information as if information obtained from one modality is noisy, and then the information obtained from other modalities can be considered as complementary information. In other words, if one modality fails to derive the conclusion, the other must generate the conclusion, according to [Jaimes A. and N. Sebe, 2007].

2.3.2 Expression and Perception of Human Emotion Recognition

As sentiments are very common in our conversation, a journal authenticated definition of human emotion. However, Kleinginna has given a series of human computer interface emotional definitions as follows:

Emotion is a intricate interplay involving both personal and impartial elements, facilitated by neural and hormonal scheme. This intricate phenomenon can result in:

a) Affective experiences encompassing sensations of arousal, preference, or discontentment;

b) Mental practices involving sensitively applicable perceptual properties, assessments, and classification progressions;

c) Broad physiological modifications in response to arousing situations; and

d) Behaviors that are frequently expressive, goal-directed, and adaptive, though not universally so.

“[Kleinginna and Kleinginna, 1981]”

Automatic human affect recognition is achieved by capturing and extraction of information from various affective modes. To present human affective data, we have a broad variety of instruments and expedients to obtain speech statistics, visual signal, language contents and anatomical signals etc. Although we have multiple modalities for emotional information, the most widely used cues reported in literature are illustrated in Figure-2.1.

i. Facial expressions

Most of the research has been performed over facial expression to detect emotion. As the facial expression has an important role to play in our social and personal communication.

Facial gestures are one of the reliable and natural sources of communicating emotion. We can easily percept the happiness, sadness, disagreement and intensions from facial expression of another person during communication. Another advantage of facial expression is that facial expression is universally displayed by any subject of any age, gender. Therefore, facial expression is the major source/channel for affective computing.

ii. Audio

Audio is the prominent medium for verbal communication. A speech conveys emotion both through linguistic and paralinguistic messages that reflect the way of communication. The emotional state of the human being can be directly taken from speech in any case. It is proved from emotion theory that the affective state involves a physiological reaction. For example, if we are walking in a jungle, we feel fear. If any wild animal appears suddenly in front of us, we start shivering and sweating. Our bodies' physiological alterations are to blame for this. The tone of speech fluctuates as per emotional state of user. If we feel happy, our speech sound healthy and if we feel mellow, our speech sounds low.

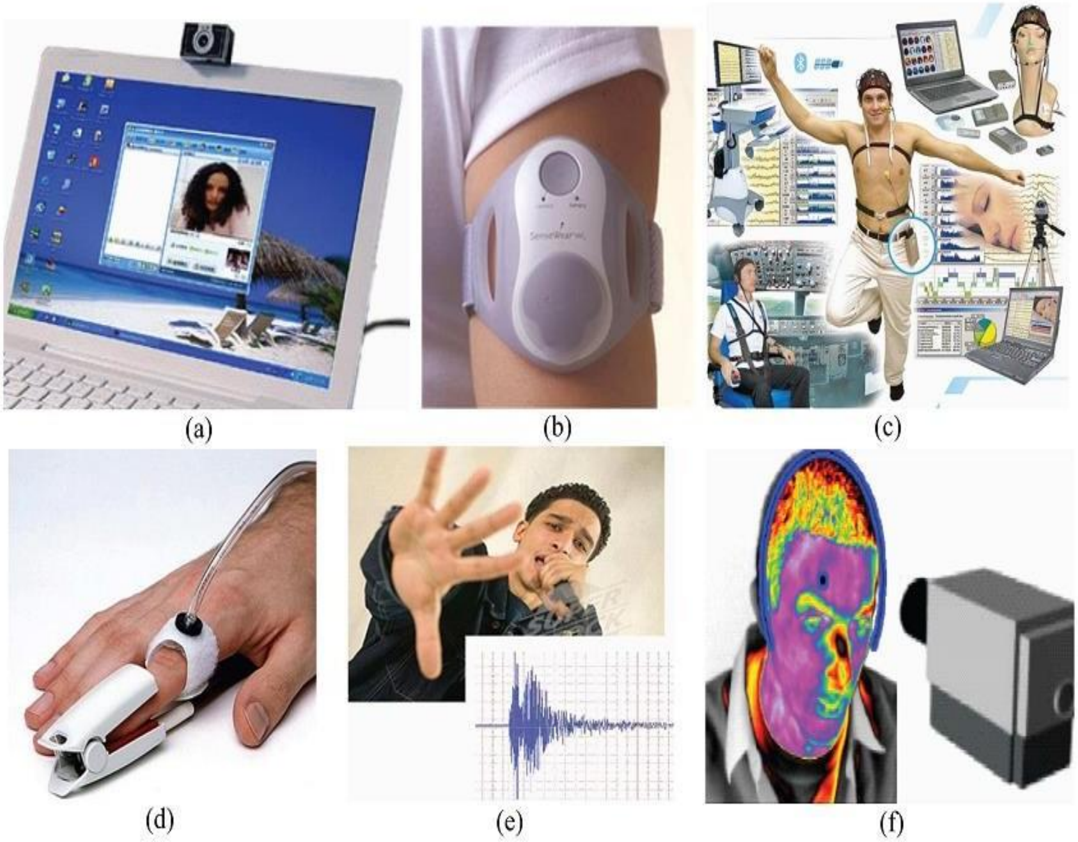


FIGURE 2.1 EXAMPLES OF AFFECTIVE MODALITIES/CHANNELS

iii. Physiological Gestures

Good number of emotional philosophies are centered on the physiological variations when sentiments are elicited. It is stated in emotion theory that the physiological changes occur when we feel emotions. Some research theories argued that we feel emotions as the alertness of the autonomic nervous system.

One of the drawbacks of using physiological signals is that measurement of physiological signal is a highly invasive way as sensors located on the user's body, which is not appropriate in the utmost accepted interaction.

iv. Gesture and Hand movement

Research in social psychology and human development has been emphasizing the fact that emotional states are expressed through body gestures [Hadjikhani, N. et al., 2003]. Many works related to analyzing and annotating bodily movement has been reported in literature, however, the complexity of mapping body gestures onto emotional facts are still an experiment. For example, after hitting the table with a sharp edge [Ekaterina P. et al., 2014].

The above-mentioned emotional channels can be studied individually or can be combined in order to merge complementary affective information. Many studies have proven that multimodal fusion increases the robustness and accuracy of appliance investigation of human sentiments. Detailed states of the art on the multimodal fusion of different affective modalities/channels are provided in next chapter.

v. Applications of Human Emotional Computing

Following are the applications/area where emotion can play a major role in communication between man and machine [Calvo R. A. et. al., 2010].

vi. ***Virtual Agents or Avatar:*** A virtual agent can express realistic facial expressions with different gestures and idle body movements. Many Hollywood movies are based on virtual agents or Avatar. There are various applications of virtual agents in the area of entertainment etc.

a. ***Robotics:*** *Todays" researchers are trying to develop intelligent robots, which can sense emotion, interpret emotion and act according to the sentiment art of stake holder.*

b. ***Marketing*** *& web- b a s e d applications are started using virtual agents to interact with customers to promote the services or sell goods.*

c. ***In gaming:*** *Treading novel interfaces are required in the design and development of multi-player games, where enormous quantity of customers interacts with each other using novel boundaries.*

d. ***Education:*** *As depicted covid-19 has taught us how important is online platform especially* ***o****nline education which is getting popular day by day as numbers of open courses are available online. In order to improve learning, it is a better idea to use the link of affect and cognition.*

e. ***Smart homes and office environments:*** *A smart home is equipped with automatic appliances, which respond according to emotional states of the user. The lighting conditions or playing music depends upon the emotion state of the user.*

f. ***In Medical:*** *There are number of applications with emotional approaches in medical science which has proved very effective.*

2.4 Identification of Facial Emotions

Identification of facial expressions is the prominent way to express emotion by human being as it plays an important role in everyday communication among humans. Whenever a person interacts to someone, the facial contact is established first during communication. It is the first and foremost observation toll which depicts on the face. The commonness in facial emotions is that the expressions keep on changing and the result to be generated out of facial expressions becomes tough. Let us explore what authors say:

2.4.1 Facial Expression

Among non-verbal communication body expressions especially facial expressions play an important role. It is one of the most research-oriented areas under human computer interaction. The beginning of research work in facial expression starts with the publication of [Charle's Darwin, 1988] book, "The Languages of the Sentiments in Man and Animal". Said work has gained the attention of many researchers towards the study of facial expression. According to Darwin non-verbal communications are "species-specific" but not "culture-specific". The summary of research work done in field of facial expression by different researchers is as follow-

- Facial expressions reveal emotional behavior of somebody or animal when he is not trying to fleece his emotions.
- Animals or person cannot be able to hide their emotions if it is not preplanned earlier.
- Observer can predict the emotion of species with greater accuracy if some possibilities.

- Species who mostly use impulsive facial expressions through statement then it is so hard for them to hide.
- Facial expressions very from species to species of different races.

By summing up, it is quite clear that expression is one of the significant factors for identification of human behavior and it is also not possible for them to hide it easily without preplanning. So, the main problem is to find out which expression belongs to which behavior and identifying them accurately.

2.4.2 Emotion Modeling

Emotion recognition involves three main stages: i) emotion modeling ii) emotion classification and iii) evaluation of the system. To model the emotion, we must know that how the emotion arises in human beings? We must explore the cognitive science in order to find out the solution of the question [Lutz et al. 1990]. Here, we are mainly focus on the emotion involved during interaction of human beings [Ekman, 1993], [Nesse, 1990]. Facial action units, given by Ekman [Ekman P. et al., 1978] are one of the most prominent visual cues to sense emotion from facial expression. A facial action unit (FAU) plays major roles for sensing the visual information for automatic affect sensing and recognition. Dimensional model is important in order to capture these facial action units effectively. Some examples of linguistic description of AUs are "inner and outer portion of the brows is raised" (AU1), & (AU2) etc. Figure-2.2 illustrates some AUs.

FIGURE 2.2 FACIAL CODING SYSTEM OF EKMAN AND FRIESEN [P. EKMAN AND W. FRIESEN, 1978], EXAMPLE OF ANTERIOR ACTION UNITS

2.4.3 Facial Expression Representation

The main parts of our body which involve showing the emotional expressions are mouth, eyes, eyebrows, cheeks and chin. Other facial expressions may be wrinkles and bulges [Bindu Maringanti H. et al., 2007], [Fellous J.M., 2004].

Most of the facial expressions are visible with other neighborhood parts, so it is required to perceive the face and calculate the corresponding features from surface only. The other way to subtract the background from facial expression is background subtraction method. The enhanced features are rich in terms of emotional information [Jingfu Y. et al., 2004], [Matthew N. et al., 2002]. Facial features are also represented by texture information, which provide information about high spatial gradients. The texture features are well able to represent emotions. Illumination and noise are the two most factor that affect the system performance. The illumination is the lighting variation which is due to uneven reflection of light from eyes, teeth and skin. It is hard to internment the emotion appearance under different clatter and illumination conditions. However, many robustalgorithms exist to reduce the effect of illumination and noise.

2.4.4 Limitations of Facial Emotion

Even though human recognize facial expression without any delay or much effort, the recognition of facial expression by machine is still a challenge in the field of outline appreciation due eco-friendly change, real- time handling requirement like interval and cosmos etc. The above problem has attracted the pattern recognition researchers towards this field. Some other key challenges are feature variety or abstraction for recognition, preparation and task segmentation in various circumstances [Axelrod, Lesley A., 2011]. Most of the emotion recognition approaches work under some constraint. Some of the constraint are-

- ❖ View or pose, position and orientation.
- ❖ Scrabbling of bean
- ❖ Scrambling in the atmosphere and illumination.
- ❖ Complex background of picture.
- ❖ Partial or full Occlusion and uncontrolled lighting condition
- ❖ Facial variability arising from gender, illness, age etc.
- ❖ Shaving or growth of beards, facial hair and make- up
- ❖ Facial deformation

Scaling, position and orientation of head are such constraint which can be avoided because there are some preprocessing techniques that are robust to translation, scaling and in plane rotation. The most difficult constraint is out of plane turning of face image which fallouts in different view of appearance. Therefore, more research is to be done for pose invariant expression recognition. Environmental clutter, Illumination, complex background and occlusion has negative effect on expression recognition because they cause the loss of some important facial features. Some authors have proposed the approaches for expression recognition under such constraint but in operation environment these constraints cannot be avoided. So, there is a need of adaptive expression recognition system which can adopt to environmental condition. But researchers have not focused their attention towards these adaptive systems. Since facial expression have some acoustic characteristics, so some authors have suggested that the combination of acoustic and visual characteristics of facial expression can be used for developing a robust expression recognition system.

2.4.5 Facial Expression Classification Techniques

To categorize the user's facial emotion, a classifier must be used with an autonomous facial model. The classifier to be used to categorize emotion may be linear and/or non-linear classifier. A linear classifier is used for input facial image or frame of a facial video however, non-linear or dynamic classifiers are used to analyze facial video sequences or the emotion information data in time series. A table of various emotion recognition systems with their accuracy and modality used has been given in Table 2.1.

TABLE 2.1 EMOTION DETECTION SYSTEMS WITH METHOD USED, MODALITIES AND ACCURACY

Author(s)	Year	Approach	Modalities	Accuracy (%)
S. Loelstra et al.	2013	PSD Facial action unit.	EEG and facial expression	86.70
S. M. Lajevardi et al.	2012	TPCF	Video	87.05
M. Soleymony et al.	2012	EEG	Video	76.40
M. A. Nicolaou et al.	2011	MFCC for audio, feature point selection for video	Audio-visual.	91.0
A. Chakraborty et al.	2009	Segmentation and localization of frames for feature	Video	88.0
Y. Wang et al.	2008	MFCC and Gabor wavelet	Audio-visual	82.14
A. Kapoor et al.	2007	Pixel difference of mouth region	Video	79.17
M. Pantic et al.	2006	Facial profile points, rule based	Video	86.3

The major techniques employed to recognize emotions are SVM, ANN etc. PCA, a dimensionality lessening technique has been widely used in several facial countenance acknowledgment [Sha T. et al., 2011], [Soyel H. et al., 2010; 2009] and [Srivastava R. et al., 2009]. SVM, a state-of-the-art ML technique has been used by several researchers in their work. **Theory of Emotions**

2.4.5.1 Categorical Approach

Ekman‟s work [Ekman P., 1993] created a base for much of the emotion research. He proposed six basic emotions, from which all other emotions are composed. The basic 6 human sentiments proposed by Ekman are as shown in Figure 2.3 and Table 2.2.

FIGURE 2.3 EKMAN‟S SIX UNIVERSAL EMOTIONS

TABLE 2.2 PRIMARY, SECONDARY AND TERTIARY EMOTION BASED ON PARROTS‟ CLASSIFICATION OF EMOTIONS (2001) [PARROTT, W., 2001.]

Basic Emotion	**Supplementary feeling**	**Tertiary Emotion**
Love	Affection	Admiration, love, affinity, enjoying, desire, concern, tendencies towards chaos, mercy, tenderness
	Lust	stimulation, want, intensity, and enthusiasm
	Longing	Pining
Joy	Cheerfulness	Pleasure, excitement, relish, brightness, joy, jubilee, exaltation, contentment, rapture, and nirvana are all adjectives used to describe fun.
	Zest	the feeling of pleasure, delight, or pleasure.
	contentment	Satisfaction and happiness
	Pride	Glory and victory
	optimism	enthusiasm, confidence, and faith
Surprise	surprise	Amazement, surprise, astonishment
Anger	Irritation	Unrest, mimicry, discontent, crankiness, and fussiness
	Exasperation	Exasperation, frustration
	Rage	Antagonism, vigor, venom, hatred, loathe, contempt, malice, animosity, dislike, and resentful are all forms of angry.
	sadness	Pessimism, anxiety, dread, futility, doom, pessimism, morose, grief, dissatisfaction, grief, pain, misery, and remorse
	disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	sympathy	compassion and pity
Fear	Horror	alarmism, surprise, anxiety, horror, dread, fright, frenzy, and un-fulfillment
	nervousness	Tension, tension, unease, concern, discomfort, and dread are some synonyms for these feelings.

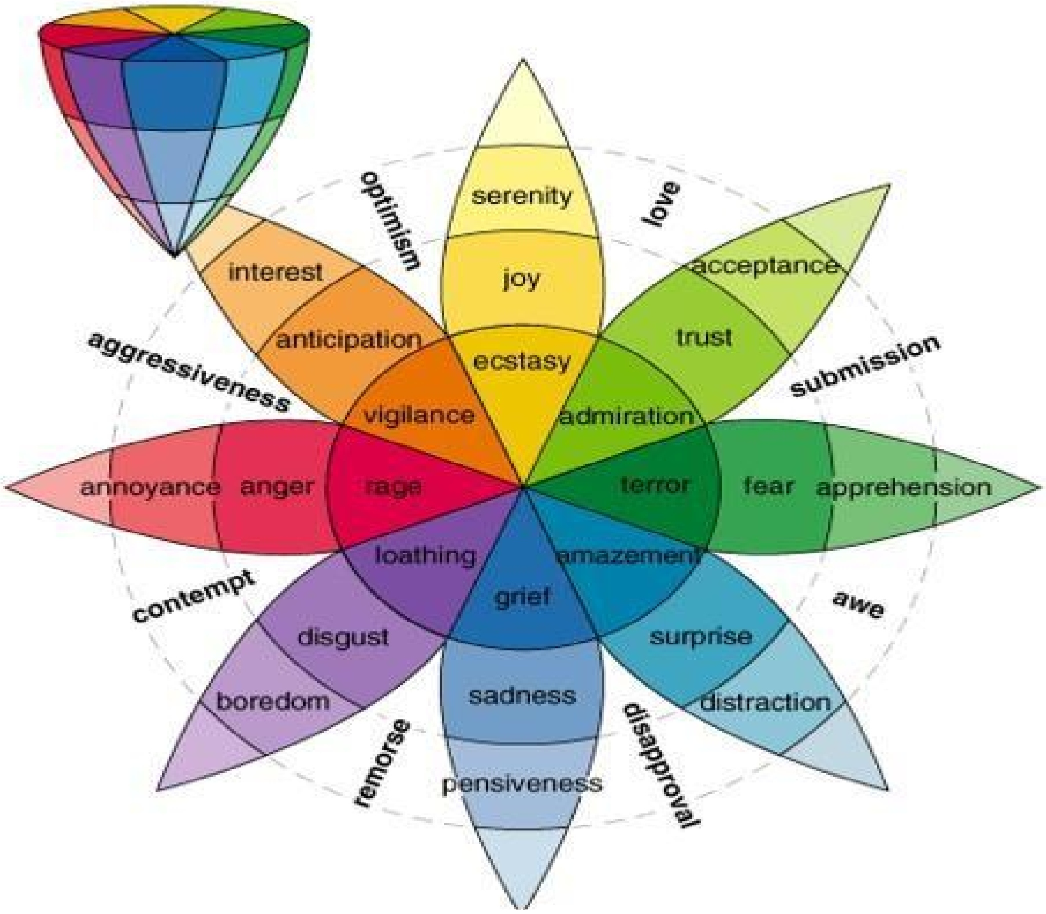


FIGURE 2.4 ROBERT PLUTCHIK'S WHEEL OF EMOTION [LUTCHIK, R., 2008.A PSYCHO-CONCEPT WITH GENETIC ALGORITHM]

2.4.5.2 Affective Information Extraction and Processing

In real life communication information is shared among many persons and ease of communication is really an important aspect. In human to human communication nonverbal cues are often used like facial expression and variations in voice tones. But unfortunately, we do not have a good human-computer interface, which can take advantage of these valuable communication mediums. Emotion recognition can be done by using single cue only or by multi-cue. In multi-cue technique, emotion information about different types of inputs is taken e.g. audio, video or image etc. We have used multi-cue sentiment recognition technique. Some researchers also applied this multi-cue technique [Koelstra S. et al., 2012].

2.4.5.3 Multimodal Information Fusion

Multi-modal information refers to the data acquired from multiple modalities such as audio, video, and text etc. Multimodal information fusion is defined as combining information from multiple sources/modalities to achieve higher performance than the performance achieved by means of a single source/modality [Verma Gyanendra K. et. al., 2011].

Information may be extracted from different source/modalities i.e. from text, image and speech separately. Low- and high-level features are to be extracted from various information sources depending on the applications. These features may be fused at low level (feature fusion) or high level (decision level) fusion. In this study we are presenting information fusion methodologies and related works captured in review.

2.4.5.4 Information fusion

At earliest fusion, data is fused in beginning i.e. at sensor or signal level. For instance, a three-dimensional image can be generated by fusion of two or more two dimensional images. [Cees G. et al., 2005]. The fusion of audio-visual information is another example of early fusion where audio and video modalities are combined into a single feature vector after applying dimensional reduction algorithms. The two modalities are fused before classification process. One of the major advantages of early fusion is that we can assign weights to each modality. This type of fusion is known as adaptive fusion. Adaptive fusion is useful where we want to assign more weightage to a modality.

There are various methods such as Bayesian Inference, Dempster–Shafer fusion, Maximum Entropy model and Naïve bay‟s algorithms reported in literature to perform fusion. A Bayesian inference-based approach was put forwarded by [Pitsikalis, V. et al., 2006] for fusion of the audio-visual features.

2.4.5.5 Intermediate Information Fusion

Here initial shortcoming of fusion techniques was its drawback with the data. Further it avoids the explicit methods of variant modalities which results as default to provide fluctuation in relative reliability as well as the asynchrony problem between the different streams [Sebe, N. et al., 2005b].

2.4.5.6 Late Fusion

Multimodal system integrates "a variety of approaches to achieve a shared understanding. It requires not only a precisely articulated procedure for incorporating the partial context but also a consistent framework for representing the common meaning of all the modes being used." [A. Jaimes and N. Sebe, 2007]. They claimed that D-S theory-based fusion outperform against the individual approaches and reduces the training time. Beal et al. [Beal, M. J. et al., 2003] proposed graphical models to process audio and video data jointly for object tracking. They applied the graphical model to combine audio and video observations and model parameters were learned from a multimedia sequence using expectation-maximization algorithm. The object trajectory is inferred from the data via bay‟s rule. An Artificial Neural Network fusion method was proposed by Gandetto et al. [Guironnet, M. et al., 2005] to detect human activities in an environment. They performed decision level fusion to fuse the sensory data obtained from a camera.

2.5 Levels of Information fusion

Sanderson and Paliwal [Sanderson and Paliwal, 2002] defined two broad categories of information fusion levels i) pre-classification and ii) post-classification fusion. Pre- classification relates to fusion before matching and post-classification relates to fusion after matching. Pre-classification schemes are subcategorized into a sensor level and feature level information fusion while post-classification schemes are subcategorized into decision levels. Pre-classification fusion is more challenging as classifiers used by the individual modality may no longer be relevant [Poh., N., 2010]. Different levels of information fusion are illustrated in Figures 2.5 to 2.7 below.

2.5.1 Blending of data or sensors

2.5.2 Synergy on a feature-level

2.5.3 Decision-level fusion

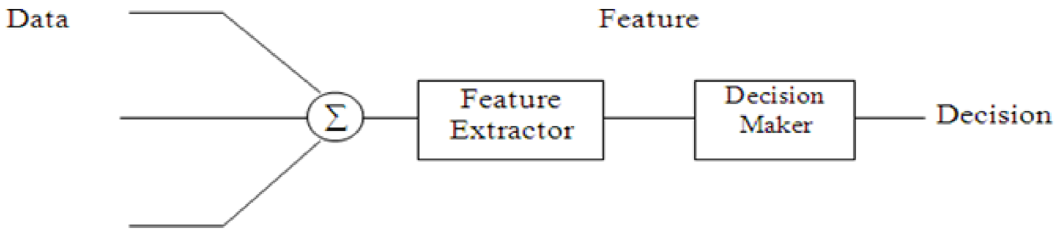


FIGURE 2.5 SIGNAL LEVEL OF GENERAL ARCHITECTURE OF DATA FUSION

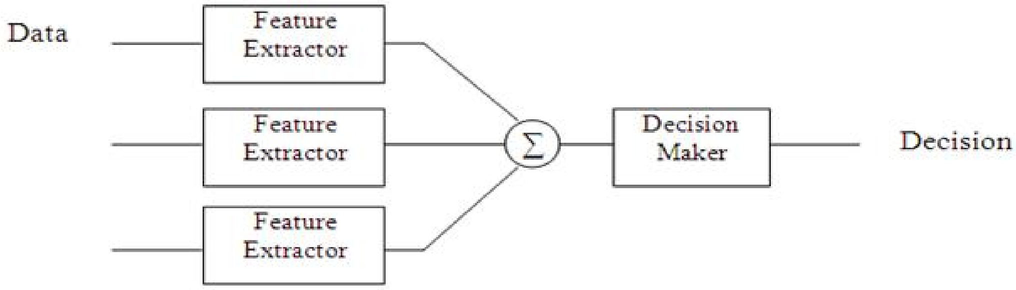


FIGURE 2.6 FEATURE LEVEL OF GENERAL ARCHITECTURE OF DATA FUSION

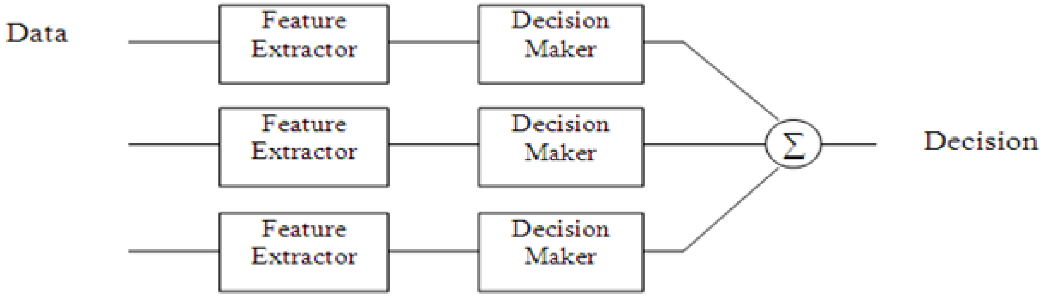


FIGURE 2.7 DECISION LEVEL FUSION OF GENERAL ARCHITECTURE OF DATA FUSION

2.5.4 Evaluation

This section summarizes the common practices in performance evaluation of emotion recognition system. Performance evaluation can be carried out at three levels: technology, scenario and operational [Phillips P. et al., 2000].

2.5.5 Evaluation Criteria

We may have three important evaluation concepts for multimodal information fusion.

- Kinds of mistakes
- benchmarks
- Presentation standards

Kinds of Mistakes

There are primarily two sorts of faults in this requirement: false acceptance rate (FAR) and false rejection rate (FRR)

False Acceptance Ratio:

FAR is the ratio of number of the acceptance (NA) to the total number of imposters (NI).

$$FAR = \frac{NA}{NI} \tag{2.1}$$

False Rejection Ratio:

FRR is the ratio of no. of reject (NR) to the total no. ofclients (NC).

$$FRR = \frac{NR}{NC} \tag{2.2}$$

2.6 Threshold Criteria

A threshold parameter is a method of selecting a barrier that must be tailored to a development set. We must establish a threshold criterion in order to choose the most appropriate threshold. Equal Error Rate (EER) and Weighted Error Rate (WER) are the most commonly used metrics.

WER is defined as:

$$WER = \alpha FAR + (1-\alpha) FRR \qquad (2.3)$$

Where $\alpha \varepsilon (0, 1)$ balances between FAR and FRR.

The value of α will be 0.5, if the no. of clients and the no. of imposter is equal. In this case

$$EER = \frac{1}{2}(FAR + FRR) \qquad (2.4)$$

2.6.1 Performance criteria

The final effectiveness of a method is evaluated by using Half Total Error Rate (HTER).

$$HTER = \frac{FAR+FRR}{2} \qquad (2.5)$$

Another performance measure is gain ratio. Gain ratio refers to the gain obtained outof a fusion experiment. Suppose we have i = 1, 2… N base-line methods.

$$\beta_{mean} = \frac{mean_i(HTER_i)}{HTER_{com}} \qquad (2.6)$$

$$\beta_{min} = \frac{min_i(HTER_i)}{HTER_{com}} \qquad (2.7)$$

Where $HTER_i$ is the HTER evaluation criterion as part of the expert I and $HTER_{com}$ is the HTER connected to the universal platform. β_{mean} and β_{min} are the proportion of the HTER of the specialist in combining the mean and the minimum HTER of the underlying experts i = 1,2,…., N.

2.7 Research Gap

After reviewing the literature from renowned digital repositories available in the form of transaction and journal papers, books, patents and web-links of scholarly articles relevant to human emotion recognition, research gaps are identified and presented as below. Researchers have done commendable work in emotion recognition for simulated emotions where only one clear and distinct hard label is associated with each input. Researchers have experimented with local as well as global features to a large extent for recognition of emotions. The induced and natural emotions are called as complex emotions as there is combination of emotions present in each utterance. Survey reveals that not much work is done for such naturalistic/complex emotions.

There is need to design approach which can address multiple challenges of emotion recognition simultaneously. These challenges are in the form of illumination variation, rotational changes, pose variations, occlusion etc. The impact of emotions on education parameters has not been studied much so far and hence there is a need to study this aspect and incorporate it in emotion recognition model which can add value to education system epically when it is in online mode.

2.7.1 Request Identification and Study Goals

Problem definition and research objectives laid down for the work undertaken are as stated below.

2.7.2 Problem Definition

"Design and development of novel emotion recognition framework for virtual learning assessment using information fusion".

2.7.3 Research Objectives

1. To study and analyse various existing technologies for virtual learning assessment model.
2. To prepare real time data-set using sensor based IOT and Machine learning algorithms.
3. To Design and Develop the proposed framework for Virtual Learning Assessment using Information Fusion.
4. To evaluate the proposed framework with the standard metrics.

To evaluate and analyse said objectives following work has been summarised in tabular form and accordingly research has been carried on to overcome the gap wherein result depends on more than one source to prove accuracy in human emotion recognition for the designated problem. Further the scope is extended for n number of source inputs to be fused for more accurate and reliable results. Table 2.3 represents findings and conclusions of recent research articles.

TABLE 2.3 FINDINGS AND IMPORTANT CONCLUSIONS OF RESEARCH PAPERS

Author (Year)	**Database Used**	**Feature Extraction Technique**	**Classification Technique**	**No. of Emotions**	**Accuracy In %**
ZHONG-MIN WANG et al. (2019)	DEAP database	SVM	NMI	32 Channels	Valence & Accuracy (75.41-74.21)
Jianzhu Guo et al. (2018)	iCV-MEFED data set	CNN - Convolution Neural Network	Neural Networ k	6 basic classes & 45+ dominant & complementary classes	Range of Rate of Miss-classificati on is 0.37-0.68
Chao_Qi et al. (2018)	CK+	Local Binary Pattern	SVM	6	84%
			Softmax		88%
Bing-Fei_Wu et_al. (2018)	CK+	Deep Convolution Neural Network	Adaptiv eFeature mapping	7	87.78%
	Amsterdam dynamic facial expression Set				90.57%
Biao Yang et al. (2018)	JAFFE Oulu-CASIA	WMDNN - weighted mix-ture deep neural network	Neural Net-works	6	92.20%
Yuanyuan Ding et al. (2017)			CNN-Convolutio nal Neural	6	91.86%

			Network		
Kamlesh_Mistry et al. (2017Y)	CK+	MLBP-		7	90.70% (SVM)
MD. Zia Uddin et al. (2017)	Locally created Video database	LDPP-Local directional position pattern	DBN-Deep Belief Network	6	92.50%
S.L. Happy et al. (2015)	CK+ Database	Facial patch based local features	SVM	6	93.09%
Kingsley Oryina Akputu et al. (2013)	CMU-PIE	SVD – Singular Value Decompositio n	HMM – Hidden Markov Model	6	90%
Usman Tariq et al. (2012)	GEMEP – Geneva Multimodal Emotion Portrayals - Facial Recognition	HG (Hierarchical Gaussianianiz ation – Patch based SIFT (Scale Invariant	SVM	6	80%
	and Analysis database	feature transform) and MF(Motion Features) – Point Based			
Ligang Zhang et al. (2011)	JAFFE	Patch based Gabor Features	SVM	6	92.93% (Linear Kernel)
Emily Mower et al. (2011)	IEMOCAP	Geometry based facial points	SVM (RBF Kernel	4	68.20%
Seyed Mehdi Lajevardi et al. (2010)	JAFFE Database	Zernike moments (order 10)	NB – Naïve Base Classifier	7	92.80%

	JAFFE Database CK+ Database				73.20%
Filareti Tsalakanido u et al. (2010)	Real Time Images	Geometry based features	Rule Based classificati on	5	75.42%
R. Ramathan et al. (2009)	JAFFE	2D Gabor Filters	SVM	7	83.30%
M. Karthigaya m et al. (2008)	SEA -South East Asian Face database		GA – Genetic Algorithm and NN- Neural Network	7	87%

2.8 Chapter Conclusion

Literature related to emotion recognition is reviewed with focus on emotion detection techniques and emotion classification techniques. Literature covering emotions & online learning is also studied. The important findings are used as supporting facts while analysing the affect online assessment data. The research gaps are identified which justify the need of an HERS model based on novel algorithm which can address the issues and challenges of emotion recognition and address the problem with greater accuracy.

CHAPTER-3

3.1 RESEARCH METHODOLOGY

The recommended architecture uses EEG plus video evidence from the DEAP database along with machine learning techniques for identifying sentiments (Moin et al., 2023). Multi-modal framework for recognition of emotions, as depicted in Figure 3.1 Thresholding in this case entails sanitizing the EEG data and selecting the best frequencies for feature extraction information. The content from the chosen sources was then rendered. The featured set combines HOG and LBP components from subjective visual data with spectral and statistical properties from EEG. Following feature set removal, the SVM, KNN, and ensemble algorithms were used to characterize the data. The learning of all classifications is completed by leveraging characteristics taken from all the respondents at the same moment because we are doing inter emotion regulation. The classifiers would have been educated on each subject's knowledge independently if it were mono emotion identification, resulting in a learned model on every subject. Figure 3.2 shows a visual representation of the desired architecture and provides a full justification.

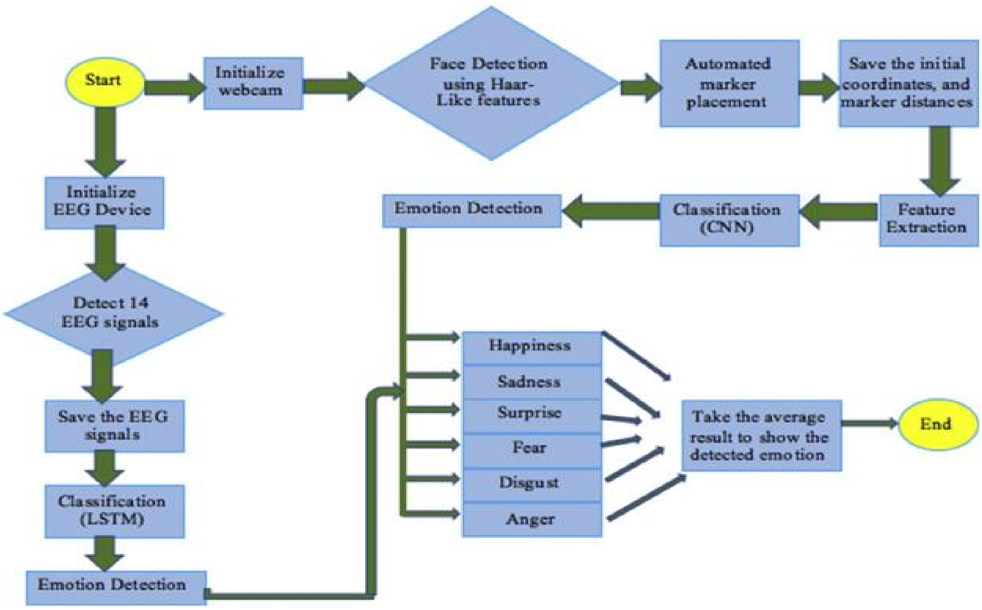


FIGURE 3.1 HYBRID EMOTION RECOGNITION STRUCTURE

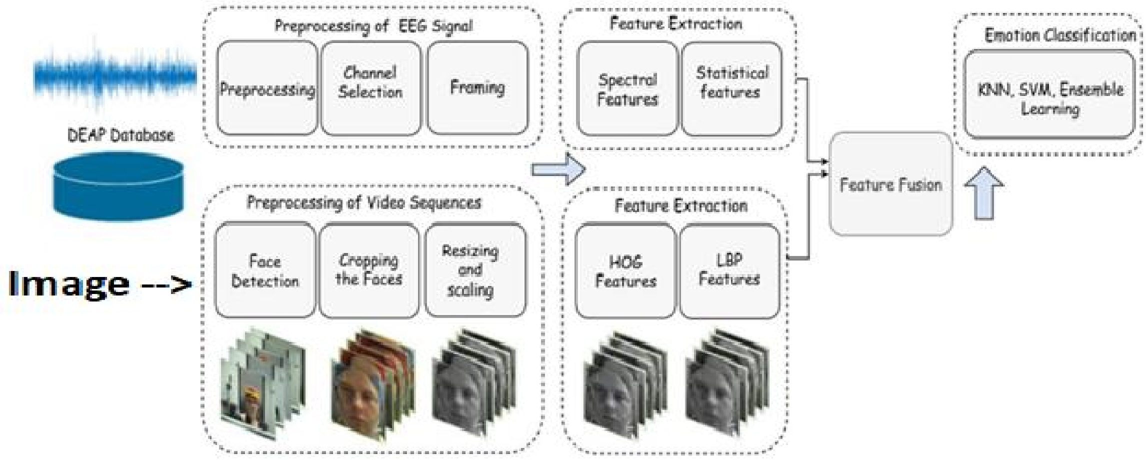


FIGURE 3.2 DIAGRAMMATICAL OVERVIEW OF OUR PROPOSED FRAMEWORK

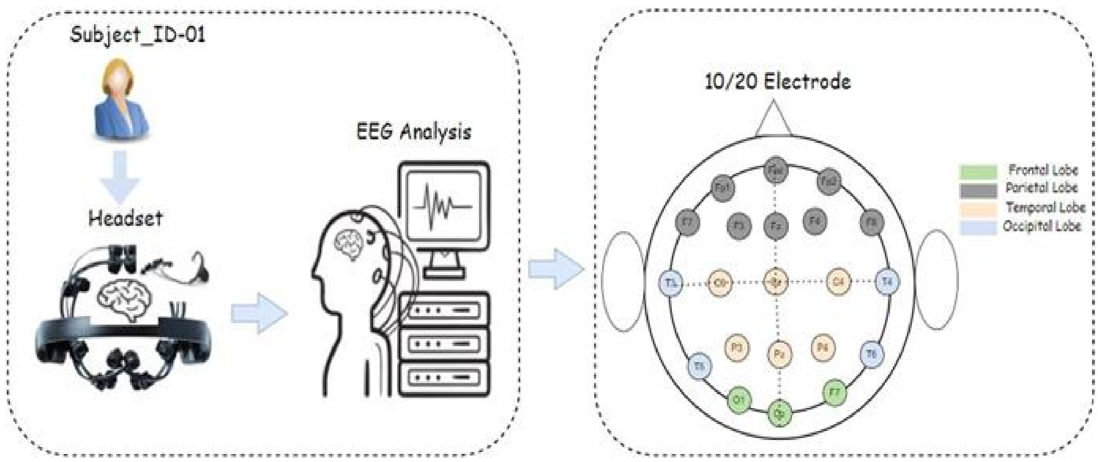


Figure 3.3 using the 10/20 electrode system for EEG data collection

3.1 Pre-processing

The data from the two distinct paradigms were standardized in the first phase to put them in a manner that could be used. This is a description of the pretreatment stages for both kinds of information.

3.1.1 Pre-processing of EEG data

As depicted in Figure 3.3, EEG data is collected using electrodes placed at various points on the skull. These sensors also pick up other artifacts including EOG artifacts in addition to ECG signals. They must be eliminated since they accumulate with eye blinking and movements. Here, these artifacts are eliminated using a blind origin technique, such as individual component analysis or EEGLAB. The EEG signals were initially down-sampled to 128 Hz. The band-pass frequency filter was then used to process signals with wavelengths between 4 and 45 Hz. The complete data also included 8064 observations, of which the first 384 specimens were eliminated because the first three seconds of data (128 x 3 = 384) was base-line data.

3.1.2 Arrangement of EEG data sources

Channel choice occurred just after EEG signals had been preprocessed in accordance to reduce the computational burden of the EEG sampling. In accordance with the 10/20 electro deposition, as illustrated in Figure 3.4, EEG sensors are attached to the skull for this purpose. Each sensor is given a unique name that included a symbol and a numeric. The number designates hemispheric locations, and the sign denotes the cerebral region. Each electrode was placed at a unique spot on the skull and was in charge of gathering data regarding a certain sentiments. The best electrodes were chosen in (Moin et al., 2023) using the technique for standardized interpolation, keeping both valence and arousal under surveillance. The prefrontal cortex was cited as being important in influencing emotional control in (Yang et al., 2018). In order to recognize emotions, (Spinelli et al., n.d.) used five sensors: P3, FC2, AF3, O1, and FP1, excellent results were attained. This shows that using all communication sources does not affect the results; rather, it lowers the results' reliability. As a result, some brain areas may be more important than others in the classification of emotions. Also, for feature extraction in (Moin et al., 2023) a grid search strategy was used. By verifying the findings of the authentication rate while applying the time-domain properties, the 32 connections were deleted one at a moment. The remaining four sensors, FP1, FP2, F3, and C4, got seen through the particle swarm optimization and were additionally employed in this investigation.

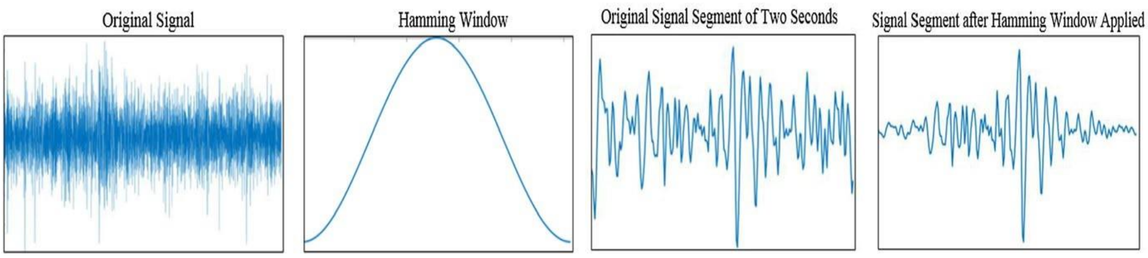


Figure 3.4 Hamming window-based reframing of EEG signals

3.1.3 EEG signal structuring

The EEG cue information subsequently transmitted via the framing procedure after the channel is chosen. The EEG signals have a tendency to be extremely complicated and unpredictable because they serve a representation of the neuronal activity taking place in the brain. Classification outcomes are subpar when stationary signal processing methods are employed to non-stationary signals. In order to solve this issue, EEG cues can be broken up into tiny chunks opting Hanning and triangle windows to get brief time frames, making the cue appear steady. The signal processing methods are then applied individually to every frame. This investigation partitioned the original signal into 30 segments using a Hamming window of 2 s with no overlapping. Characteristics were then extracted from the frames to create the feature maps. Figure 3.5 shows the structuring of the signals.

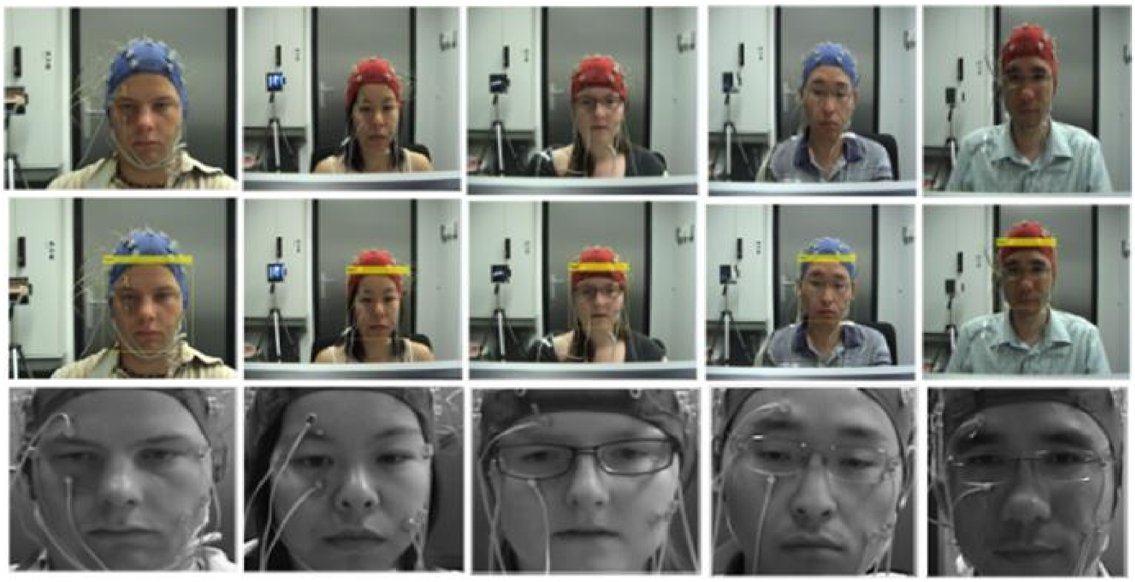


FIGURE 3.5 FINDINGS OF VIDEO SEQUENCE PRE-PROCESSING. THE FIRST ROW DEPICTS TO THE FACES OF VARIOUS OBJECTS, THE MIDDLE ONE DEPICTS RESULTS OF FACE RECOGNITION, AND THE THIRD ROW TO CLIPPED GRAYSCALE FACE FEATURES.

3.1.4 Pre-processing of Video data

The information from the movie segments needed to be preprocessed, just like the EEG signals did. At this phase, the frames taken from video sequences were trimmed to remove the unnecessary areas and only the human facial region's features could be recovered. Each video in this study has a frame rate of 50 frames per second. We recovered 30 frames from each movie and 1200 frames per participant, from each of the 3000 overall footage in each film. The Viola-Jones procedure was then used to carry out face detection on the photos.

The faces were then clipped and made into grayscale photographs. All edited images were finally scaled to a fixed size of 64 x 128. The collection of characteristics from these images was done after that. Figure-3.7 displays the outcomes of preprocessing of the video sequences.

3.1.5 Feature extraction

Many characteristics were gathered from the EEG data after the artefacts in the EEG signals were removed, the channel was chosen, and the EEG signal was framed. Table 3.1 is a list of the capabilities. The pre-processed photos' features were also extracted in a similar manner. A multifunctional approach yielded a total of 19 attributes, of which 17 came from the EEG cues and two from the streaming data. To create a single feature representation, all the retrieved features from the two separate modalities were combined. The subsections that follow offer a step-by-step clarification.

TABLE 3.1 EXTRACTIONS OF ATTRIBUTES FROM EEG & SEQUENCES OF VIDEO

Type of Feature	Feature Expected
EEG. spectral-features	i)Coordinates ii)Skewness iii)Kurtosis iv)Spread v)Entropy vi)Flatness vii)Slope viii)Crest ix)Decrease x)Roll-off point
EEG. statistical-features	i)Mean ii)Variance iii)Kurtosis iv)Skewness v)MAD(Mean-Absolute-Deviation) vi)RMS(Root-Mean-Square) vii)IR(Interquartile-Range)
Features-extracted out of video-frames	1. Variability in an Ordered Histogram 2. Individual Binary Sequence

3.1.6 Information Fusion Methodology

The act of combining data from many sources of data to produce information deemed more trustworthy, precise, and valuable than that provided by any one data source alone. Based on the operating stage at which fusion actually occurs, data fusion

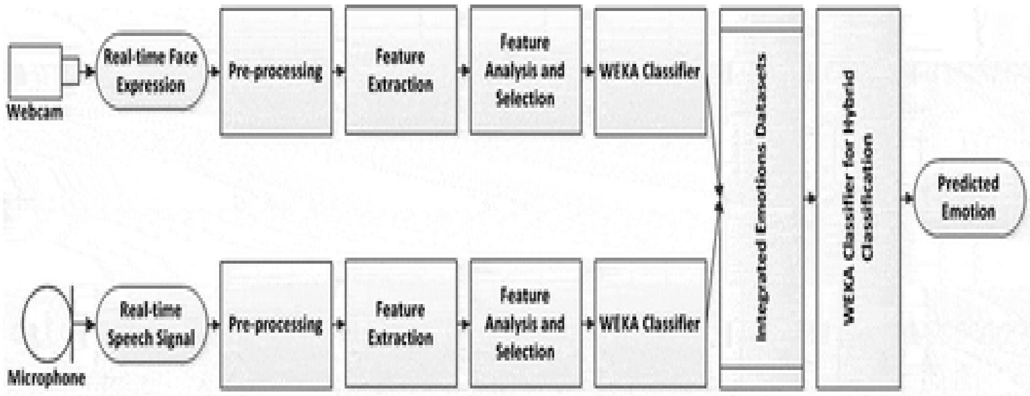


FIGURE 3.6 DATA FUSION COMBINING FACE/VOICE EMOTIONS

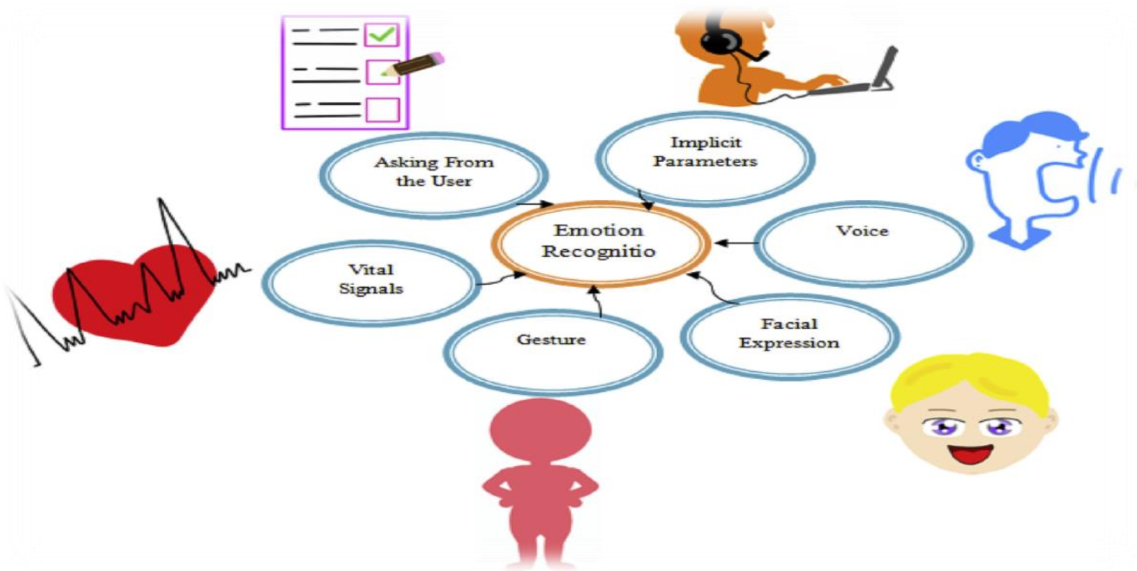


FIGURE 3.7 MULTIPLE SOURCE OF EMOTION RECOGNITION

technologies are also classified as low, intermediate, or large. In order to generate new raw data, low-level data fusion incorporates many sources of raw data. Fused data is supposed to be more descriptive and simulated than the initial inputs. Sensor fusion, for instance, is often referred to as (multispectral) information fusion and is a subset of data fusion. The theory of data fusion derives from the emerging capacity of humans and animals to combine multi-sensory information to strengthen their ability to survive. A blend of sight, touch, smell, and taste, for example, can indicate whether a substance is acceptable. In my study of Novel Emotion recognition framework, the combination of data is required, which is obtained from multiple sources for better results with high accuracy. The hybridized approach for data fusion of the integrated datasets for automatic emotion recognition as well as facial character recognition using software modules is shown in the Figure 3.6 above.

3.1.2 Novel Emotion Recognition

The modal of real time e-learning emotion tracking system used in my study. In this model, the tasks to be undertaken are like gathering emotional signals, labelling and selecting unlabelled emotional datasets from different edge-cloud sources, hybrid human emotion recognition, data fusion recognition, plus analysing the AI model in the remote cloud, and required sentiment input for decision-making to test whether the learners have understood the concept or an improvement is required in teaching via virtual mode. Figure-3.17 shows different sources of emotion confirmation and Figure-3.18 shows the example of real time sentiment gathering and feedback analysis.

Understanding that education is every society's most critical foundation where communications among nations and their citizens is improved so the world with its notable processes in the lifelong richness of knowledge & specializations. The actual learning environment, i.e., the human instructor (mentor) has a face-to - face interaction with the learner (scholar), should anticipate the most suitable reaction based on the learner's behaviour or affection, and also offers the most suitable instructional points. As in the electronic education system, the learning environment and material structure are typically portrayed in a static manner, without considering the sentiments of the learners and without interactivity and input from the teacher. Nevertheless, feelings will improve the level of learning when considering love. To address this issue, an interactive electronic learning system compatible with learner's affections was suggested. This strategy enhances learning efficiency, increases attention and strengthens memory. The emotion recognition methods already in place have been divided into **7** groups. Each of the seven has its own advantages and disadvantages. Their implementation part is coincided by different algorithms. But from the perspective they deal with emotion detection in 3 key stages, where they're similar:

1) Data preparation
2) Exploring the powerful interface for emotion detection.
3) Classification based on features derived or identified for their assignment of acceptable emotional states.

Here in my research, I have tried to touch good number of research areas via diagrammatically or with some explanation to enlighten hidden difficulties in e-learning and to have a model which proves effective results by combining multiple sources as an input for fusion to prove the accuracy of the model.

3.2 Feature extraction from EEG data

Human Emotion detection is a challenging method on account of the complexity, nonlinearity, and non-stationary of EEG recordings. So, for precise emotion recognition, relevant and discriminative feature extraction is necessary. 10 & 07 spectral / statistical features where taken respectively from EEG samples in this study. To perform the emotion categorization tasks, this recovered information was combined with visual features.

❖ **Accurateness**

This indicator evaluates the overall number of accurate predictions the approach has produced across all of the current classes (Moin et al., 2023). Typically, it is calculated as described in Eq. (3.1) by dividing the total number of assumptions that the system accurately predicted by all of the forecasts.

$$\text{Accurateness} = (T_{P.} + T_{N.}) / (T_{P.} + T_{N.} + F_{P.} + F_{N.}) \text{ ---------------(3.1)}$$

3.2.1 Spectral features

By translating time-domain to the frequency-domain via a Fourier-transform, one can obtain spectral-features, which are frequency dependent-properties. These characteristics define the signal's form. The ten spectral features that were taken out of EEG data are as follows:

3.2.2 Spectral centroid

A spectral centroid was used to symbolize the gravitational center of the spectrum. It is referred to as normalizing the frequency's weighted sum by its un-weighted sum, as indicated in Eq. (3.2).

$$\mu_1 = \frac{\sum_{k=b_1}^{b_2} f_k s_k}{\sum_{k=b_1}^{b_2} s_k} \text{---(3.2)}$$

In the above scenario, k stands for the bin, fk for frequency-corresponding to the bin, Sk for the spectral value at the bin, and b1, b2 for band borders in bins that need to be calculated for the spectral nucleus.

3.2.3 Spectral coverage

The instantaneous bandwidth of the spectrum is represented by the spectral spread. This is referred to as the standard deviation over the spectral centroid as per Eq. (3.3).

$$\mu_2 = \sqrt{\frac{\sum_{k=b_1}^{b_2} (f_k - \mu_1)^2 s_k}{\sum_{k=b_1}^{b_2} s_k}} \text{------------------------------(3.3)}$$

Where k stands for the bin, fk for the frequency that corresponds to the bin, Sk for the spectral value at the bin, b1, b2 for band-boundaries over which the spectral spread must be determined, and μ1 is the spectral centroid.

3.2.3 Skewed spectrum

The centroid's symmetry was represented via spectral skewness. It is a metric for determining how different from structure across mean frequency a spectrum's structure is below the gravitational center. It is computed using the third-order moment, as shown in Eq. (3.4).

$$\mu_3 = \frac{\sum_{k=b_1}^{b_2} (f_k - \mu_1)^3 s_k}{(\mu_2)^3 \sum_{k=b_1}^{b_2} s_k} \text{-------------------------(3.4)}$$

3.2.4 Spectral Kurtosis

The spectral kurtosis is a measure of the visual flattening or non-Gaussianity near the center. Additionally, the signal's-impulsiveness is quantified in frequency-domain by fluctuating-frequency. It is calculated using the 4th-order moment, as seen in equation (3.5).

$$\mu_4 = \frac{\sum_{k=b_1}^{b_2}(f_k - \mu_1)^4 s_k}{(\mu_2)^4 \sum_{k=b_1}^{b_2} s_k} \text{-------------------------(3.5)}$$

3.2.5 Spectral Entropy

The degree of disorder is assessed using spectral entropy. For a rate-map with multiple spectral peaks, the resulting feature is low; for a flat rate-map spectrum, it is considerable. Additionally, it represents the peak-ness of the spectrum and is computed using Eq. (3.6).

$$\text{Entropy} = \frac{-\sum_{k=b_1}^{b_2} s_k \log(s_k)}{\log(b_2 - b_1)} \text{ --------------------(3.6)}$$

3.3 Statistical features

EEG signals can be analysed using various statistical features to extract information about the brain's activity. Some common statistical features used in EEG analysis include, Mean, The average amplitude of the EEG cue over a identified time slice, Standard deviation, They are often used in combination with other techniques, such as wavelet analysis and machine learning algorithms, to improve the accuracy of EEG analysis. Statistical features use mathematical formulas that are applied to signals in order to extract information that is helpful. From the EEG data, we deduced seven statistical traits, which are described in more detail in the sections below.

3.3.1 Mean

As shown in Eq. (3.7), the mean is calculated by dividing the total-number of samples by their summation of all samples provided in a signal:-

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \text{ --------------------------------(3.7)}$$

Where N is the total number of samples, and Ai stands for the ith signal sample.

3.3.2 Variance

Applying equation 3.8, the variance for a signal A with N samples was calculated

$$V = \frac{1}{N-1}\sum_{i=1}^{N} |A_i - \mu|^2 \text{-------------------------(3.8)}$$

In the above scenario, μ stands for average of the signal A

The following statistical characteristics are also present:

Skewness, Kurtosis,

Mean Absolute Deviation, and

Interquartile Range.

3.4 Extracting features out of the video-clips

Extracting features from video clips is an important step in video analysis and processing. The type of features extracted will depend on the specific application and goals of the analysis. Here are some common types of features that can be extracted from video clips like, Color histograms; this feature describes the color distribution of the video clip. It can be used to identify objects in the video based on their color, Motion vectors; these features describe the direction and magnitude of motion in the video. They can be used to track objects in the video and analyze motion patterns, Optical flow, this feature describes the movement of pixels between consecutive frames in the video. It can be used to detect and track motion, and to identify regions of the video with different motion patterns, Texture features, and these features describe the texture of different regions in the video. They can be used to identify objects based on their texture and to analyze changes in texture over time, Shape features, these features describe the shape of objects in the video. They can be used to detect and track objects, and to analyze changes in object shape over time, Facial features, these features describe the characteristics of human faces in the video. They can be used to detect and track faces, and to analyze facial expressions and emotions, Audio features, These features describe the audio content of the video, including speech, music, and ambient noise. They can be used to analyze the audio content of the video and to detect changes in sound patterns over time. These are just a few examples of the many features that can be extracted from video clips. The choice of features will depend on the specific analysis goals and the available resources for processing the video data.

We extract the characteristics from the human face expressions after extracting the various spectral and statistical features from the EEG signals. The frames were initially taken out of the video sequences and scaled to the predetermined dimensions using software. Then, using two well-known image descriptors, the features from these gray scale images (frames) with human-cropped faces were extracted. The following provides an explanation of the HOG and LBP image descriptors.

1. HOG

A feature called Histogram of Oriented Gradients (HOG) is used to glean details about an entity's structure from an image. It works by obtaining information about the gradient (also known as magnitude) and approach (also known as direction) of the borders of the objects in the image. The gradient and direction are first established for each region once the input image has been divided up into smaller sections. The propensity and inclination of the pixel values are then used to create a histogram for each area. Equations (3.9) and (3.10) are used to compute the magnitude and orientation:

$$\text{Magnitude} = \sqrt{(G_x)^2 + \left(G_y\right)^2} \text{-----------------(3.9)}$$

$$\text{Orientation} = \left|\tan^{-1}\left(\frac{G_y}{G_x}\right)\right| \text{----------------------(3.10)}$$

Also between G_x and G_Y stand for changes in the x- and y-axes.

2. LBPs

LBP is a straightforward texturing manager that uses threshold holding near each pixel to identify the pixels in the image and outputs the result as a binary integer. The entire image is then separated into 3 3 cell-sized windows, known as cells. The neighbors of the core value of this 3 3 matrix were given new binary values using the central value as a threshold criterion. After completing this process, we were able to create a new image that accurately captured the traits of the original. Equations (3.11) and (3.12) illustrate how the LBP features are mathematically formulated.

$$\text{LBP} = \sum_{i=0}^{p-1} s(n_i - G_c)2^i \text{--------------(3.11)}$$

$$S(X) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases} \text{-----------------------------(3.12)}$$

Where p indicates the overall number of neighbour pixels, n_i is the ith neighbour, and G_C stands in for the central pixel value.

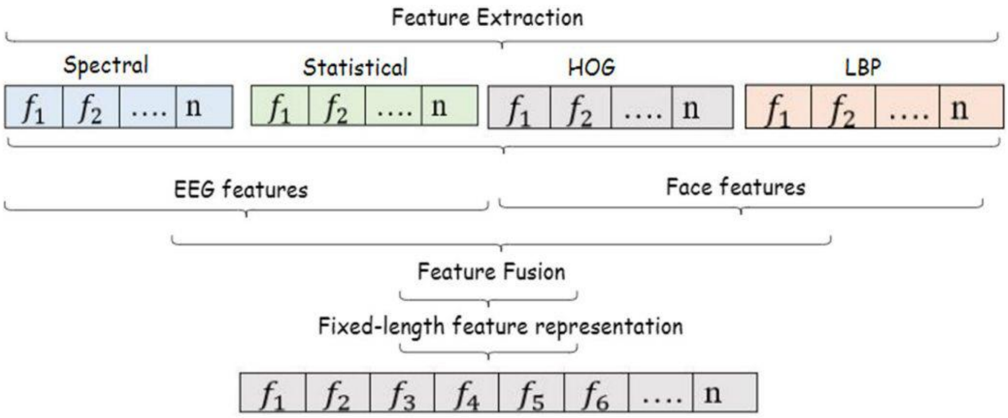


FIGURE 3.8 FEATURE FUSION PROCESS TO CLASSIFY EMOTIONS

3.4.1 Fused features

Fused features are a technique used in data analysis and ML to combine multiple sets of features extracted from the same or different data sources. The objective feature fusion is to enhancing efficiency of the analysis or machine learning model by providing a more comprehensive representation of the data. There are several approaches to feature fusion, including Early fusion, In this approach, while entering the characteristics from several sources into the research or computer vision framework, they are integrated into a single feature vector, Late fusion, In this approach, the features from different sources are analyzed or processed separately, and the outputs are combined at a later stage in the analysis or machine learning pipeline, Hybrid fusion, This approach combines both early and late fusion techniques to leverage the benefits of both approaches, Multi-modal fusion, In this approach, features from numerous modalities, such as audio, visual, and text data, are fused to provide a more complete representation of the data.

As illustrated in Figure 3.8, In order to execute sentiment analysis, the characteristics from each paradigm of numerous selected features were integrated into a single feature set. This feature representation was then utilized as entry to the sentiment detection algorithms. The final article of vectors for the multimodal approach were created by combining the temporal and statistical characteristics gathered after each two-second frames of each EEG experiment with the HOG and LBP patterns produced after each screen of each two-second video sequence of each experiment. We omitted the EEG data from several individual trials since video sequences weren't available, allowing all the characteristics to be properly concatenated.

3.4.2 Emotion classification

We carried out data standardization and class balance prior to classification. In order to standardize our data, we employed a z-score. Following is a full explanation of each of the three recommended frameworks created on these classifiers that were put into practice:

3.4.3 SVM

SVM is a linear prototype that can handle classification and regression issues that are both linear and nonlinear. It uses a hyper-plane to divide the data into two independent categories. Locating a hyper-plane (an ideal hyper-plane) that maximises the margins suggested by support vectors is more explicitly the focus of SVM. The margin is also the separation between these support vectors. Eq. (3.13) provides the mathematical description of the hyper-plane for the data point:

$$W^T x + b = 0 \text{ -----------------------(3.13)}$$

In the above scenario, w^T is stands for the slope of line whereas b is intercept. The SVM can handle a high-dimensional input space as well, as shown in Figure-3.9 However, If the data are nonlinear, the SVM uses a kernel function to map them to high dimensions and turn them into linear data. The nonlinear data that were collected for this investigation were transformed into linear data by using the Gaussian kernel function, which is specified in Eq.:

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \text{-----------------(3.14)}$$

Where is the variance, x y 2 denotes the Euclidean distance between points and y, and and and y are two points. Additionally, in all experiments, we used a fine Gaussian SVM with a Gaussian kernel function and a box constraint level of 1. All EEG-related investigations used as kernel scale of 2.1, while single-modality video sequences and multimodal trials both used a kernel scale of 16..

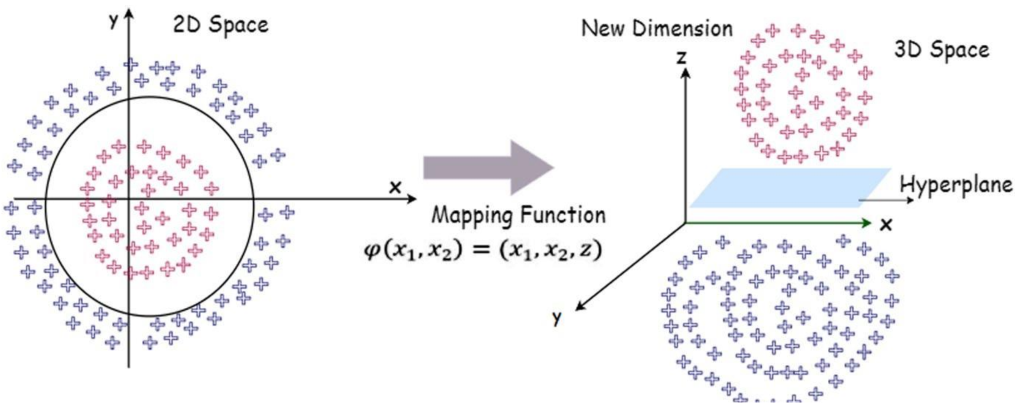


FIGURE 3.9 KERNEL PROCEDURE PROJECTION IN SVM

3.4.5 KNN

The best well-liked ML classifiers for classification and regression applications is KNN. The nearby data points form the basis of the KNN classifier's operation. The KNN algorithm's main goal is to deduce the tight relationships between similar items. To determine the closest neighbours, the resemblance or separation among datasets is determined. This study used the Euclidean distance, which is given below in equation (15) to calculate the distance:

Algorithm 1 Multiple-Modality-based Emotion Detection System for Effective Human-Computer Interaction

Input: EEG signals and Video sequences
Output: Emotion Classification (e-g If valence then class='High" or 'Low')
For i in range (1, n):
    $x \in EEG\ signals$
    EEG_final = Preprocessing (x)
    $y \in Video\ Sequences$
    $Video\_data$ =Preprocessing(y)
    Feature Vectors f_1= Feature Extraction (EEE_final)
        $f_1 = (Spectral \cup Statistical)$
    Feature Vectors f_2= Feature Extraction $(Video\_data)$
        $f_2 = (HOG \cup LBP)$
    Final Features $v \in \{v_1, v_2, v_3, \dots \dots v_n\} = Fusion\ (f_1, f_2)$
    Result $r_1 \leftarrow SVM\ (v \in \{v_1, v_2, v_3, \dots \dots v_n\})$
    Result $r_2 \leftarrow KNN\ (v \in \{v_1, v_2, v_3, \dots \dots v_n\})$
    Result $r_3 \leftarrow Ensemble\ (v \in \{v_1, v_2, v_3, \dots \dots v_n\})$
end for
Return result r_n

$$\text{Euclidian Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \text{------------(3.15)}$$

where the two data points are x1, y1 and x2, y2. One was chosen as the value for the total number of nearest neighbours, k. As a result, the label of the test instance was established at the time of testing by comparing its Euclidian distance to that of all training instances. The query instance's class label was determined in the following phase by choosing the top k = 1 instances. With one neighbour as the number of neighbors, Euclidean distance as the distance function and an equal weight for distance, we have utilized Fine KNN for all trials..

3.4.6 Ensemble strategy

The ensemble strategy uses a group of classifiers that work together to learn, and the final outcome is made by weighing or not weighing each classifier's unique predictions. The ensemble-based techniques include bagging, boosting, and stacking, among others. In this research, decision trees were used in conjunction with the bagging approach for individualized learners. Bagged trees, in which the results of multiple decision trees are combined using bootstrap aggregation, are another method for addressing the issue of over-fitting with individual decision trees. This method reduces over-fitting and improves the generalizability of the model. Additionally, we used an ensemble method called bagged trees with a mild boost, in which the decision tree served as the learner. Additionally, 40 students were included in the settings for the ensemble technique, and the learning rate was set at 0.1. Each experiment took place in the same environment..

3.4.7 Experimental discussion

Now that the information used in the research for sentiment detection and cauterization has been described, let's discuss the performance metric that was employed to assess how well the suggested technique performed. The outcomes are then provided for the various experimental contexts.

3.4.8 Compilation of a dataset

We used the DEAP database (Moin et al., 2023) for the compilation, which includes video clips as well as EEG and physiological signals. 32 people, ranging in age from 19 to 37, provided information. 40 movies lasting 60 seconds each were given to the participants to watch. Additionally, the EEG data were shown in two arrays: one with a length of 40 X40 8064, representing the total number of channels and videos, and the other with a length of 8064 representing the amount of samples each EEG electrode was able to retrieve while watching a one-minute video clip. For each video clip, the second array includes labels for valence, arousal, liking, and dominance on a scale of one to nine. Because only the first 22 individuals had access to video data for this investigation, we used their information. We ignored the physiological signals and just used EEG and face data. Only the classes of valence and arousal were used for emotion recognition, with the classes of dominance and liking being eliminated as well. The rating system runs from one to nine, with 6.5 being the highest and 1-3.5 being the lowest. The middle or neutral class of data, which is comprised of values between 3.5 and 6.5, was likewise removed from the proposed framework. According to (Moin et al., 2023), 1–3 are low, 4–6 are medium, and 7–9 are high.

3.5 Assessment metrics

Performance metrics included accuracy, precision, recall, and f-measure. A confusion matrix was used in the calculation of these metrics. Below is a detailed discussion of each of these.

- ❖ **Precision**

Precision calculates the proportion of true positives among all positive cases. This is calculated mathematically using Eq. (3.16):

$$Precision = TP/(TP+FP) \text{ -----------------------}(3.16)$$

❖ **Recall**

The recall is the proportion of correctly predicted positives among all observations in an actual class. Simply dividing the total number of true positives by the total number of valid estimates yields the recall. This is calculated mathematically using Eq. (3.17):

$$Recall = TP/(TP+TN) \text{----------------------------}(3.17)$$

❖ **F-measure**

Precision and recall were used to calculate the F1 measure or F1 score. This demonstrated the general accuracy of the advised method. It is calculated mathematically using Eq. (3.18).:

$$F1= 2x\ (Precision\ x\ Recall) / (Precision + Recall) \text{ ------}(3.18)$$

Here, TN stands for true negatives, TP for true positives, FN for false negatives, and FP for false positives in equations (3.15 to 3.18).

3.6 Data integrating & standardisation

For a classifier to produce accurate results, the normalization or standardization of the data is necessary. We used the z-score approach to normalize the data. The data were rescaled using this method, which used a Z-score with a zero mean and one standard deviation. We estimated the z-score with a value of using Eq(3.19) as below:-

$$Z = (x- \mu)/\delta \quad \text{---------------------------}(3.19)$$

where the variable's standard deviation is and the average value is μ. Before classifying, we also took into account the issue of class misbalancing. It is essential to balance the training dataset first since the degree of imbalance affects how accurately a classifier makes predictions. Up sampling and down sampling are the two methods for class misbalancing that are most frequently utilized.

3.6.1 Results and discussion for the same

The recommended model was rigorously tested under diverse experimental conditions using tenfold cross-validation to assess its accuracy in sentiment revealing from EEG signals. The study aimed to identify the most informative time segment for valence classification, segmenting the data into 10-, 20-, and 30-s intervals. Multimodal characteristics of human emotion were employed across these segments and the entire signals, validating three classifiers (SVM, KNN, and ensemble). The ensemble classifier consistently outperformed, achieving the highest accuracy across all segments, with a peak accuracy of 93.1% in whole signal classification. These results emphasize the superior performance of the ensemble approach in valence classification, findings which were consistent in arousal classification on the entire signal.

TABLE 3.2 RESULTS WITH THE OUTPUT OF EACH CLASSIFIER WITH VARIOUS SEGMENT LENGTHS

Section Length	Segment No.	Classifiers accuracy in %		
		KNN	SVM	Ensemble
10's	1	70.8	76.4	80.5
	2	71.6	78.2	82.2
	3	72.4	76.9	80.8
	4	72.6	77.7	82.7
	5	73.3	77.5	82.1
	6	71.8	76.3	82.1
20's	1	72.5	76.9	86.4
	2	72.8	76.5	86.1
	3	74.6	78.5	87.0
30's	1	73.5	77.6	89.3
	2	75.4	79.1	89.8
60's	1	74.8	79.1	93.1

Table 3.2 indicates that the ensemble classifier, with the highest accuracy for a full 60s signal, performed the best among classifiers. In this experimental scenario, Table 3.3 shows that KNN achieved accuracy of 74.8% for valence and 73.2% for arousal in the single modality (EEG). SVM displayed valence and arousal recognition rates of 79.1% and 76.9%, respectively. The ensemble technique outperformed both SVM and KNN, achieving accuracy rates of 93.1% for valence and 91.5% for arousal. When considering only video sequences, KNN and the ensemble had the same average accuracy values for both valence and arousal. However, when combining EEG and

video sequence modalities, the ensemble approach once again yielded the best results, with valence and arousal scores of 97.2% and 96.1%, respectively.

Table 3.3 Outcomes of all classifiers with different Modalities

Modality	Evaluation metrics	Valence			Arousal		
		KNN	SVM	Ensemble	KNN	SVM	Ensemble
EEG	Precision	0.78	0.74	0.92	0.73	0.73	0.89
	Recall	0.68	0.87	0.94	0.71	0.84	0.93
	*F*1	0.73	0.80	0.93	0.72	0.78	0.91
	Accuracy %	74.8	79.1	93.1	73.2	76.9	91.5
Video	Precision	0.97	0.91	0.97	0.95	0.85	0.93
	Recall	0.95	0.97	0.95	0.93	0.97	0.93
	*F*1	0.96	0.94	0.96	0.94	0.91	0.93
	Accuracy %	96.5	94.2	96.6	94.7	90.5	93.5
Combined (EEG & Video)	Precision	0.97	0.91	0.97	0.97	0.82	0.97
	Recall	0.95	0.97	0.96	0.94	0.97	0.96
	*F*1	0.96	0.93	0.97	0.95	0.90	0.96
	Accuracy %	96.9	93.6	97.2	95.2	89.5	96.1

The confusion matrix that were used to calculate the accuracy, recall, f1, and precision are also shown in Figure 3.10.

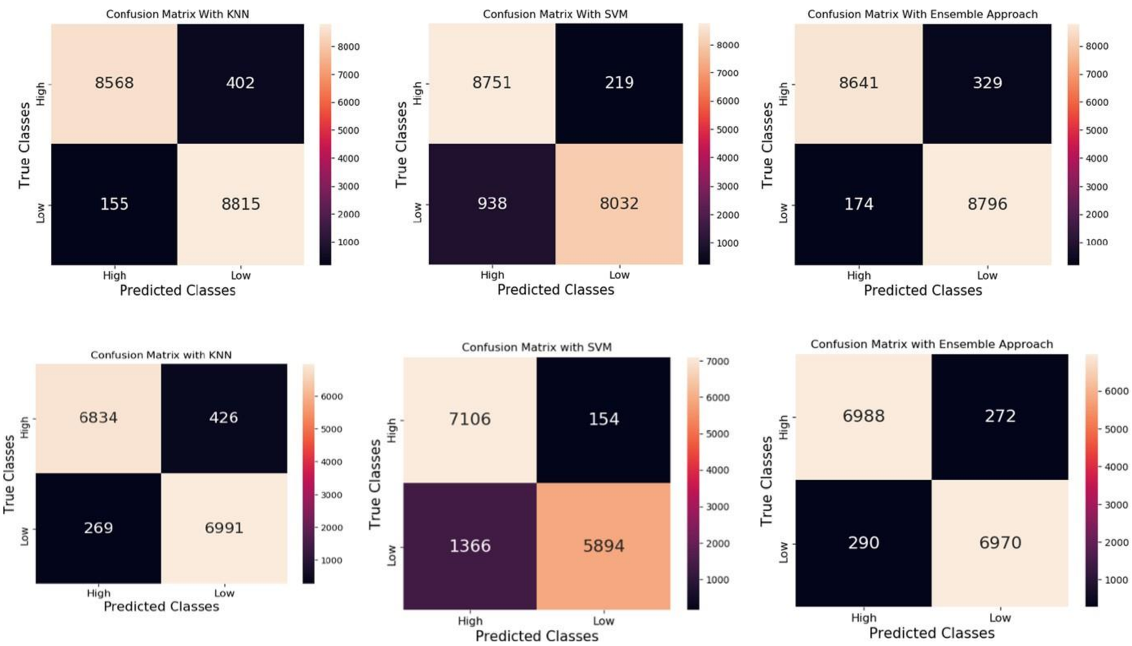


FIGURE 3.10 CONFUSION MATRES UNDER DIFFERENT EXPERIMENTAL SETTINGS; THE FIRST ROW REACTS TO CONFUSION MATRES WITH VALENCE CLASSIFICATION, AND THE NEXT ROW REACTS TO CONFUSION MATRES WITH AROUSAL CATEGORIZATION

The proposed framework, which is based on the ensemble approach, produced the highest accuracy levels across all of the experiments. This high accuracy is due to the ensemble approach's many decision-tree classifiers, which select the classes based on the voting process from all classifiers' predictions. More specifically, using maximum voting and training on many models boosted the ensemble's capacity for prediction. For arousal and valence, respectively, Figures 3.11 & 3.12 show graphically how all the classifiers compare in terms of the results. The accuracy levels achieved by the KNN, SVM, and ensemble classifiers for single modalities and multimodal systems are shown in these graphs. The accuracy of the multimodal system was shown to be higher than that of the two single modalities. For example, the ensemble approach's accuracy for valence and arousal using a single modality and EEG data is 93.1% and 91.5%, respectively. Similarly, if only video-based modality is taken into account, the results are 96.6% and 93.5%. However, the proposed architecture combines both video and EEG signal modalities, which enhances the model's performance. The multimodal system improved the accuracy for the valence and arousal classifications to 97.2% and 96.1%, respectively. Therefore, incorporating different modalities can significantly improve the performance of the model for emotion recognition in cross-subject analysis.

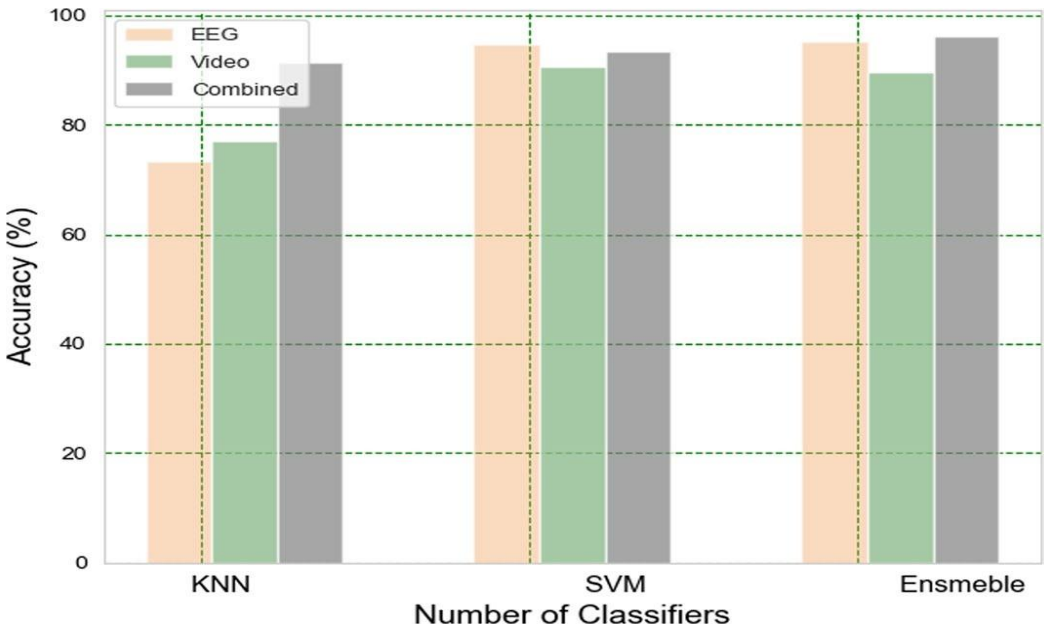


Figure 3.11 Performance evaluation in relation to valence categorization

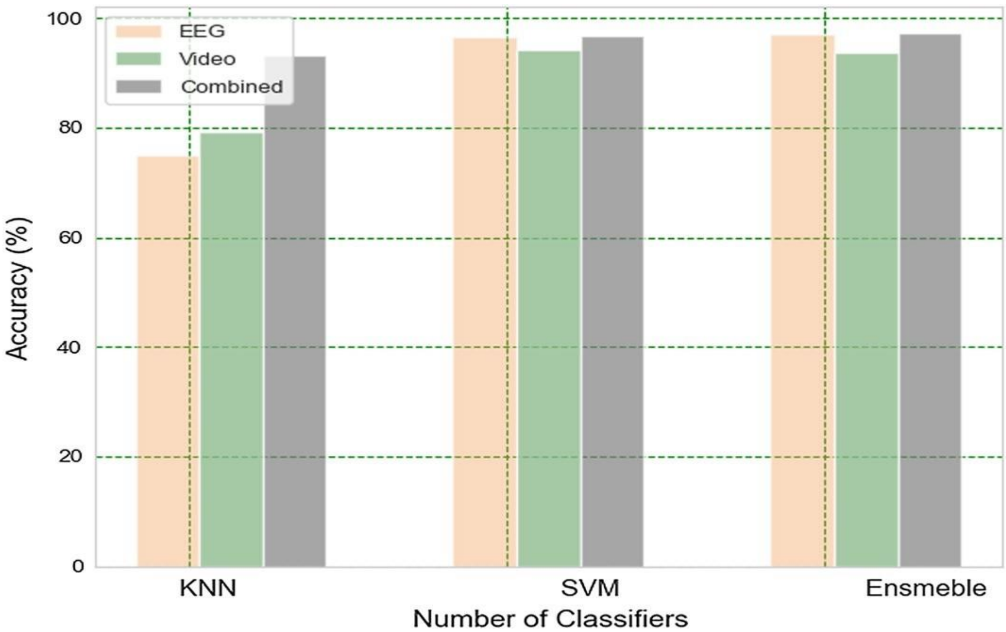


FIGURE 3.12 RELATIONSHIPS BETWEEN PERFORMANCE AND AROUSAL CATEGORIZATION

3.6.2 Comparison with present approaches

Table 3.4 provides a comparison of the suggested strategy to several currently in use techniques. We have offered a comparison with other studies that also takes EEG data into account because our main goal is to increase the accuracy of classifiers utilising EEG signals. They combined short-time Fourier transformations with electrode frequency distribution maps in their research to further examine EEG signals. The accuracy of their suggested method was 82.84%. In order to examine and choose the most effective set of features for emotion recognition, (Moin et al., 2023) selected features from the time, frequency, and wavelet domains. They classified emotions into nine different classifications using the SVM-based technique, and their accuracy rate was 65.92%. (Wang et al., 2020) used audio data to classify emotions into four groups with an accuracy rate of 78.11%. Additionally, (Cheng et al., 2021) created 2D frame segments by taking the spatial relationships of the channels into account. A deep forest classifier was then used to classify these sequences. The accuracy rates for this method are 97.69% for valence and 97.53% for arousal. It should be highlighted that, when compared to the subject-independent technique, the proposed approach had the highest accuracy. Even while some of the previously published work had higher accuracy, they only looked at one topic when evaluating it. We aggregated the data from all subjects in this study and reported accuracy on the combined subjects' data. We found that choosing the right channels and features could considerably aid in subject-independent emotion identification.

Table 3.4 Comparison with existing methodologies

Authors	Channels	Stimulus	Classifier	Emotions	Subjects	Accuracy (%)
Sharma et al.	32	Video	LSTM	4	32	84.02
Wang et al.	32	Video	CNN	3	14	84.83
Khateeb et al.	4	Videos	SVM	9	32	63.91
Bhatti et al.	1	Videos	MLP	4	30	78.12
Cheng et al.	32	Videos	Deep Forest	Val-ence Aro-usal	32	96.78 (Val_ance) 97. 54(Aro_usal)
Proposed EEG	4	Videos	Ensemble	Val-ence Aro-usal	22	92.81(Val_ance) 91.61(Aro_usal)
Proposed Video	---	Videos	Ensemble	Val-ence Aro-usal	22	96.63 (Val_ance) 94. 52(Aro_usal)
Proposed (EEG & Video Combined)	4	Videos	Ensemble	Val-ence Aro-usal	22	97.19 (Val_ance) 96. 11(Aro_usal)

3.6.3 Conclusion based on DEAP database

Cross-subject analysis of EEG cues for emotion detection are still difficult. In this study, we proposed a multimodal system that combines EEG signals with video data to provide an effective method for emotion identification. Pre-processing was performed on the data from the two separate modalities, after which the various features were collected, combined, and used as input by the 3 classifications based on machine learning. The EEG data yielded spectral and statistical characteristics, while the video sequences yielded features based on HOG and LBP. Additionally, combining facial expressions with EEG inputs may greatly enhance performance. The DEAP database was used to validate the proposed framework, and the ensemble-based approach outperformed the other three machine-learning classifiers. However, more advancements are needed, such as raising the number of recognized emotions, to create a real-time emotion identification system. A whole signal provides more useful information, and neglecting any time segment decreases accuracy, as we have demonstrated, In the future, we want to test deep learning on data from unseen participants.

3.7 Objectives of the proposed work:

The main goals/objectives of the proposed research are:

5. To study and analyse various existing technologies for virtual learning assessment model.
6. To prepare real time data-set using sensor based IOT and Machine learning algorithms.
7. To Design and Develop the proposed framework for Virtual Learning Assessment using Information Fusion.
8. To evaluate the proposed framework with the standard metrics.

3.7.1 Summarised objective first and second

A conference paper was published with titled as "Real time data evaluation with wearable devices:

Smart watch technologies are transforming the ecosystem of transmission and surveillance for investors and study participants who want to give real-time information to be assessed. An Effect of Artefact Calibration Technique on Emotion Classification. Smart watches have a variety of sensors that can track physical activity and locations. Real-time surveillance of physical and maybe emotional growth is made possible by the integration of all of these components, which enables the data to be delivered from the data collection device to a distant computer. A simple and affordable optical measurement technique called photoplethysmography is frequently employed for heart rate monitoring. PPG is a non-invasive device that measures the volumetric fluctuations in blood flow by using a light beam and a photo detector at the top layer of skin. Approaches for HRV (Heart Rate Variability) assessment are being researched in a number of fields, including the identification of human emotions (HER). Photoplethysmographic (PPG) data are commonly assessed for this assessment, hence smart-watch as sensor-based gadget play a crucial role. However, because these signals are susceptible to a number of factors, including motion blur, light sources, pressure distribution, racial background, and environmental factors, their quality (in terms of additional disruptions) might not always be ideal. In this case, artefact correction techniques are important and have an effect on the result. In order to increase the effectiveness of sentiment recognition and categorization using PPG signals during auditory stimuli and an SVM classifier, this study suggests a novel data distortion reduction method. The presented scheme offers an enhanced characterization in trigger sensation, i.e., 68.75 percept, as compared to data which was previously conducted utilising a traditional toolset, i.e., 48.81.

PPG could be used in collaboration with other indication, such as electroencephalographic action, to make even more progress.

3.7.2 PPG Sensing flow

The Support Vector Machine (SVM) procedure was employed to determine if stimuli were present or absent by taking as input parameters any values that a baseline t-test determined to be statistical significance.

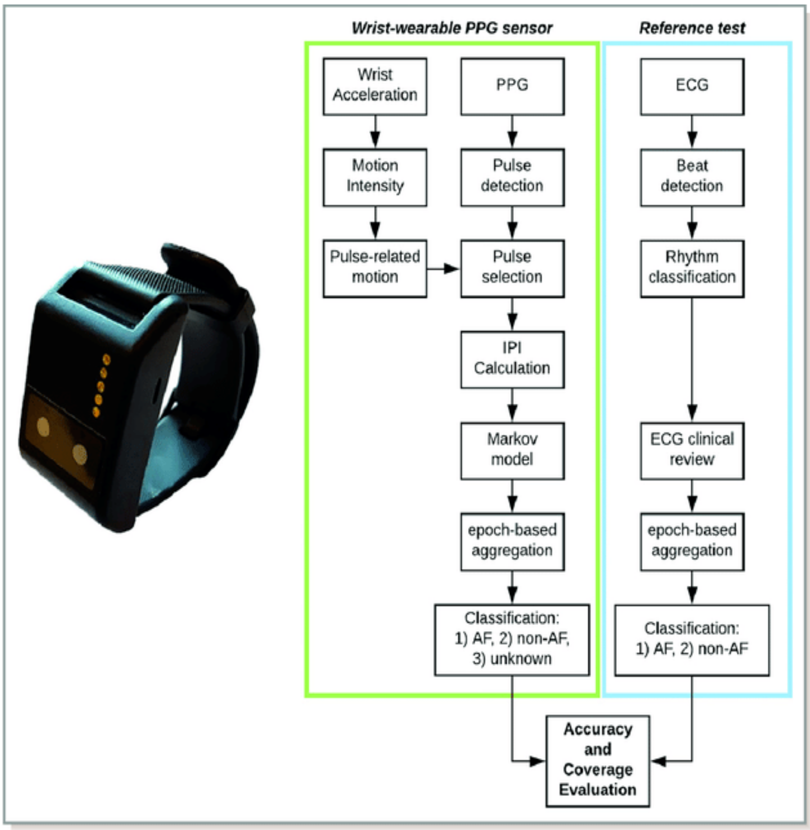


FIGURE 3.13 PPG SENSING FLOW

With the aim of identifying an auditory stimulus, an investigation has used a forearm, multimodal smart-watch (Empatica E4) to record a variety of physiological indicators from which features are derived. The procedure of the intended operation is shown in Figure 3.13.

3.7.3 Motivation and related work

Novel wearable technology frameworks consist of sensors are gaining traction. Due to the varied, mixed, and unpredictable networks, which have a various applications prospect in the age of modern sensor-based intelligent machines, and inconsistencies between service offering and application demand. In this regard, researchers think that, in addition to existing solutions, smart sensor-based gadgets can provide an unanticipated source of innovation for overall systems such as human emotion identification systems.

3.7.4 Protocol for testing

May refer to Figur-3.14 and Figure-3.15 (Fayaz et al., 2021). In order to accomplish this, two male and five female healthy individuals with ages ranging from 36 to 18 (average absolute variation) and BMIs between 22.7 and 2.1 kg/m2 participated in the experiments. The initial calibration time for PPG data is approximately 15 seconds, as seen, after which the results from this time period are disregarded for further analysis. Here, three stimuli of varying intensities are displayed.

- ✓ Pleasant sample of approx. 810
- ✓ Natural stimulus of approx. 715
- ✓ Unpleasant sample of approx. 300

FIGURE 3.14 EMPATICA E4 MEASUREMENT SETUP

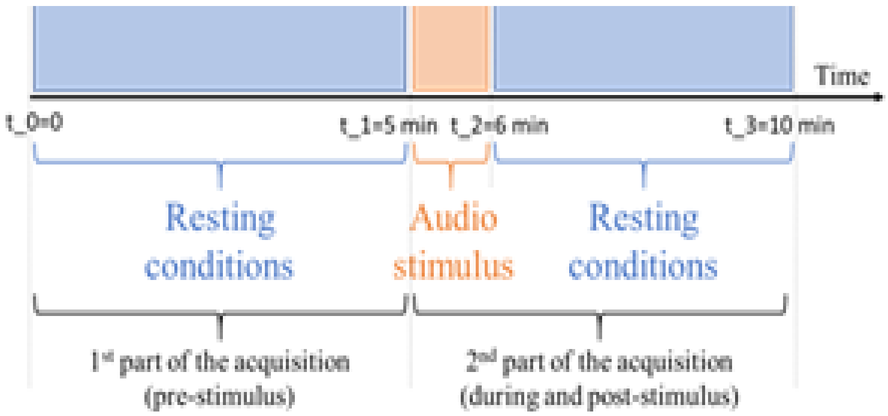


FIGURE 3.15 TWO-PART HRV ANALYSIS

3.7.5 Method and result

The estimation method provided inside the Kubios tool and a novel approach put forth by the scientists independently were both assessed in the latest study to remove spikes which would produce improper frequency components. (Lipponen & Tarvainen, 2019) comprises of the phases seen in Figure 3.16 and results in Figure-3.17 & Figure-3.18.

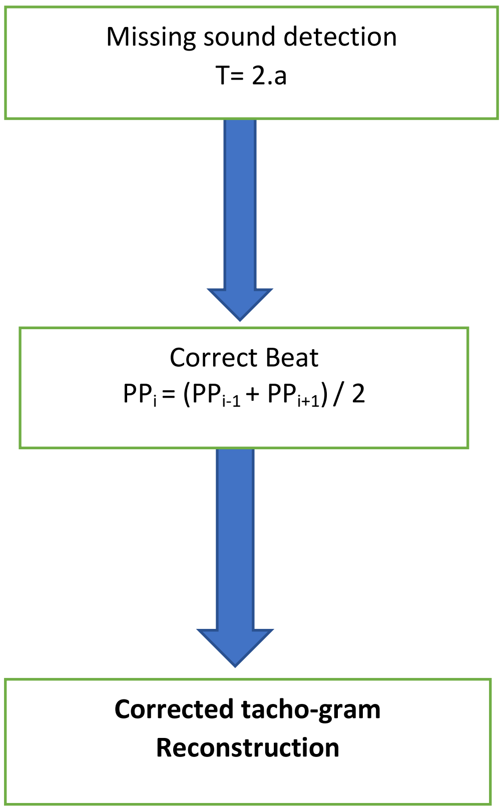


FIGURE 3.16 CORRECTION TECHO-GRAM FLOW

Results

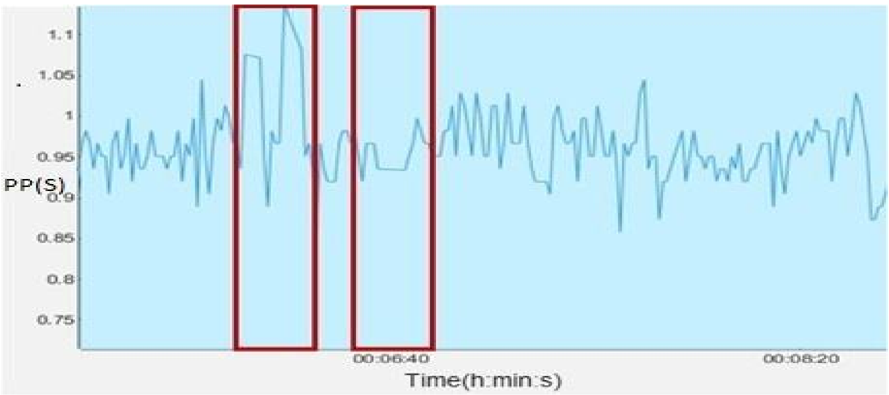


FIGURE 3.17 ***SHOW PROPOSED METHODS TACHOGRAMS COMPARISON***

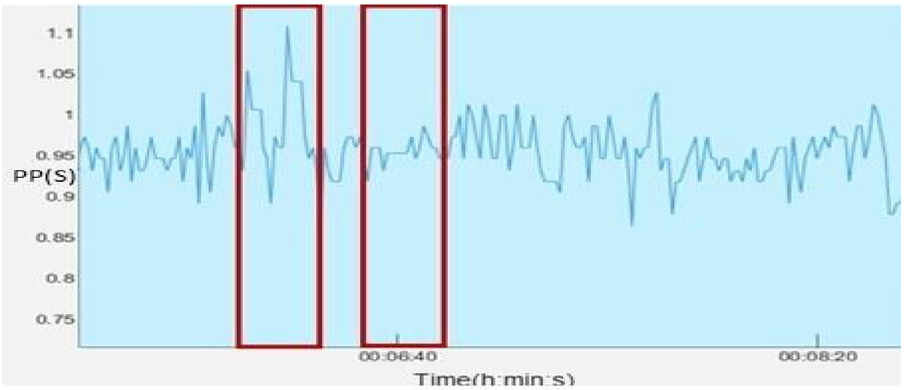


FIGURE 3.18 ***SHOWS PROPOSED METHODS TACHOGRAMS COMPARISON***

3.7.6 Proposed methodology for achievement of the objectives:

To develop a novel emotion recognition framework for virtual learning assessment

Figures 3.19 to 3.22 with following technologies are to be considered:

❖ Information fusion which is the technique of integrating set of data sources to obtain consistent, useful, and accurate information.

❖ Novel Emotion Recognition

- EEG
- Data pre-processing
- Discovering powerful interface for emotion detection.
- Unlabelled emotion data annotation layer
- Expression data

- The Internet of Things refers to objects that are individually recognizable and their virtual representations as constructs in the World.
- Deep learning a powerful AI feature that imitates human brain activity to process real time data that can be used in emotion recognition, speech recognition, language translation, object detection and decision taking. Without human oversight it can learn.

3.8 Objectives third and fourth

In continuation with objective number third and fourth the published results with graphs are depicted below.

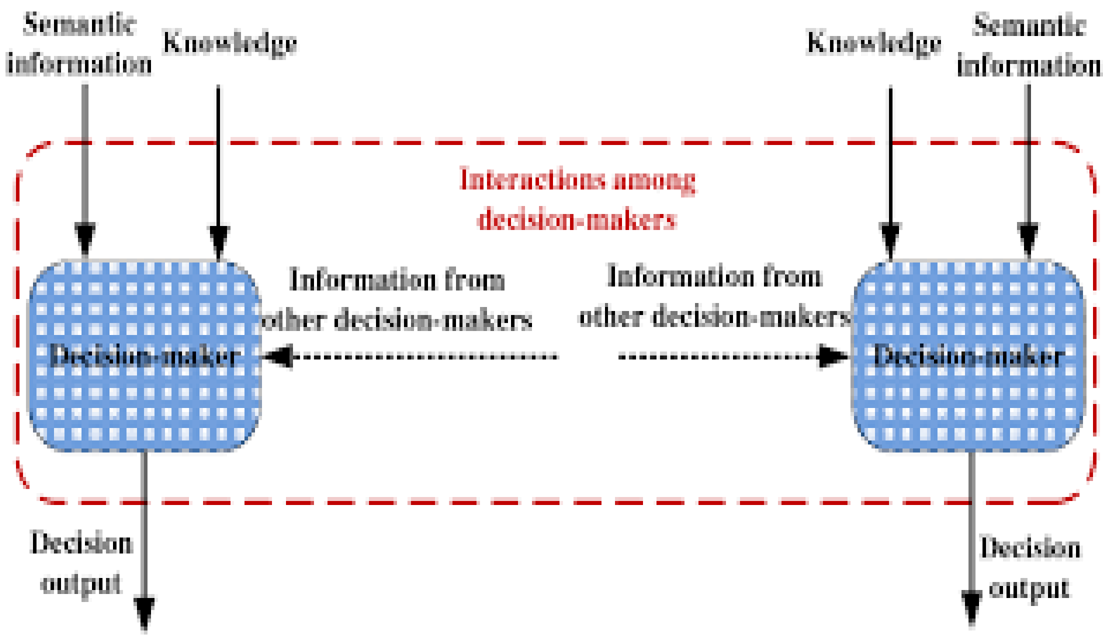


FIGURE 3.19 PROTOTYPE OF INTELLIGENT SENSING INTELLIGENCE SYSTEM(ISI) (PARK ET AL., 2019)

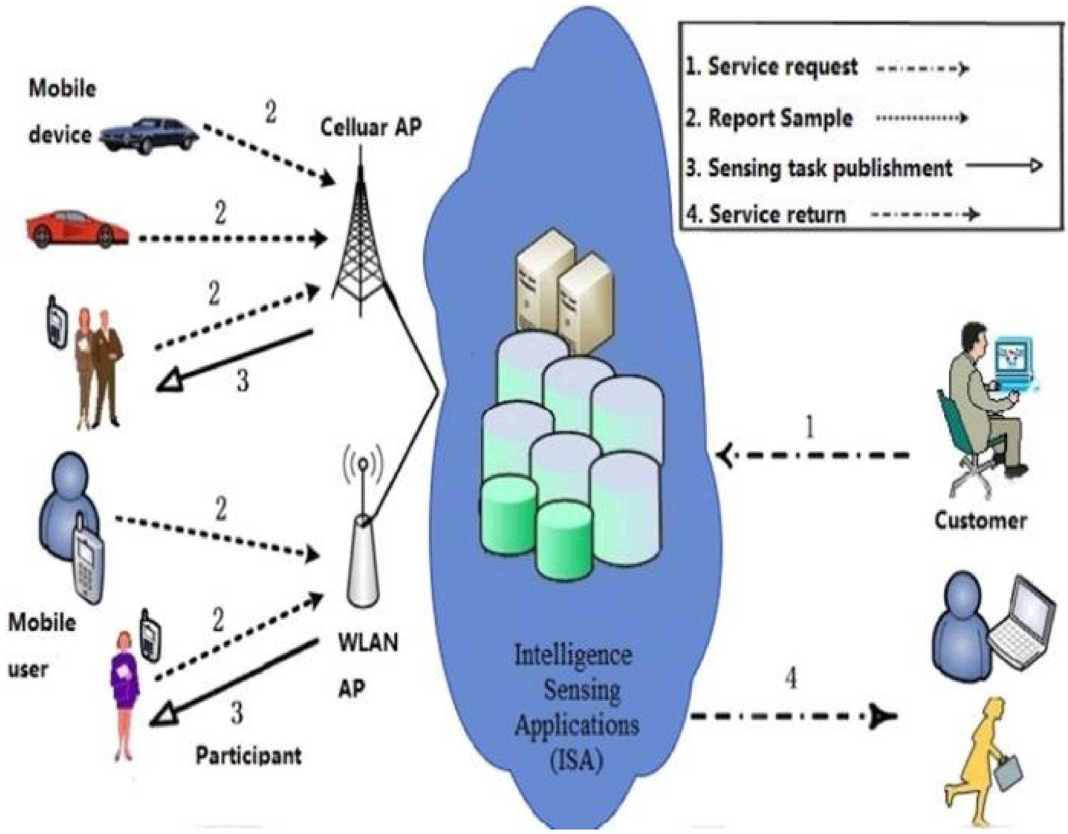


FIGURE 3.20 PROTOTYPE INCORPORATING ISA (FERNÁNDEZ-CABALLERO ET AL., 2016)

3.8.1 Graphs with observations

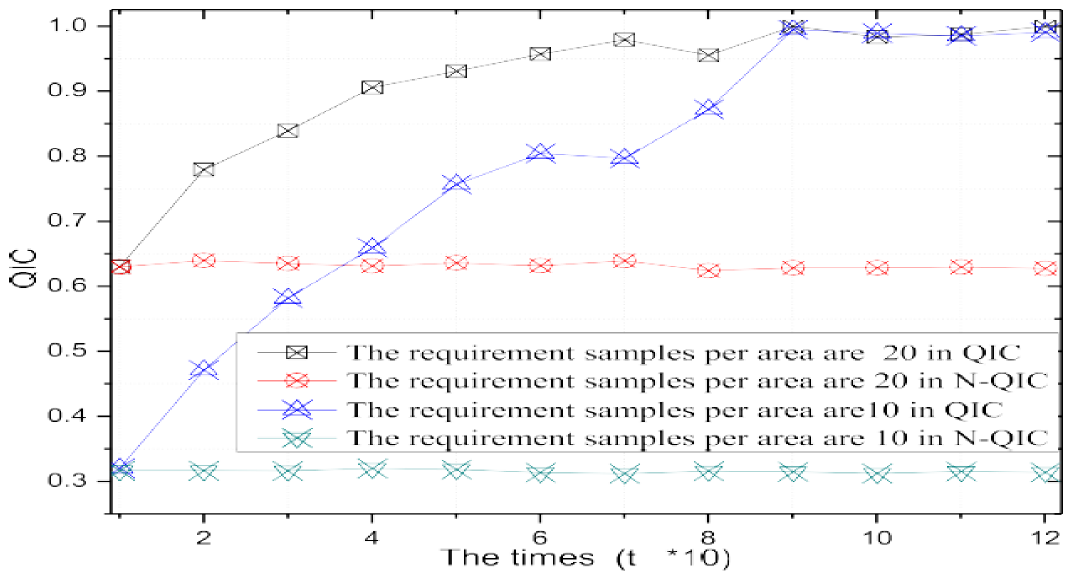


FIGURE 3.21 GRAPHICAL REPRESENTATION OF QIC SOURCE (THOITS, PEGGY, 2011)

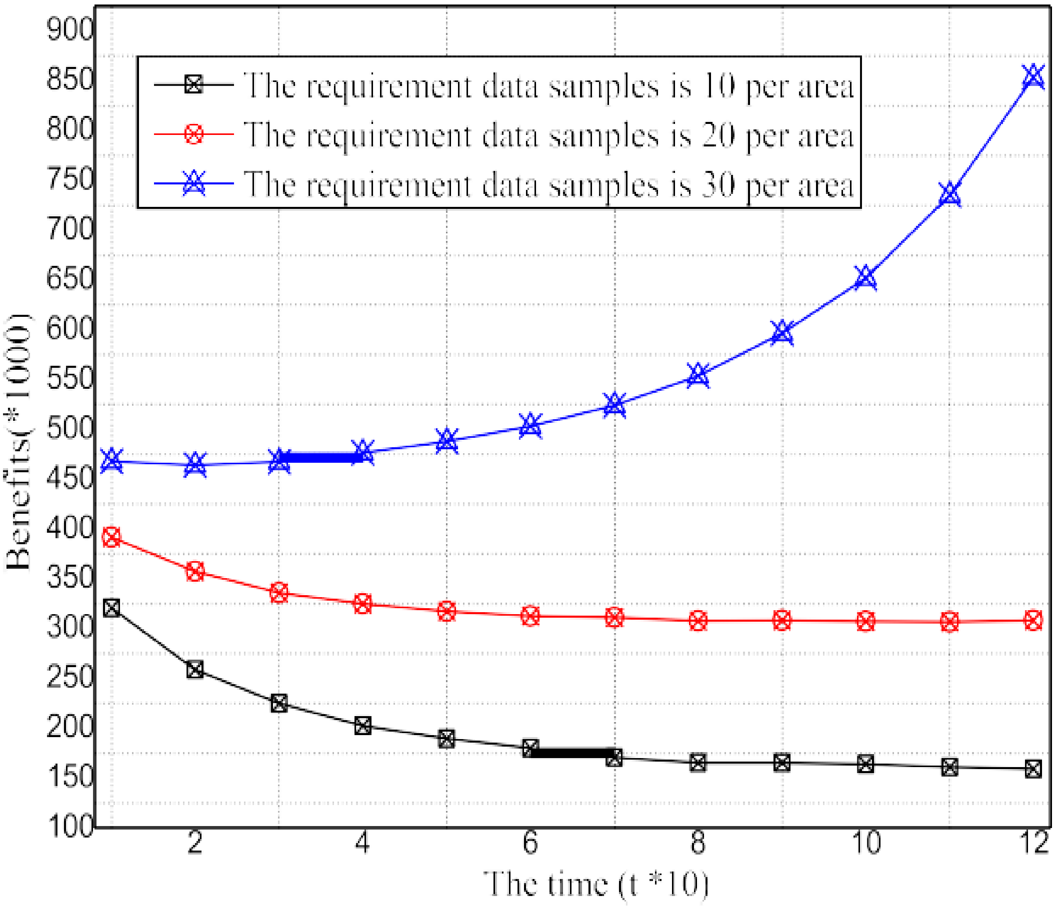


FIGURE 3.22 GRAPH VARIATIONS WITH TIME (T) OF ISA

3.8.2 Potential area of potential work

The requirements for CIoT data samples are established to maximise financial gains through market mechanisms. On the basis of ISA's disclosed prices, reporters then develop reporting strategies to maximise payoffs. Via games that incorporate ISA As a result, we point out that where reporters do not enforce privacy security features, they will adjust payouts to maximise the effectiveness of a structure and sample data. They'll have difficulty. Reporters will therefore anticipate our potential employment. In order to accelerate the development of the Internet, privacy regarding this Content.

3.9 Neural system design bottlenecks in machine-human mapping:

As the abstract depicts the comprehensive knowledge of human intellect is still continuing its development, i.e., humans and evidence assurance is not yet seamlessly aligned. Designers will be able to design human tailored Cognitive Information

Systems (CIS) by comprehending the cognition process depicted in Figure 3.23. The need for this study is justified because today's corporate Decision-Makers (DM) are confronted with problems that they cannot answer in the allotted time without the use of cognitive information systems. The researchers' goal is to show the response choices to increase the efficiency of Human-Computer Interaction (HCI), which leads to a better cognitive information system strengthening with a greater cognitive threshold by demonstrating the resilience of cognitive resonant frequency and the role of info-communication via HCI, such as connectivity, relation, and influences. The practical research approach comprises research analyses and a review of existing articles to pursue a comparative study patterns; after that, a model development paradigm was employed to observe and supervise the progress of a CIS during HCI. Our study range gives a broader perspective of how various disciplines influence HCI and how the human cognitive model is reinforced to enrich the addition. We uncovered a significant gap in the current literature about mental processing produced by the vast spectrum of theory and practice".

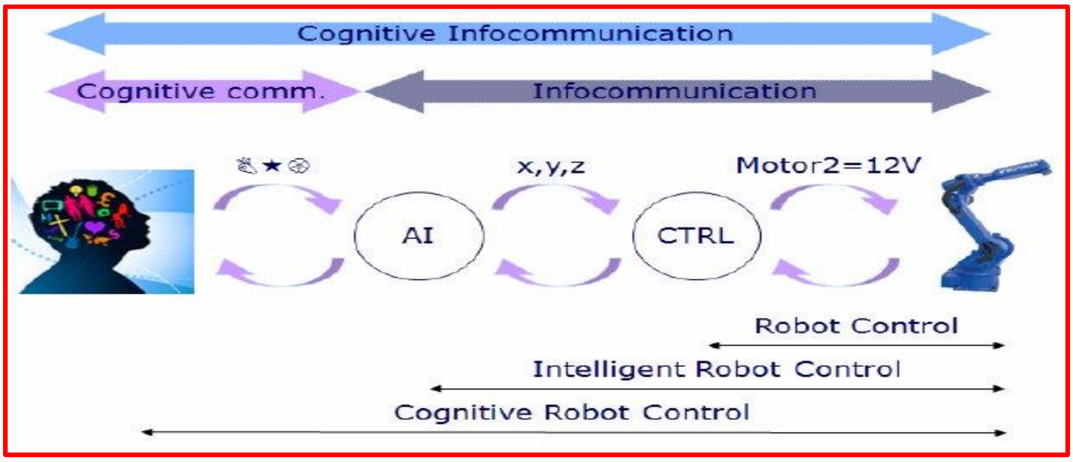


FIGURE 3.23 COGNITIVE INFO COMMUNICATION

There are three types of social cognitive processes: Observation, evaluation, and management.

- Observation: Aspects of the compliance that are perceptual or representational of Neuro-cognitive factors include observing face and verbal information, an impression of bodily movement, and other elements.
- Evaluation: Emotion Detection (ED), emotional and cognitive sensitivity and

other similar processes are examples of assessment or interpretation processes.

- Management: Management is a set of regulatory activities that include behaviours, self-awareness, observation, cognitive control, and assessment, among other activities.

The complexity of ANN could be even more significant than the complexity of the brain, which has approx.1011 neurons and approx. 1015 trillion interconnections. It all depends on who is designing the network(Gold & Shadlen, 2007). There is no limit to what may be accomplished with an ANN-supported CIS that could more correctly predict and support humans while also replacing human decision-making and improving cognition(Ferentinos, 2018). The artificial neuron has a structure that is quite close to that of a biological neuron in terms of function, as shown in the Figure 3.24 to Figure 3.26.

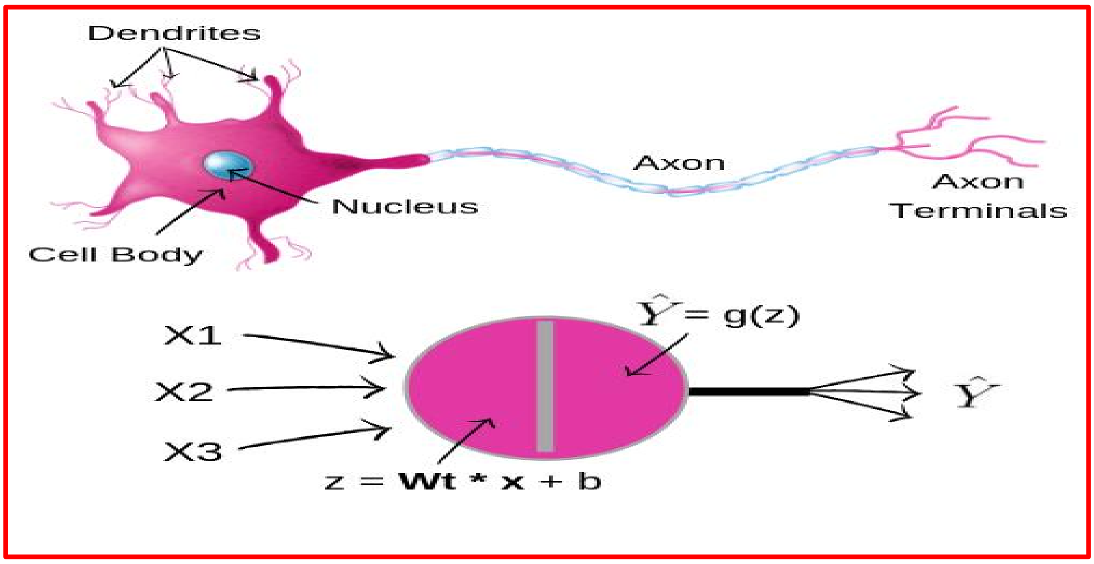


FIGURE 3.24 ARTIFICIAL NEURON STRUCTURE

Dendrites such as weight w1 plus bias b1, which originate from synaptic vesicles x1 and move to postnatal neuron y1, constitute the cell's inputs. The neutron's entries as an axon, and the cell body (nucleus) as f(x) serves as the kernel function among them.

$$Y1 = \sum_{i=0}^{n}(w1i \times x1i) + b \rightarrow Y1 = f(\sum_{i=0}^{n}(w1i \times x1i) + b1)$$

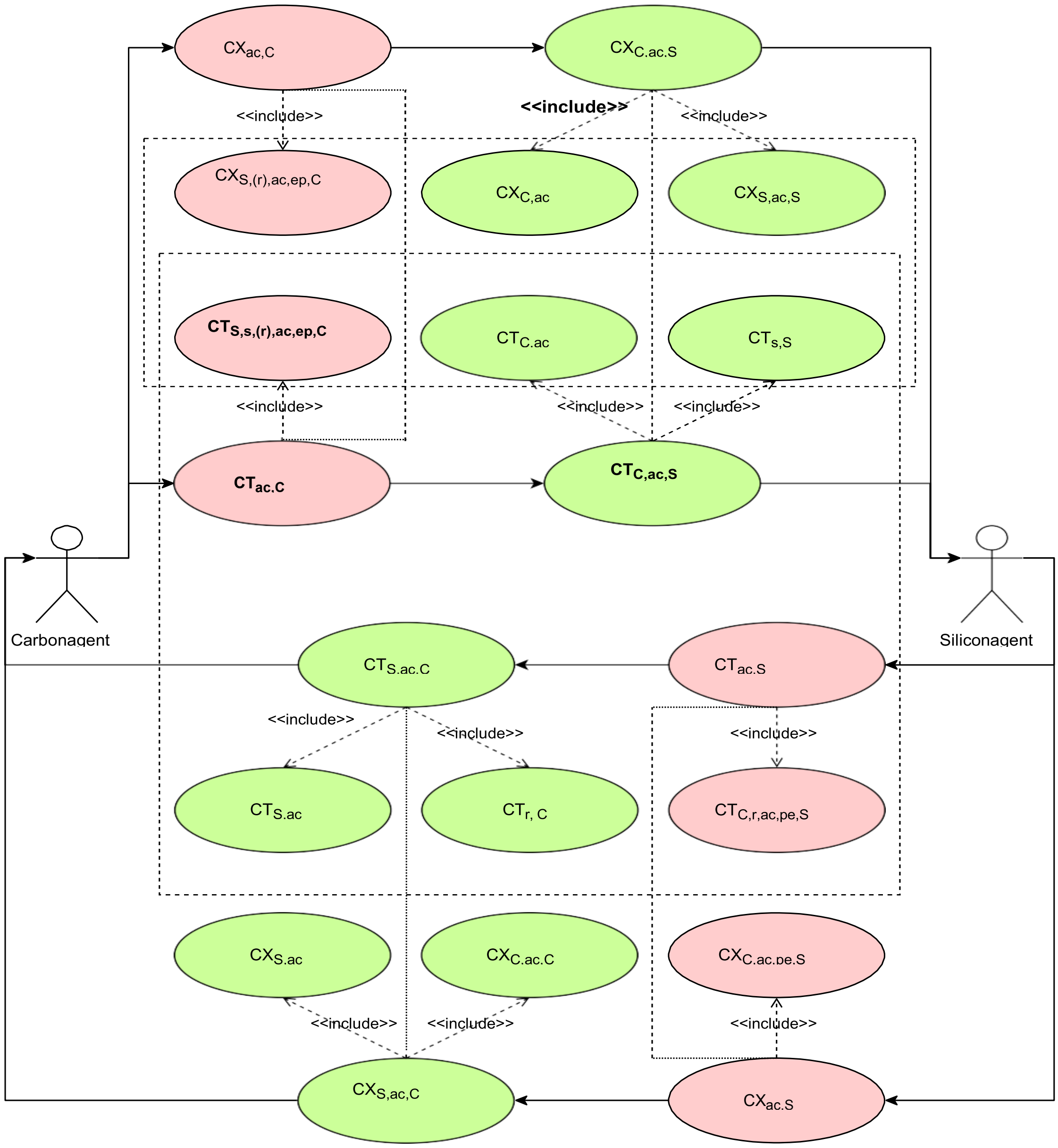


FIGURE 3.25 ARTIFICIAL NEURAL NETWORK

The fundamental formula shows ANN solution with or without activation function, which can be viewed as an instance of a basic ANN approach to be applied in this situation, based on neural network(Wolfert et al., 2022). In this scenario, the input and output have a linear relationship, which could be used as a multiple regression model if the activation function is not linear. ANN can thus be adapted to nonlinear data using the neurons nonlinear function.

In humans, nonlinear synchronization mechanisms are required for information flow across cortical structures during the conscious brain(Hagarty & Morgan, 2020). Intelligent technology algorithms focused on humans can enhance cognitive dispensation in CIS stages.

Nevertheless, the result of Y1 is derived from f (neutron). Based on the objective of the neural network, many classes of ANNs are employed to represent it in HCI.

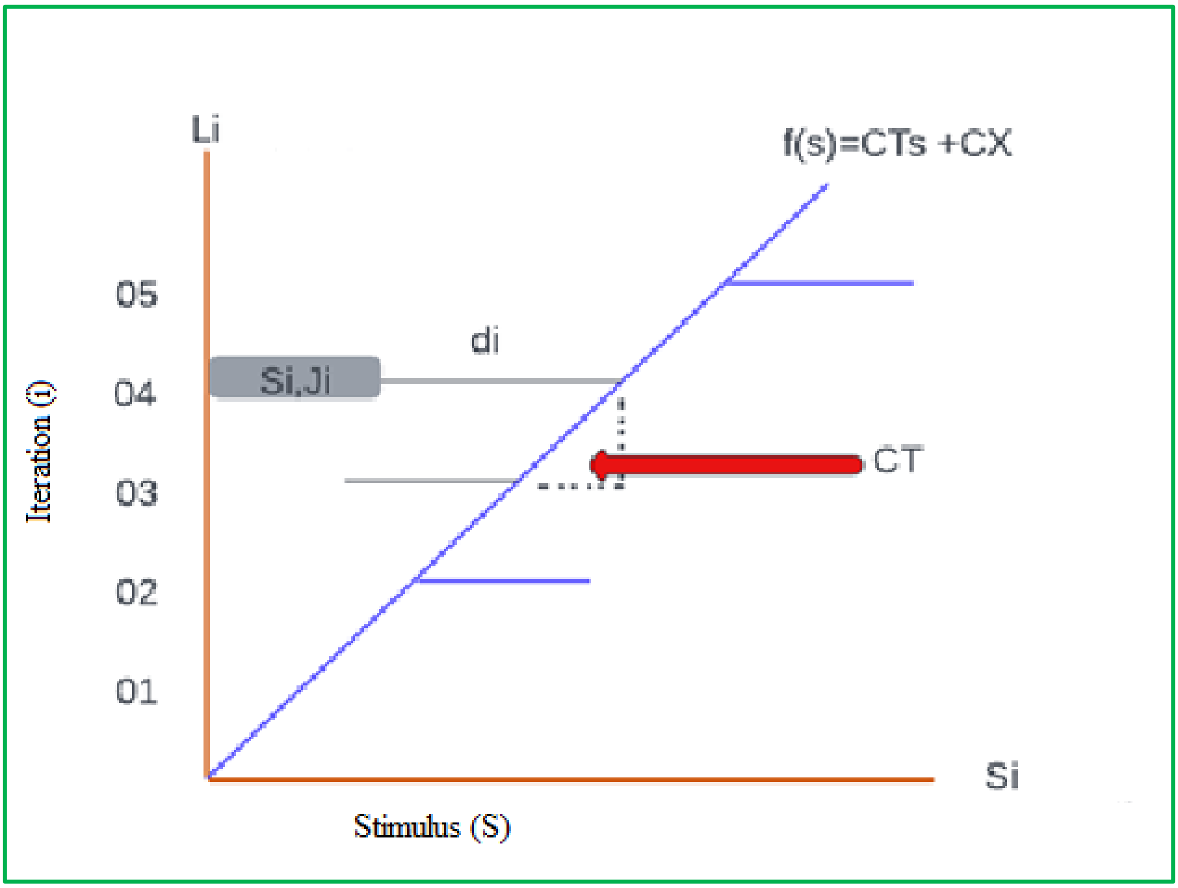


FIGURE 3.26 COGNITIVE RESONANCE DISTANCE

3.9.1 Info-Transportations and Cognitive Info-Communication

With the recognition of themes those can promote thinking prototype among carbon and silicon agents, an effective data system should be placed between both to facilitate appropriate information communication link among the human aspect and the prototype parts within the computing structure is required(Wright & Liley, 1996). Data communication is critical in various settings, including commerce and the industry environment as shown in Figure 3.27.

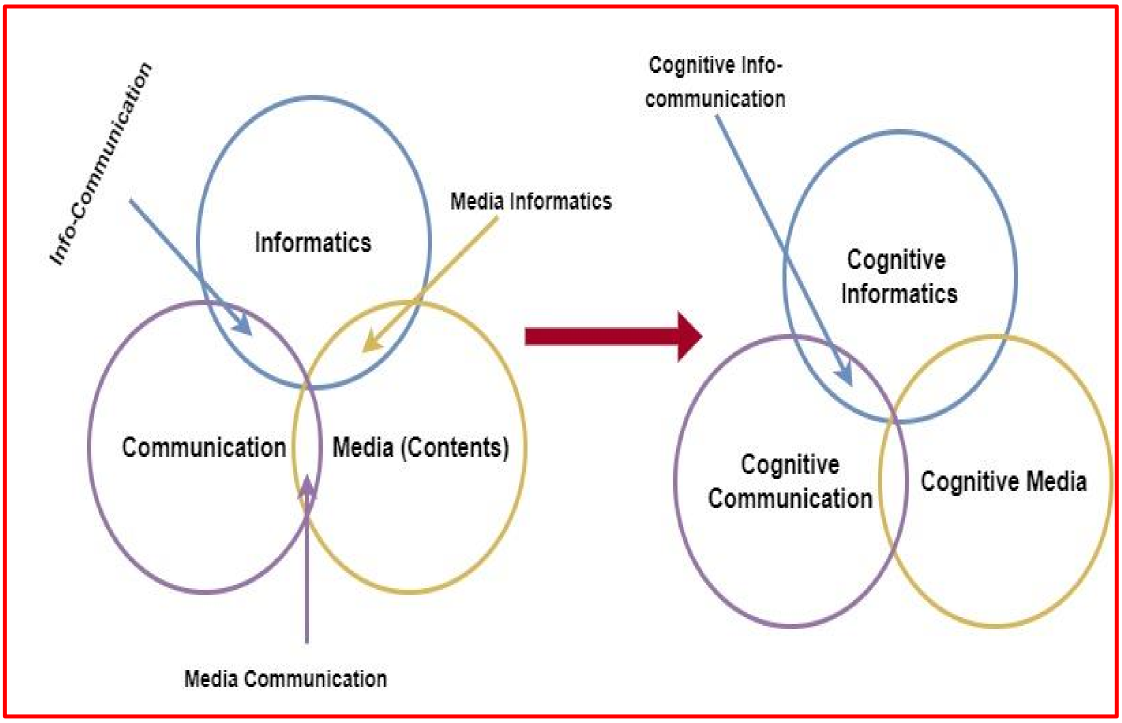


FIGURE 3.27 INFO-COMMUNICATION AND COGNITIVE INFO-COMMUNICATION

From a corporate decision standpoint, data in an analytical manner must be examined, comprehended, and organized as per corporate requirement(Ko & Fujita, 2013). The information is created at three tiers, each used to build the expertise, which serves as the foundation for Business Analytics (BA). BA is a set of tools relying on cutting-edge automation technologies that may be used to gain a competitive advantage in various businesses. Various apps, best - practices, techniques, and other resources are examples of tools. Financial, data information, and awareness of marketplaces, for example, is all part of the market. At different stages of DM, business analytics has a positive impact which is: a decision based on short-term evaluation, Strategic choices, and choosing a methodology.

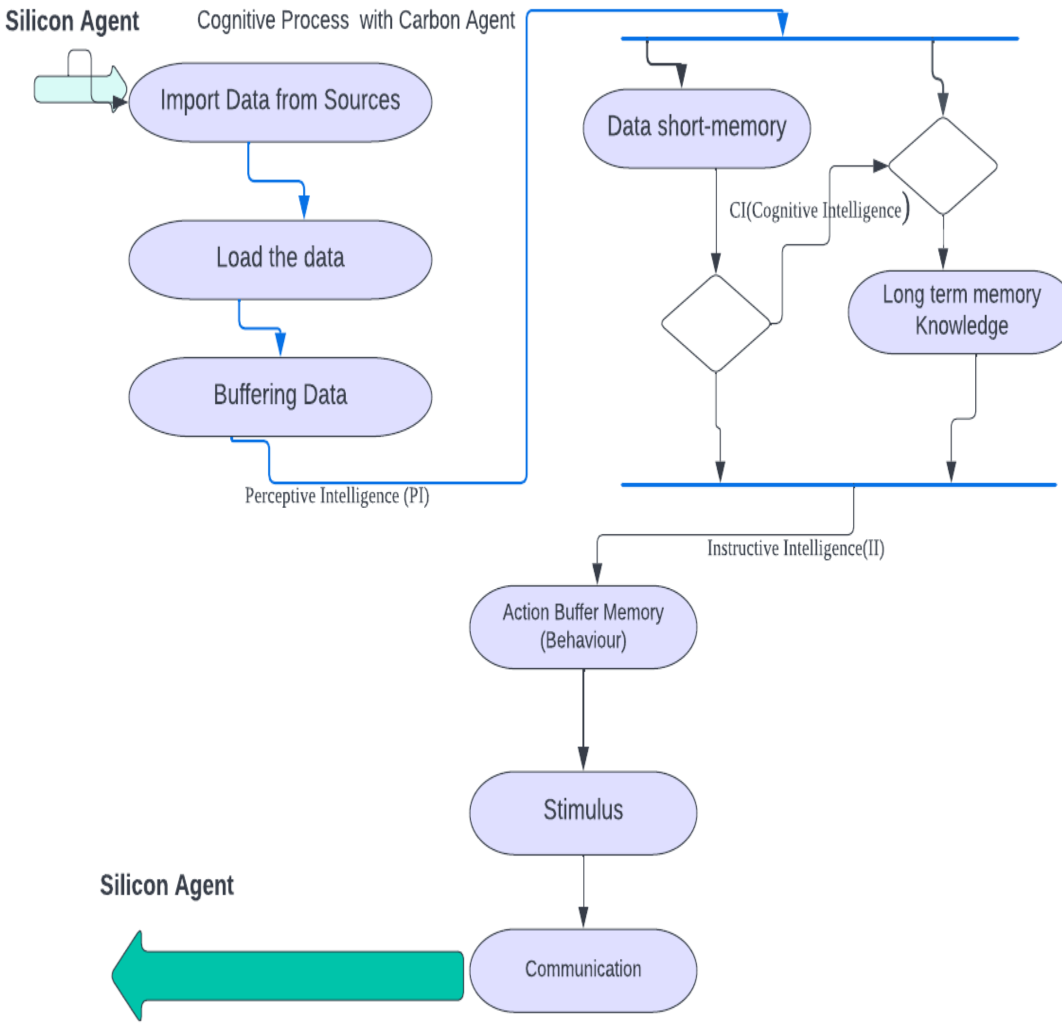


FIGURE 3.28 AN ACTIVITY DIAGRAM SHOWING THE STRUCTURE FOR BOTH VERBAL AND NONVERBAL CARBON-SILICON AGENTS.

The efficiency of management choices at 3 managerial levels could be measured using mixed information, like context, as detailed underneath in the CS info-communication paradigm(Tomasello et al., 2005). The level of information communications has a direct influence on the quality of decision-making and its repercussions. Silicon agents can be used to mimic, analyse, and replace managerial actions which affect an organization. The basic concept of cognitive resonance could be used to develop modelling techniques that can facilitate carbon agent sensing(Molnar & Mattyasovszky-Philipp, 2019). One of the possibilities could be to propose a framework that can allow the creation of a personalized user interface and appropriate material for each carbon agent. You'll need to develop customized material utilizing various technologies to enhance administrative understanding and strategy formulation and judgments, thus aiming for a high level of customization, which is equivalent to hyper-personalization.

The framework has been changed via modifications to become a versatile and adaptable platform that can meet the needs of different levels of leadership. Adjustment and adaptability are essential qualities in light of the passage of time and energy, the increased amount of available knowledge, and the complexities of the options. Data science, machine learning, and artificial intelligence are used for creating data, visualisations, and an administrative environment for decision preparedness and making choices(Zhu et al., 2022). As a result, the purpose is to promote the algorithm's interpretations, openness, interpretability, and ease of understanding. As depicted in figure above, the cognitive info-communication mechanism, which includes internal information, interpretation and the variables from Figure-3.28 from carbon-silicon info-communication framework, are mathematical theoretical proofs incorporated in the model.

3.9.2 Theme Discussions

Research activities and real extrapolation assessments are examples of mathematical techniques that can be applied to cognitive resonator data given in frameworks based on knowledge transfer among the carbon and silicon actuators. In order to develop an RN or ANN, a dataset must be built using the above parameters prior to introducing cognitive rich to silicon and carbon components for lateral construction. By the use of this ML technique, a thinking data model framework may be constructed that uses an underlying bio. foundation to analyse, optimize, enhance, and replace human cognition in a manner by using the silicon agent with Intelligent Process Automation, which is linked to computer vision, AI, and BI based Robotic process automation (RPA) (Thoits, Peggy, 2011). Various frameworks provide differing degrees of information about intellect and DM during the investigation. This, in our opinion, is a focus area that may be utilized to derive new findings and, based on that, to conduct additional research to guarantee that each framework includes not just sections of the flow but, the full process. Identical frameworks to be improved onwards; however, they are currently considered basic models. The basic prototypes can be extended to acquire a deeper understanding of a certain area in the long run. A professional choice or a tricky decision is too conveyed to the DM process. Different variables can be included in the design via a quantitative formula to enhance and improve the accuracy of the basic models. Leveraging tailored material and context, and reasoning, the silicon agent, can help with decision-making(Patel et al., 2018). Before making a decision, a complete in and out environment and affecting threat evaluator is required. Machine intelligence and business intelligence systems have already examined data to make practical judgments. Many projections and scenarios based on big data are reviewed, calculated, and then developed to help the company to make the result-oriented decision practically possible for their financial circumstances. Such techniques are among the most recently studied cognitive information management tools.

3.10 Chapter Conclusion

Throughout the study, various models provided diverse degrees of detail about human thought and decision-making. This, in our opinion, is a primary concern that may be utilized to derive new findings and, depending on those findings, to do additional studies to guarantee that all designs encompass not just sections of the flow, but the full procedure. Certain designs can be improved in future; though, they are presently regarded as a base design. In the long run, the base designs can be prolonged to acquire a deeper understanding of a precise sub-area. While making a decision, a complete macro and micro environmental and affecting risk assessment is essential. Knowledge that has previously been examined by AI and BI systems is required for strategic decisions(Wang, 2007). Numerous models based on big data are developed to aid companies in optimal decision-making within their financial contexts, representing advanced cognitive information system solutions. This study delves into the creation of multidisciplinary frameworks for physiological and human-based perceptual system engineering. These frameworks, integral to human-computer interaction and emotion recognition, heavily incorporate cognitive resonance. Mental knowledge plays a pivotal role in disseminating cognitive resonance, contributing to the advancement of models that depict the intricacies of the human mind and its activities, guiding future research endeavours.

CHAPTER-4

Results & Discussions

4.1 Introduction

This section studies the spontaneous human emotion detection from input signals. The hybrid modal emotion recognition approach in this chapter is based on multi firmness analysis of acoustic & filmic signals. Multi-resolution analysis (MRA) refers to analyzing the signal in different resolution. The features are extracted by MRA using DWT and classification is done using multiple classifiers i.e. SVM, MLP and K mean classifier. The system based on MRA has been developed in this thesis by extending the previous work in emotion recognition using MRA [Verma, G. K. et al., 2011c]. This chapter includes emotion recognition from audio and visual modality.

Here we summarize the different approaches reported in literature for human emotion detection. Further this chapter describes the human sentiment recognition based on audio, visual signal and some of its key components on a development set of eNTERFACE [Martin, O. et al., 2006] and RML [Wang & Guan, 2008] multimodal emotion database. Multimodal fusion is then studied for emotion recognition system. Experimental results are finally given on eNTERFACE and RML database independently for audio and video cue, as well as their combination using feature level fusion.

4.2 Spontaneous Affect Recognition

Within Human Computer Interaction (HCI), Affect recognition has been an emerging research area because the various emotional states like cognition, embarrassment and depression can be considered complex emotional states and expressed via dozens of possible facial expressions(Fayaz, 2021). The complex affective states may not be accurately expressed by a single signal; therefore, researcher used multimodal cues to detect emotion [Gunes H. et al., 2011].

Different methodologies have been conveyed in the study in order to extract features from audio and visual cue(Fayaz et al., 2023). They can broadly be divided into 1) dimension-based approach, in which the emotion is represented in 2D or 3D space in terms of valence, arousal and dominance. [Nicolaou, M. A. et al., 2011], [Schuller B., 2011], [Koelstra S., et al., 2012] 2) Categorical approaches, in which the emotion is categories into six basic emotions namely, Happy, Sad, Surprise, Anger and Disgust as proposed by Ekman [Ekman P., 1999]. and 3) Appraisal based approach, in which we deal the emotion which are generated through continuous evaluation of internal states as well as the external state of the world [Gunes H. et al.,2011].

4.3 Databases

Recent advances in human emotion recognition have prompted many researchers to create emotion database. Some of the emotion databases are MIT [Healey J. A., et al., 2005], MMI [Pantic, M. et. al., 2005], HUMAINE [Douglas-Cowie, et al., 2007], VAM [Grimm, M. et al., 2008], SEMAINE [McKeown, G. et al.], MAHNOB-HCI [Soleymani, M. et al., 2012a], and DEAP [Koelstra, S. et. al., 2012]. DEAP database is being used in this study. These databases contain speech, visual or audio-visual and physiological emotion data. In this work, we have used eNTERFACE and RML audio-visual database. The facial expression of different emotions for eNTERFACE and RML database are shown in Figure 4.1 and the database content for both databases are summarize in Table 4.1.

4.4 The eNTERFACE Database

The eNTERFACE"05 is a benchmark emotion database contains audio and video samples. Despite requesting confession, the database comprises 6 fundamental sentiment expression varieties to be recorded for evaluation. The voice samples being collected by as many as 42 subjects (34 male & 8 female).

FIGURE 4.1 FACIAL EXPRESSIONS FOR DIFFERENT EMOTIONS FIRST ROW - ENTERFACE DATABASE, SECOND ROW- RML DATABASE

Table 4.1 Database content summary of eNTERFACE and RML database

eNTERFACE	
Database type	Audiovisual
No. of subjects	44
Language	English
# Emotion	6 Universal emotions
#Video	1320
RML	
Database type	Audiovisual
No. of subjects	8
# Emotion	6 Universal emotions
#Video	720
Image frame size	720*480
Frame rate	30
Audio sampling rate	22050

4.5 RML Database

RML emotion database is also an audio-visual emotion database contains six basic universal emotions. The videos recorded from eight subjects in different languages. Each subject was asked to act naturally while reading emotional sentences.

4.5.1 Emotion recognition system based on Audio cue

Emotion recognition from audio cue is based on acoustic characteristics of the audio signal. There are several techniques such as Mel Frequency Cepstral Coefficient (MFCC), LPC etc. to capture the acoustic characteristics. MFCC is a state-of-the-art technique which is used in this thesis. Different speech information is represented by different speech features like emotion, speaker audio and both in overlapped manner. They find out recognition rate of 89.20% for speaker dependent and 48.18% for speaker independent case.

Pitch is the most extensively used prosodic feature for sentiment acknowledgment due to its high discriminating power than other features. Besides pitch log energy [Kammoun & Ellouze, 2006] is one most considered parameter to examine speaking styles and emotions. In review, a number of researchers (Cummings and Clements, 1995 and Ling et al., 2005) worked on excitation source information features for developing emotion recognition system so we have not reviewed the excitation source features here and they are not considered in our research. For feature extraction, we have used prosodic and spectral feature of an audio signal.

4.5.2 Experiments

Experiments were carried out into two emotion databases i.e. eNTERFACE and RMLdatabase.

4.5.3 System DevelopmentAudio Features

The initial configurations for the experiments are as

follows:MFCC 12

Level of decomposition 5

Coefficients : detail coefficient at each level and approximation at 5thlevel

Feature vector dimension 32

The results for audio modality are shown in Table 4.2 & 4.3 and the confusion matrices for audio cues are shown in Table 4.4 and 4.5.

Table 4.2: Classification accuracy of Audio cue over eNTERFACE and RML database

		A	D	F	H	S	Sp
eNTERFACE DATABASE	MLP	62.5	20.0	40.0	62.5	60.0	67.5
	SVM	60.0	37.5	35.0	72.5	45.0	40.0
RML DATABASE	MLP	74.17	62.50	46.67	55.83	70.83	68.33
	SVM	67.50	65.83	40	57.50	77.50	68.33

Table 4.3: Audio cue results for eNTERFACE and RML database

		SVM			MLP		
		Accuracy	Recollection	F-measures	Precision	Recollection	F-measures
ENTERFACE database	A	0.499	0.61	0.549	0.635	0.635	0.635
	D	0.358	0.38	0.365	0.358	0.21	0.255
	F	0.328	0.45	0.334	0.51	0.41	0.45
	H	0.645	0.744	0.684	0.59	0.635	0.63
	S	0.545	0.551	0.594	0.45	0.61	0.512
	SP	0.593	0.4	0.478	0.551	0.675	0.607
RML Database	A	0.681	0.675	0.678	0.712	0.742	0.727
	D	0.594	0.658	0.625	0.591	0.625	0.607
	F	0.485	0.4	0.438	0.505	0.467	0.485
	H	0.556	0.575	0.566	0.545	0.558	0.551
	S	0.699	0.775	0.735	0.726	0.708	0.717
	Sp	0.732	0.683	0.707	0.701	0.683	0.692

Table 4.4: Confusion metrics for Audio based on a) MLP b) SVM (RML Database)

	A	D	F	H	S	Sp
A	**74.17**	3.33	0.83	6.67	0.00	15.00
D	1.67	**62.50**	15.00	15.83	4.17	0.83
F	9.17	16.67	**46.67**	7.50	12.50	7.50
H	5.00	16.67	6.67	**55.83**	10.00	5.83
S	0.83	5.83	12.50	10.00	**70.83**	0.00
Sp	13.33	0.83	10.83	6.67	0.00	**68.33**

(a)

	A	D	F	H	S	Sp
A	**67.50**	3.33	5.00	5.83	0.00	18.33
D	0.83	**65.83**	10.83	15.00	6.67	0.83
F	9.17	18.33	**40.00**	9.17	20.83	2.50
H	5.83	17.50	10.83	**57.50**	5.00	3.33
S	0.00	5.00	9.17	8.33	**77.50**	0.00
Sp	15.83	0.83	6.67	7.50	0.83	**68.33**

(b)

Table 4.5: Confusion metrics for Audio based on a) MLP b) SVM (eNTERFACE Database)

	A.	D	F	H	S	Sp
A	**97.6**	0.1	0	0.1	1.25	1.25
D	2.5	**77.5**	8.75	3.75	5	2.5
F	1.25	1.25	**91.25**	0	3.75	2.5
H	0	1.25	1.25	**95**	1.25	1.25
S	2.5	2.5	0	2.5	**88.75**	3.75
Sp	3.75	1.25	1.25	2.5	10	**81.25**

(a)

	A.	D.	F.	H.	S.	Sp.
A.	**95**	0	0.1	0	5	0
D.	0	**97.5**	0	0	0	2.5
F	0	7.75	**92.5**	0	0	0
H	0	0	0	**100**	0	0
S	0	2.5	0	0	**97.5**	0
Sp	2.5	6.25	2.5	0	0	**88.75**

(b)

4.6 Emotion recognition system based on Visual cue

A face contains the following components; Lips, nose, eyes, eyebrows, chin, cheeks play a key role to produce facial expressions or appearance. The Facial expression can be determined by single or combinations of these face components. Ekman categorize the muscular activity of above face components and named 'action units' (AUs), See Tables 4.6 to 4.9.

4.6.1 Experiments

Experiments have been carried out into two emotion databases i.e.

eNTERFACE andRML database.

4.6.1.1 System Development

Visual Feature

The initial configurations for the experiments are as

follows:Function : DB4

Level of decomposition 5

Coefficients : detail coefficient at each level and

approximation at 5th level

Feature vector dimension 32

Table 4.6 depicts the outcome aimed at visual modality.

Table 4.6: Video cue results for eNTERFACE and RML database

		SVM			MLP		
		Accuracy	**Recollection**	**F-measures**	**Precision**	**Recollection**	**F-measures**
ENTERFACE database	**A**	0.5	0.589	0.547	0.836	0.813	0.8
	D	0.5	0.54	0.61	0.78	0.3	0.8
	F	0.54	0.62	0.5	0.751	0.715	0.72
	H	0.74	0.67	0.72	0.754	0.938	0.912
	S	0.43	0.558	0.71	0.78	0.81	0.791
	Sp	0.536	0.375	0.441	0.75	0.75	0.75
RML database	**A**	0.531	0.65	0.584	0.793	0.8	0.797
	D	0.621	0.683	0.651	0.785	0.792	0.788
	F	0.657	0.592	0.623	0.819	0.792	0.805
	H	0.678	0.667	0.672	0.805	0.825	0.815
	S	0.653	0.55	0.597	0.769	0.75	0.759
	Sp	0.719	0.683	0.701	0.828	0.842	0.835

Table 4.7: Video cue results for ENTERFACE and RML database

		A	D	F	H	S	Sp
eNTERFACE DATABASE	**MLP**	0.756	0.819	0.825	0.85	0.838	0.75
	SVM	0.625	0.744	0.631	0.706	0.694	0.65
RML DATABASE	**MLP**	0.8	0.792	0.792	0.825	0.75	0.842
	SVM	0.65	0.683	0.592	0.667	0.55	0.683

Table 4.8: Confusion metrics for Video based on a) MLP b) SVM (RML Database)

	A	D	F	H	S	Sp
A	**81.00**	2.50	2.50	5.00	1.67	8.33
D	4.17	**79.18**	7.50	4.17	4.17	0.83
F	5.00	5.83	**79.18**	0.00	5.83	4.17
H	5.00	6.67	1.67	**82.51**	3.33	0.83
S	3.33	5.00	3.33	10.00	**74.00**	3.33
Sp	3.33	1.67	2.50	0.83	7.50	**85.17**

(a)

	A	D	F	H	S	Sp
A	**64.01**	7.50	10.00	7.50	4.17	5.83
D	13.33	**68.33**	5.83	4.17	5.00	3.33
F	14.17	11.67	**59.27**	4.17	3.33	7.50
H	10.00	6.67	4.17	**65.67**	9.17	3.33
S	10.83	10.83	3.33	13.33	**55.20**	6.67
Sp	9.17	5.00	7.50	2.50	7.50	**68.41**

(b)

Table 4.9: Confusion metrics for Video based on a) MLP b) SVM (RML Database)

	A	D	F	H	S	Sp
A	**73.34**	7.5	7.5	4.375	3.125	3.125
D	3.75	**88.74**	3.125	1.25	1.25	1.875
F	4.375	5.625	**81**	0.625	4.375	5
H	3.75	1.875	1.125	**85.75**	0	4.375
S	3.125	1.125	5	1.125	**85.25**	3.125
Sp	5.625	3.75	3.75	3.75	3.75	**79.375**

(a)

	A	D	F	H	S	Sp
A	**64.75**	4.375	8.75	6.875	7.5	8.75
D	15.625	**65.75**	4.375	5.625	3.75	1.875
F	13.75	5.625	**62.25**	1.875	8.125	9.375
H	11.25	11.875	1.25	**71.00**	0.625	5.00
S	10.00	4.375	13.1258	1.125	**69.4**	3.125
Sp	12.5	3.750	3.750	3.750	3.750	**62.51**

(b)

4.7 Experiments based on proposed fusion framework of audio-visual cues

Audio and visual cues are used for these experiments. The experiments are performed for both databases i.e. RML and eNTERFACE database. For RML database, total 1440 audio and video samples have been used. Similarly, for eNTERFACE database 2640 audio and video samples used to predict emotion. The audio and video features are combined together and deed to feature selection system to select most discriminating features. We used Fisher Discriminate Analysis (FDA) to select top 50 features. A brief description of fisher's approach is described under section 4.5.2.

A. = Anger

D. = Disgust

F. = Fear

H. = Happy

S. = Sadness

SP. = Surprise

Table 4.10: Results based on single and combined cues for eNTERFACE and RMLdatabases
(A, D, F, H, S, Sp)

			SVM			MLP		
			Accuracy	Recollection	F-measures	Precision	Recollection	F-measures
ENTERFACE DATABASE	Audio	A	0.97	0.98	0.964	0.93	0.98	0.942
		D	0.852	0.972	0.933	0.95	0.77	0.84
		F	0.9342	0.9451	0.943	0.89	0.913	0.901
		H	1	1	1	0.916	0.95	0.933
		S	0.951	0.975	0.963	0.807	0.888	0.845
		Sp	0.973	0.888	0.928	0.878	0.813	0.844
	Video	A	0.495	0.588	0.537	0.826	0.713	0.765
		D	0.564	0.663	0.609	0.774	0.813	0.793
		F	0.533	0.6	0.565	0.75	0.75	0.75
		H	0.73	0.675	0.701	0.788	0.838	0.812
		S	0.606	0.538	0.57	0.78	0.8	0.79
		Sp	0.536	0.375	0.441	0.75	0.75	0.75
	Fusion	A	0.786	0.528	0.805	0.89	0.913	0.901
		D	0.71	0.825	0.763	0.87	0.838	0.854
		F	0.818	0.9	0.857	0.9	0.9	0.9
		H	0.823	0.813	0.818	0.89	0.913	0.901
		S	0.836	0.763	0.797	0.869	0913	0.89
		Sp	0.873	0.688	0.769	0.88	0.825	0.852
RML DATABASE	Audio	A	0.681	0.675	0.678	0.712	0.742	0.727
		D	0.594	0.658	0.625	0.591	0.625	0.607
		F	0.485	0.4	0.438	0.505	0.467	0.485
		H	0.556	0.575	0.566	0.545	0.558	0.551
		S	0.699	0.775	0.735	0.726	0.708	0.717
		Sp	0.732	0.683	0.707	0.701	0.683	0.692
	Video	A	0.531	0.65	0.584	0.793	0.8	0.797
		D	0.621	0.683	0.651	0.785	0.792	0.788
		F	0.657	0.592	0.623	0.819	0.792	0.805
		H	0.678	0.667	0.672	0.805	0.825	0.815
		S	0.653	0.55	0.597	0.769	0.75	0.759
		Sp	0.719	0.683	0.701	0.828	0.842	0.835
	Fusion	A	0.789	0.808	0.798	0.798	0.792	0.795
		D	0.721	0.817	0.766	0.807	0.8	0.803
		F	0.695	0.683	0.689	0.715	0.733	0.724
		H	0.795	0.742	0.767	0.791	0.758	0.774
		S	0.839	0.825	0.832	0.851	0.858	0.855
		Sp	0.858	0.808	0.943	0.813	0.833	0.823

Table 4.11: Multiple cue results for ENTERFACE and RML database

			A	D	F	H	S	Sp
ENTERFACE DATABASE	MLP	Audio	62.5	20.0	40.0	62.5	60.0	67.5
		Video	75.6	81.9	82.5	85.0	83.8	75.0
		Fusion	91.25	83.75	90	91.25	91.25	82.5
	SVM	Audio	60.0	37.5	35.0	72.5	45.0	40.0
		Video	62.5	74.4	63.1	70.6	69.4	65.0
		Fusion	79.17	80.0	73.33	75.83	85.83	83.33
RML DATABASE	MLP	Audio	74.17	62.50	46.67	55.83	70.83	68.33
		Video	80.0	79.17	79.17	82.50	75.0	84.17
		Fusion	79.17	80.0	73.33	75.83	85.83	83.33
	SVM	Audio	67.50	65.83	40.0	57.50	77.50	68.33
		Video	65.0	68.33	59.17	66.67	55.0	68.33
		Fusion	80.83	81.67	68.33	74.17	82.50	80.83

4.7.1 Results

Results in Table 4.10 & 4.11 notify the performance results for audio, video and fusion (feature level), audio-video cues using MLP and SVM classifier for eNTERFACE and RML database. The results are given for MLP and SVM classifier. The system based on video cue outperforms the audio cue. The combined system achieved good performance rather than single audio/video cue based system. Confusion matrices are shown in Tables 4.12 & 4.13 with multiple cue results for RML database in Figure 4.2.

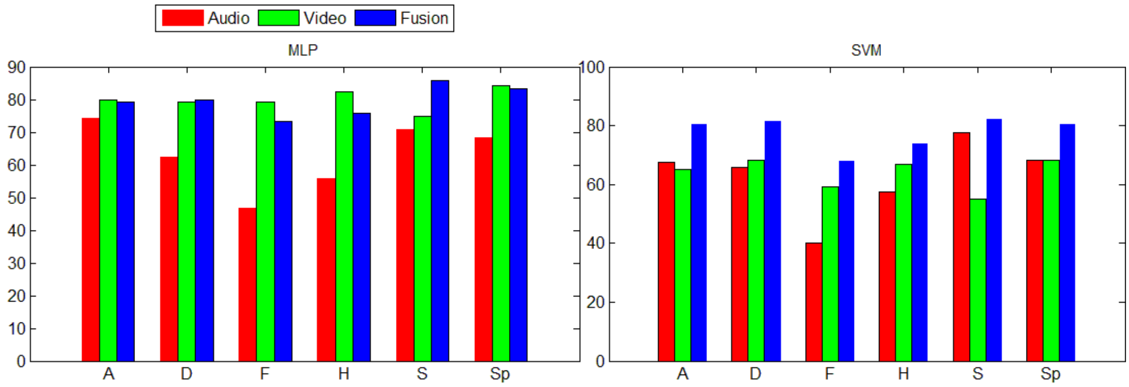


FIGURE 4.2 MULTIPLE CUE RESULTS FOR RML DATABASE

Table 4.12: Confusion metrics for Fusion of audio-video based on a) MLP b) SVM (RML Database)

	A	D	F	H	S	Sp
A	**79.17**	1.67	2.50	4.17	0.00	12.50
D	1.67	**80.00**	10.00	5.83	2.50	0.00
F	5.00	9.17	**73.33**	1.67	7.50	3.33
H	5.00	6.67	5.00	**75.83**	4.17	3.33
S	0.83	1.67	5.83	5.83	**85.83**	0.00
Sp	7.50	0.00	5.83	2.50	0.83	**83.33**

(a)

	A	D	F	H	S	Sp
A	**80.83**	1.67	3.33	5.00	0.00	9.17
D	1.67	**81.67**	9.17	4.17	3.33	0.00
F	4.17	15.00	**68.33**	3.33	7.50	1.67
H	4.17	10.00	4.17	**74.17**	5.00	2.50
S	0.00	4.17	7.50	5.83	**82.50**	0.00
Sp	11.67	0.83	5.83	0.83	0.00	**80.83**

(b)

Table 4.13: Confusion metrics for Audio based on a) MLP b) SVM (eNTERFACEDatabase)

	A	D	F	H	S	Sp
A	**91.25**	2.5	1.25	2.5	1.25	1.25
D	3.75	**83.75**	2.5	2.5	2.5	5
F	1.25	2.5	**90**	0	3.75	2.5
H	2.5	3.75	0	**91.25**	1.25	1.25
S	0	0	3.75	3.75	**91.25**	1.25
Sp	3.75	3.75	2.5	2.5	5	**82.5**

(a)

	A	D	F	H	S	Sp
A	**82.5**	5	2.5	6.25	2.5	1.25
D	5	**82.5**	2.5	3.75	3.75	2.5
F	1.25	3.75	**90**	0	2.5	2.5
H	6.25	6.25	1.25	**81.25**	3.75	1.25
S	3.75	8.75	5	3.75	**76.25**	2.5
Sp	6.25	10	8.75	3.75	2.5	**68.75**

(b)

4.7.2 Comparisons with related research

Accuracy comparison with other systems based on eNTERFACE databases are given in Table 4.12. It is observed that the average accuracy of our system (based on MLP) is 88.3%, which is much more compared to the performance of other systems as shownin Table 4.14.

TABLE 4.14 ACCURACY COMPARISON WITH OTHER SYSTEMS BASED ON eNTERFACE

Systems	**Year**	**Features**	**Accuracy (%)**
Schuller B. et al.	2009	Arousal- valance	80.2 for SVM & 80.5 for HMM/GMM
Datcu D. et al.	2009	Facial features	85 through SVM
Mansoorizadeh M. et al.	2010	Asynchronous Feature level fusion	84
Paleari M. et al.	2010	Neural network	71
Wang Y. et al.		Modalfactor analysis	72.4
Our study	2023	wavelet analysis	88.3

4.7.3 Fusion in Hybrid Sentiment Recognition Framework

To progress the precision and resilience of sentiment identification, the hybrid emotional recognition paradigm makes use of information fusion approaches. The result entails integrating information from several modalities or sources, like voice tones, facial expressions, and physiological signs, to create a thorough picture of a person's emotional state. This structure seeks to address the drawbacks of single-modal methods by utilising a variety of data varieties.

Figure-4.3 depicts the key aspects of the framework include feature extraction, where relevant emotional features are derived from each modality, and classification, where machine learning algorithms analyse the fused features to identify the emotional state accurately. Information fusion plays a pivotal role in addressing challenges like ambiguity and variability in emotional expressions.

This approach acknowledges that emotions are multi-dimensional and can be better captured through a holistic analysis of various cues. The integration of information from different modalities not only improves accuracy but also provides a more nuanced and context-aware illustration of emotions. Overall, the hybrid emotion recognition framework showcases the power of information fusion in creating a more comprehensive and effective system for understanding human emotions.

Proposed model

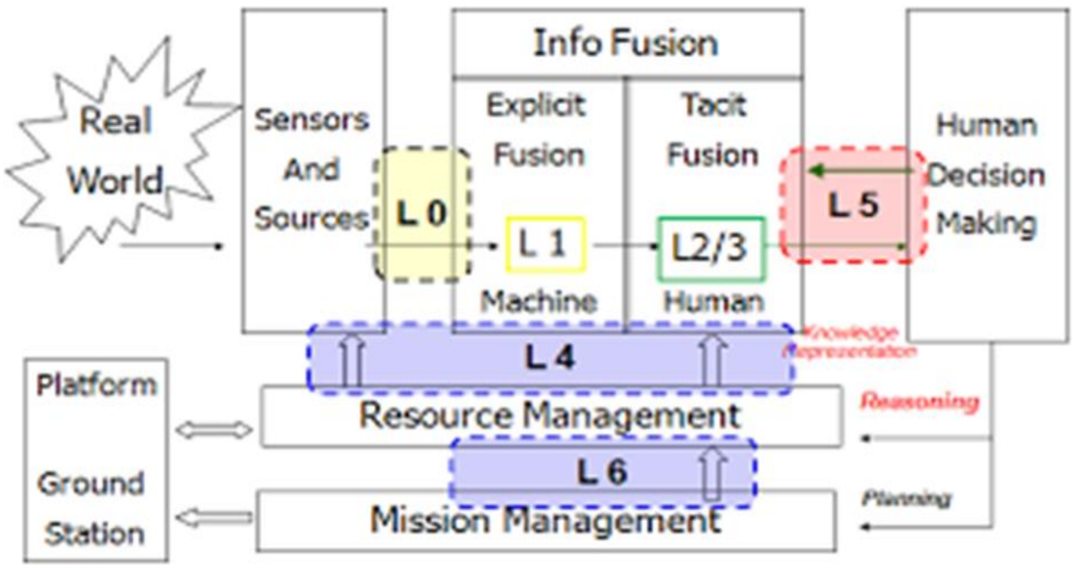


FIGURE 4.3 HUMAN DECISION MODEL BASED ON INFORMATION FUSION

4.8 Chapter Conclusion

The study presented a hybrid emotion recognition framework employing multi-resolution analysis (MRA) of audio and visual signals, with features extracted through discrete wavelet transform (DWT) and classified using SVM, MLP, and K-means classifiers. Using eNTERFACE and RML databases, the system demonstrated effective emotion recognition from both modalities individually. The fusion of audio and pictorial cues at the feature level significantly improved performance, surpassing single-modal systems. The proposed framework achieved notable accuracy, outperforming related systems, emphasizing the efficacy of a hybrid approach in capturing the complexity of human emotions through diverse modalities.

CHAPTER-5

5.1 Discussion and Conclusions

A Vast array of questions are raised throughout the difficult process of choosing measuring instruments and sensors. Both physical techniques used to collect signals and several physiological characteristics to evaluate are available. There is a wide range of options from which to choose thanks to measurement technology in respect to certain sensors refer to Figure 5.1. There have been numerous attempts to categorize emotions, sensors, and global selection methods(Feidakis et al., 2011). We make an attempt to close this disparity with the current plan and some approaches are described here. In comparison to physical observations, determining emotions from measurements of a person's physiological parameters is just a challenging issue. Owing to insufficiency of categorization techniques and working connections among sensors and desirable sentiments, fundamental sensor methodologies become ambiguous. In order to fulfil the two-step selection technique suggested, we provide taxonomy of emotion recognition methodologies in our conclusion. The first phase entails choosing measuring parameters and approaches, while the step two involves choosing sensors.

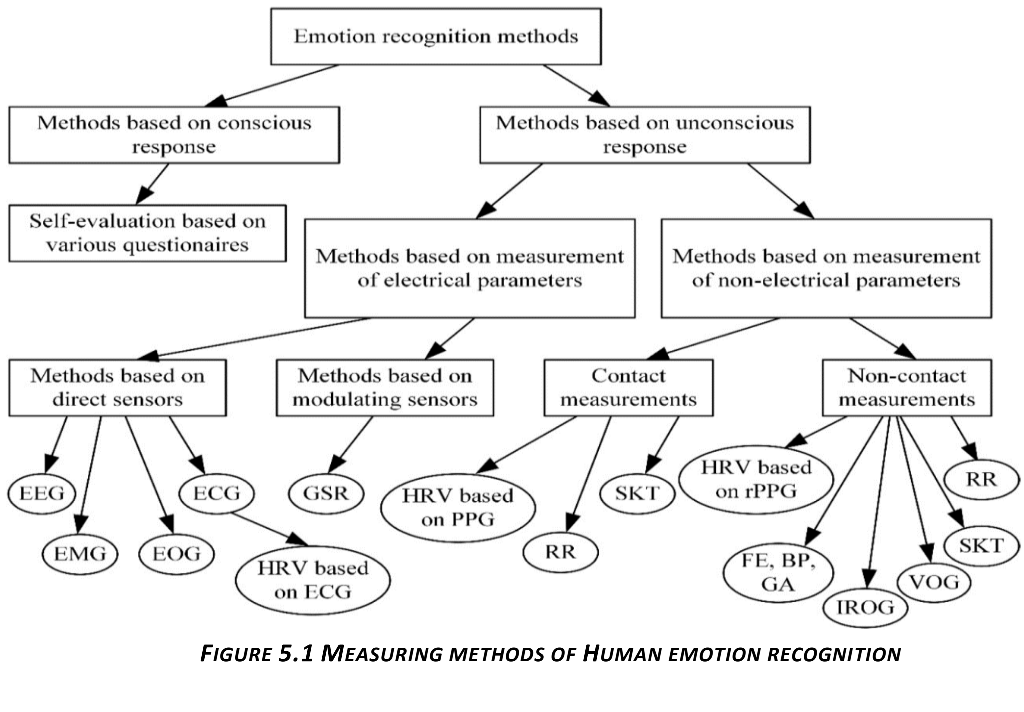


FIGURE 5.1 MEASURING METHODS OF HUMAN EMOTION RECOGNITION

Researchers suppose that in order to choose a strategy, it is first important to specify whether ourselves are concerned to reply, or perhaps in both concurrently(Takahashi, 2004). Conscientious reply research is generally straightforward and does not need any specialized equipment, although creating surveys demands careful consideration. Results from self-evaluation, however, are less trustworthy since there is a chance that a person could misjudge their own emotions or give vague replies to challenging questions(Lin et al., 2019). Techniques usually on unconscious reactions often yield more accurate findings; however they raise high hardware compatibility and necessitate more tries at measuring techniques. Intuitive response-based methodologies offer a variety of options; thus, we suggest choosing between electrical and non-electrical parameter measurements. Electric characteristics are main entities because they deliver results that are generally accurate and because all interactions in humans are governed by electric cues produced in the nervous system. The assessment of non-electrical cues reveals how the body reacts to electric cues. On the contrary, electrical cues can only be analysed by contact assessment techniques(Zucco et al., 2017). Cordless techniques can, of course, be used to transmit a signal to an acquiring machine, but there are still some restrictions on human activity throughout measurements.

It is feasible to use direct (self-generating) detectors to assess electrical characteristics whenever the measured signal is created either by brain's neocortex (EEG, ECG, HRV, EMG, EOG), as well as evaluations based on modulator detectors when features of the sensing vary as the human body transforms (GSR)(Wu et al., 2010). Theoretically, direct detectors are still more accurate, but because sensors absorb some of the signal's strength, they can have a minor impact on the signal (particularly one with a modest magnitude). Synchronizing sensors, on the other hand it will experience some latency, which will vary depending on the characteristics of each device. The fundamental benefit of monitoring systems that utilize measurements of non-electrical factors is the ability to conduct non-contact procedures without restricting human behaviour(Picard, 2003). These techniques are also more suited for practical applications and for approximating emotional situation evaluation. The optimum answer, according to current studies in the field of sentiment evaluation, is multisensory evaluation since the techniques complement one another and increase overall dependability of findings. There is not a methodology that is perfect for every situation.

Dearth of a shared dataset conceptualization is a glaring methodological issue in all emotion detection systems. The number, make-up, duration, and durations of the control groups are chosen by researchers at random or based on feasibility.

Although the characteristics used by each emotion recognition algorithm vary, a clear specification for dependable criteria that address dataset concerns is necessary. This might free up resources that might otherwise be spent for study with solid outcomes(Poria et al., 2017).

Approaches for signal analysis and processing are indeed crucial when choosing methodologies and sensors. The bulk of the time, the effectiveness of signal processing and evaluation techniques is what determines how well emotions are recognized. For instance, respiration and HRV data can be gleaned from ECG data. According to published findings, multi-standards assessment based on statistical methodologies (ANOVA) or on ML procedure are among the most effective strategies used for emotion classification.

In conclusion, we can say that curiosity in emotion recognition and the actual use of this approach is gradually growing and finding use in new fields. The physiological aspect of this item is the main focus of the in-depth study that is now available in public archives. We discovered a dearth of research and unifying categorization cantered on the engineering aspect of this inquiry, such as a lack of justification regarding measuring devices, measurement uncertainly, and straightforward specification as to which strategy, sensor, preparation, and investigation methods are suitable for knowing a strong feeling. Intelligent systems that recognize emotions may have a background thanks to this type of investigation. Alongside computer vision, natural language processing, deep learning, and other related technologies, sensors and techniques for recognising human emotions have made significant advancements in the IoT space. Because of this, there has been a noticeable advancement in our knowledge of human sentiments(Kaklauskas et al., 2019).

Human emotion computing is the study for the development of tools & techniques. This is a way to identify, comprehend, analyse, and simulate the effects on individuals. Number of scholars throughout the world has developed machinery for understanding, expressing, analysing, communicating, and reacting to data regarding feelings as well as for some examples of affective computing. With development of affective computing technology has made it possible to have a novel comprehension of oneself and to have more effective, enhanced human interactions. This promises future technological advancements that will lessen stress rather than increase anxiety. As is often remarked, leadership includes measurements. Computers offer real-time skills, but they are difficult and sophisticated. These abilities enable them to comprehend human emotions better and to react intelligently to them. Human emotions are complex, but they arise naturally and are expressed organically for the same reason.

Their application spans a variety of disciplines, including those in the natural sciences, such as neurology, physiology, and psychiatry. However, in surveys of the published scientific literatures, there is a dearth of detailed analyses of a current state of the science for hybrid frameworks. Yearly, it is anticipated that emotional sensor technology will become a reality. Considering the fact that the majority of currently available human emotion detection sensors, methodologies, and technologies need on-body gadgets and/or voice/facial identification programs, studies and development will focus heavily on contactless technology involved for monitoring sentiments(Kaklauskas et al., 2019). Despite these attempts, the IoT is now only very slowly being humanized with human sentiment detection techniques and/or sensors. As a result, by integrating human sentiment detection sensors and techniques into academic and corporate communities, this study demands an attempt to introduce the concept of humanizing the Internet of Things with affective computational capabilities. The IOT & emotion computing systems, which have been created by the study's authors, serve as validation of something like this fact.

5.2 Future directions

Through order to create most promotions or educational advertising, emotion acknowledgement is a potent and incredibly valuable approach for assessing human emotional responses and forecasting their behaviour. Additionally, the process of developing diverse human machine interface systems can benefit greatly from emotional evaluation and recognition.

Despite a number of still-uncertainties in choosing measuring and data processing methodologies, relationships among specific emotions and bodily responses in humans have long been documented. There are a staggering number of data analysis techniques and efforts at real world applications, but the eight methods that are most frequently employed in that sector are all based on observations of varying factors. In this assessment, we looked good number of scientific articles and classified the AEE approaches used in a synopsis of popular emotion identification techniques and numerous initiatives for increasing the outcomes' correctness. The dependability, sensibility, and sustainability of AEE approaches are also discussed from an engineering standpoint in this work. A combination of these techniques and the use of computer vision for data gathering appear to be a very potent combination in the coming years that will lead to advancements in real-world applications across all industries, beginning with advertising and marketing and ending with manufacturing engineering disciplines etc.

5.3 Feidakis, Daradoumis Rearks

Feidakis, Daradoumis, as well as Cabella's study(Feidakis et al., 2011), which presents a categorization of feelings based on simple designs, states that there are 66 emotional responses, which can be further divided into 10 basic emotions (distress, excitement, suspicion, afraid, joy, joy, love, despair, amaze, believe), and fifty six secondary emotions. It is incredibly challenging to assess such a large number of emotions, particularly when automatic identification and assessment are needed. In addition, overlapping metrics can be assessed for similar emotions. In order to address this concept, the majority of research on sentiments assessment concentrates on other categorizations, which typically contexts to include the aspects of valence & arousal, and only examine the most fundamental emotional responses that can be characterized more simply.

The bulk of studies adapt Russel's circumplex paradigm of emotions see Figure 5.2, which distributes fundamental emotions in ways of valence and arousal in two dimensions. Hence a technique enables the defining of a desired emotions and the assessment of its strength using only two-dimensional analysis. The below described methodology makes it clear how to categorise and evaluate feelings, but there are numerous problems with this approach, particularly in terms of how to measure things, how to evaluate the results, and how to choose measuring hardware and software. The problem of sentiment detection and assessment is further complex by its multidisciplinary approach: philosophy disciplines are concerned with emotion detection and strength assessment, while medical disciplines and measuring device engineering are involved in the measurements and evaluations of human's overall specifications, and mechatronics is concerned with the analysis and resolution of sensor information.

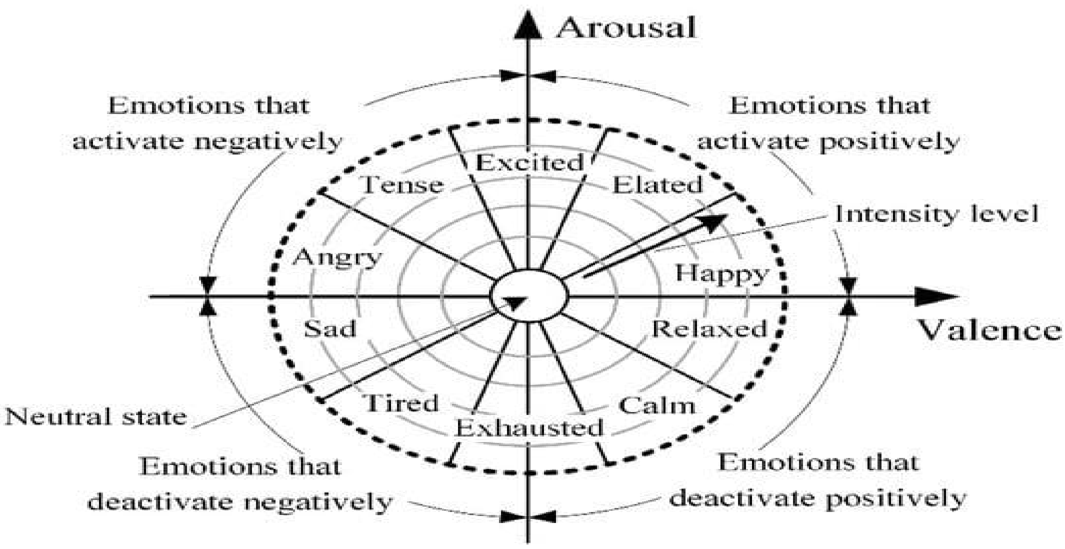


FIGURE 5.2 AROUSAL VALENCE

The technologies and techniques for automatic emotion detection are indeed the main topics of this study. Such techniques can be employed in machine learning processes that use obtained empirical information analysis and automatic methods based on the findings of these investigations. The concept of 'humanizing' the IOT & emotion computing systems, which are supported by specific researchers involved in it.

The universe will undoubtedly become healthier thanks to intelligent devices that have empathy for humans. Due to advancements in voice recognition, computer vision, DL, and associated systems, the Internet of Things (IoT) sector is undoubtedly making headway when it comes to comprehending human emotions.

5.4 Tools for Emotional Assessment

Based on the fundamental strategies used for emotion detection, the emotion estimation techniques some of which are documented in the review can have distinct groups: self-report computational and communication on emotions self-assessment by filling out various survey questions, and machine evaluation procedures involving measurements of various human body characteristics(Wallbott & Scherer, 1989).

In order to maximize the dependability of the outcomes acquired, multiple approaches are frequently used simultaneously. Likewise, emotions can be assessed by examining five main aspects of sentiments (behavioural tendencies, physical response, gestures, cognitive appraisals, and other experiences), but only the first four of these aspects can be assessed to provide cues about such a user's emotional state without interfering with their conversation. Self-assessment approaches are typically the sole way to analyse emotional responses. Automatic emotion recognition is often carried out by evaluating changes in a variety of bodily characteristics or electrical impulses in the nervous system. Electroencephalography, assessments of skin resistance, blood pressure, heart rate, eye activity, and pose estimation are the most often used methods.

5.5 Electroencephalography (EEG)

This is a straightforward electrophysiological method to capture when a person's brain produces electrical impulses. Hans Berger, a German psychiatrist, introduced EEG for mankind during his first publication on the application of EEG technology(Gonçalves et al., 2017). Typically, EEG signals are recorded using an electroencephalogram, a specialized tool Figure 5.3.

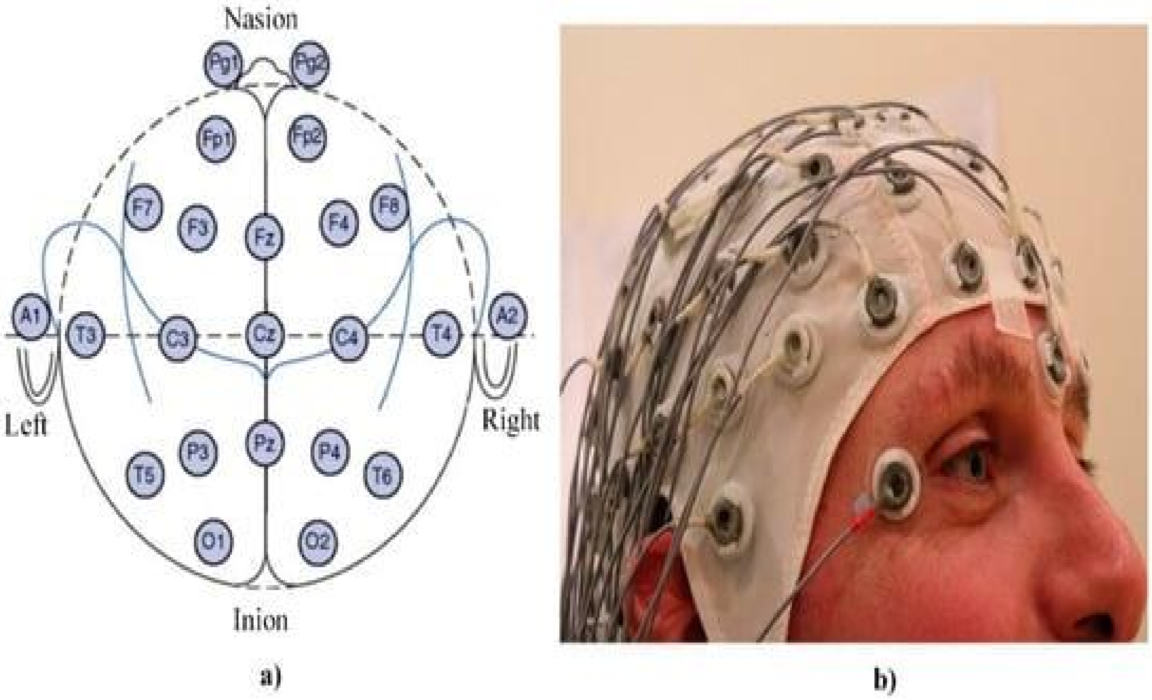


FIGURE 5.3 EEG DISTRIBUTION A) ELECTRODES B) SKILL SET

Knowing that components of such apparatus are unique metal plate sensors that must be applied to a person's scalp; alternate needle electrodes may also be applied in certain circumstances. Most frequently, the nasion, strategically placing, and right and left in preauricular spots on the skull are the locations of 8, 16, or 32 pairs of electrodes. A specific helmet with electrodes affixed or adhesive-conducting gel can be used to adhere conductors to a person's skull. The EEG cue is a variation in voltage across 2 paired sensors with regard to time and the maximum average method is typically used to measure signal amplitude Figure 5.4.

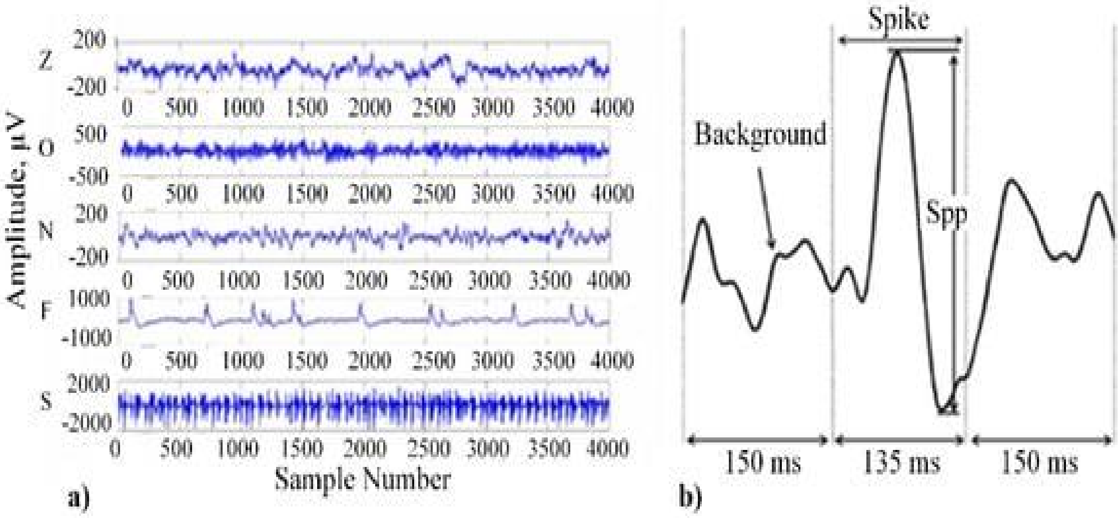


FIGURE 5.4 EEG SIGNALS A) RAW DATA B) SIGNAL AMPLITUDE EVALUATION TECHNIQUE

5.5 Electrocardiography (ECG)

Among the important parts in the human body is the heart, and ECG is an important most effective diagnostic techniques in healthcare, frequently used to evaluate the heart's performance. Since an ECG is a physiological signal, it is typically utilized as a non-invasive approach to evaluate the heart's electrical activity in real time. ECGs are helpful not only for studying the heart's activity but also for identifying emotions because cardiac function is connected to the human brain's central nervous system. Numerous studies go into great detail about the ECG recording process(Goshvarpour & Abbasi, 2017). The 12-lead ECG technique is often utilized method. Nine sensors are implanted on the humans as depicted in the figure for this procedure. The three primary senses are split between the left leg (LL). The only connection to the right-leg (RL) is a wire, whereby the sensors are linked to it should utilise as their common ground. With only these triple sensors, doctors can utilize a technique known as a 3-lead ECG, which has limitations in that it does not provide data on all areas of the heart but is still helpful in emergency situations where speedy analysis is necessary.

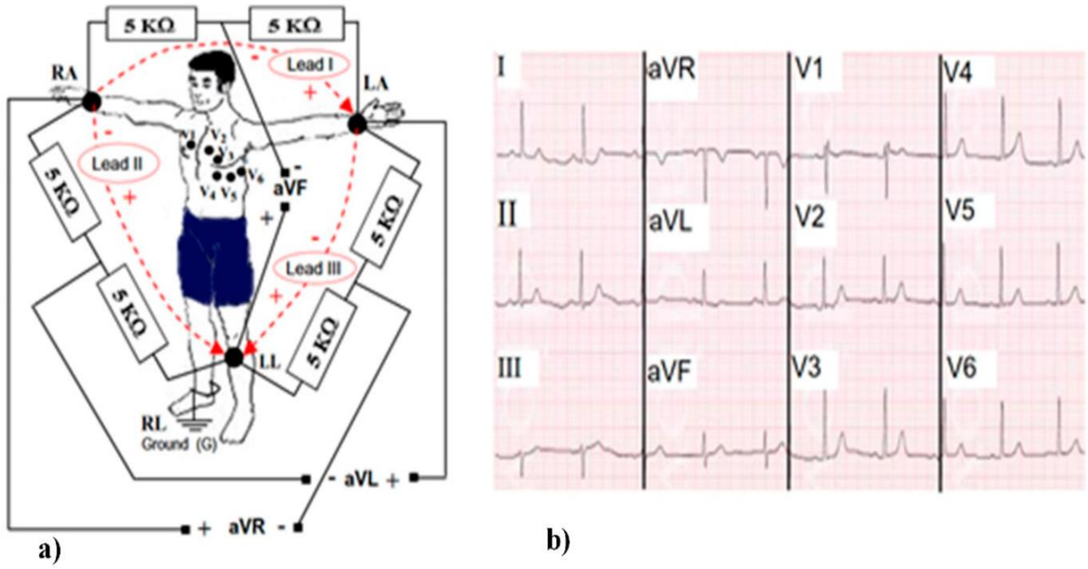


FIGURE 5.5 ECG REPRESENTATION A) LEADS OF ECG B) QUARTER SIGNALS OF ECG

The 12-lead ECG produces 12 signals when all nine sensors are used, which are referred to in biomedicine as Lead-I, Lead-II, Lead-III, aVR, aVL, aVF, V-1, V-2, V-3, V-4, V-5, and V-6 as shown in Figure 5.5.

5.6 Electro-dermal Sensitivity

A cumulative assessment of the electrical characteristics of human skin is the skin temperature responses (GSR), consequently called as electrodermal activity (EDA). The main factor in this technique is most frequently skin conductions(Wu et al., 2010). Skin's electrical properties are not under the consciousness of a person. since, in accordance with conventional wisdom, they are dependent on variations in sweating, which are thought to be a reflection of alterations in the sympathetic nervous system. There is evidence to support the claim that few cues from sympathetic nervous burst are accompanied by variations in skin currents. The primary relationship between electro-dermal and arousal is that when arousal rises, so does skin conductance.

The acquired measurement outputs correlate with the self-assessed appraisal of arousal, and GSR signal amplitude is related to stress, willingness, involvement, anger, and rage(Critchley, 2002). Both the frequency and the amplitude of GSR increases, simultaneously in response to attention-demanding activities and attention-grabbing stimuli. GSR may thus automatically identify the decision-making process in addition to recognizing emotions(Ayata et al., 2017) Figure 5.6 and Figure 5.7 depicts GSR attached and function accordingly.

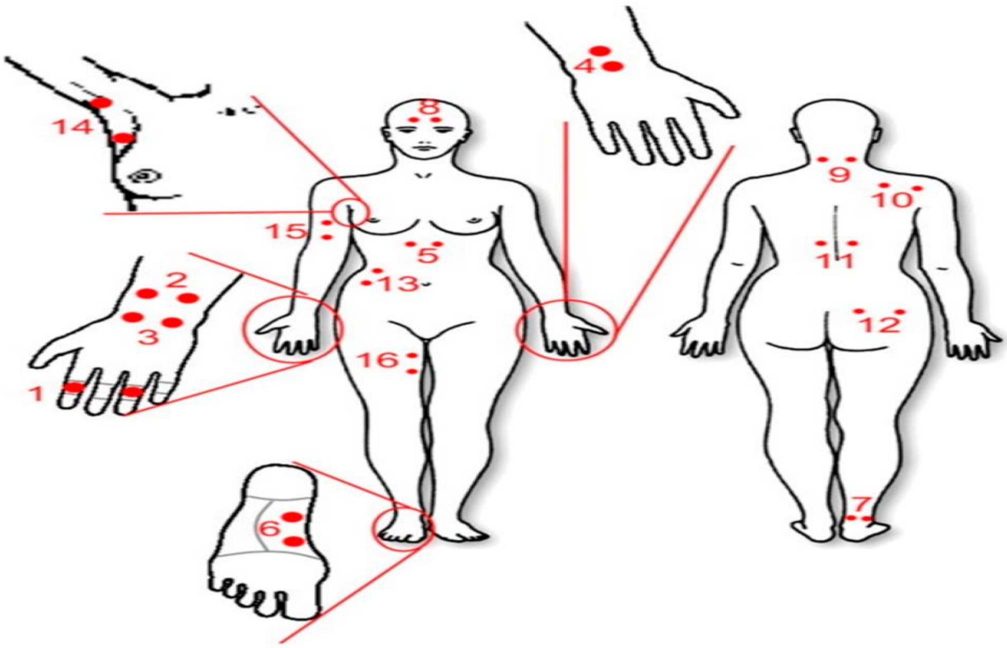


Figure 5.6 GSR Electrodes attached at varient places

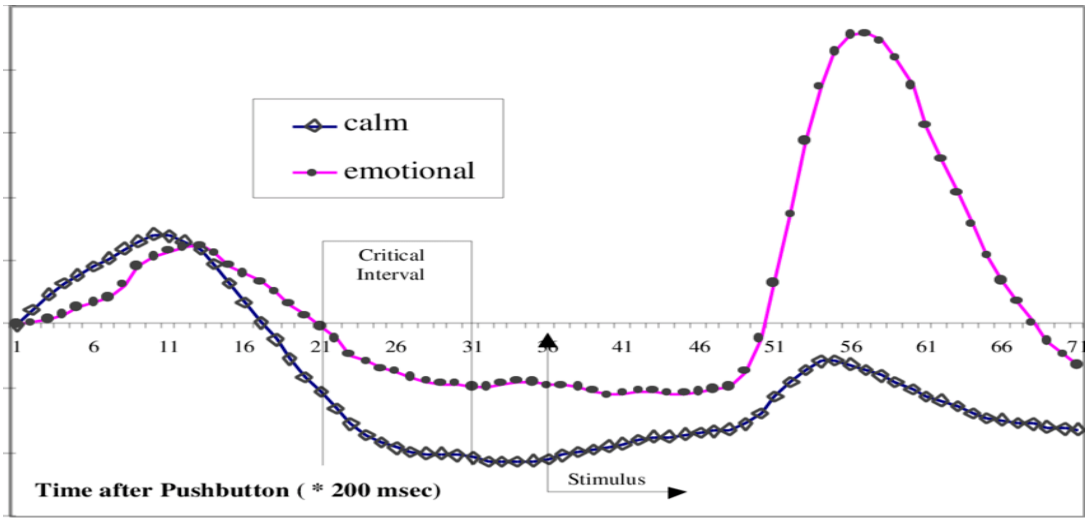


Figure 5.7 GSR for function calm and emotional states

5.7 Comparison with related research

Comparing of reliability with other systems based on eNTERFACE databases are given in Table 4.14. It was conceived that the average performance of our system (based on MLP) is 88.3%, which is much more compared to the performance of other systems as shownin table 5.1.

Table 5.1 Accuracy comparison with other systems based on eNTERFACE database

Systems	Year	Features	Accuracy (%)
Schuller B. et al.	*2009*	*Arousal- valance*	*80.2 for SVM &* *80.5 for HMM/GMM*
Datcu D. et al.	*2009*	*Facial features*	*85 through SVM*
Mansoorizadeh M. et al.	*2010*	*Asynchronous* *Feature level fusion*	*83*
Paleari M. et al.	*2010*	*Neural network*	*71*
Wang Y. et al.	*2012*	*emotion framework*	*72.41*
Our study	*2014*	*wavelet analysis*	*88.3*

5.8 Conclusion and Future Work

The study underscores challenges in emotion recognition, emphasizing the need for a comprehensive engineering-focused approach. It proposes a two-step selection technique, highlighting the importance of signal analysis and processing for effective emotion recognition. The future direction suggests the potential of emotion recognition in promotions and human-machine interface systems. Tools like EEG, ECG, and Electro-dermal Sensitivity are discussed for emotional assessment. The study's proposed system outperforms others with an average accuracy of 88.3%, showcasing its potential in emotion detection. Overall, the study advocates for a multidisciplinary and technology-driven advancement in emotion recognition systems. However, more advancement is needed, such as raising the number of recognized emotions to create a real-time emotion identification system. A whole signal provides more helpful information, and neglecting any time segment decreases accuracy, as we have demonstrated; in the future, we want to test deep learning on data from unseen participants.

ABBREVIATIONS

ECG	Electrocardiogram
ANN	Artificial Neural Network
ML	Machine Learning
BBN	Bayesian Belief Network.
BEs	Basic Emotions
BP	Blood Pressure
CAE	Convolutionary Auto Encoder
CBF	Cloud Based Function
CEs	Compound Emotions
CNN	Photoplethysmography
CU	Control Unit
CSRC	Contributing Source
CV	Computer Vision
DL	Deep Learning
FACS	Facial Action Coding System
FAM	Fuzzy Art Map
FAS	Facial Action System
CIO	Cognitive Internet of Things
HER	Human Emotion Recognition
HHER	Hybrid Human Emotion Recognition Recognition.

FEP	Facial Expression Parameters
FER	Facial Emotion Recognition
FLs	Facial Landmarks
FN	False Negative
FP	False Positive
GCN	Global Comparison Standardization
GPU	Graphical Processing Unit
HMM	Hidden Markov Model
IoT	Internet of Things
KNN	k-Nearest Neighbor
LCD	Liquid Crystal Display
LDP	Local Directional Pattern
LFDA	Local Fisher Discriminatory Facial Discrimination
LFW	Labeled Faces in the Wild
LIPO	Lithium Ion Polymer
LTP	Local Ternary Pattern
MAOP-DL	Multi-Angle Pattern-Based Deep Learning
MEs	Micro Expressions
ML	Maximum Likelihood
MLP	Multilayer Perceptron
MTCNN	Multi-Task CNN

NCS	Neural Compute Stick
NIR	Near-infrared
NN	Nearest Neighbor
NS	Nearest Subspace
Open VINO	Open Visual Inference and Neural Network Optimization
PDM	Point Distribution Model
P-MLBP	Polytypical Local Multi Block Binary Pattern
DOS	Distributed Operating System
RBF	Radial Base Function
SVM	Support Vector Machine
TAN	Tree-improved Naive Bayesian
USB	Universal Serial Bus
VIS	Visible Light

Bibliography

Chen, M., Hao, Y., Lin, K., Yuan, Z., & Hu, L. (2018). Label-less Learning for Traffic Control in an Edge Network. *IEEE Network*, *32*(6), 8–14. https://doi.org/10.1109/MNET.2018.1800110

Chen, M., Herrera, F., & Hwang, K. (2018). Cognitive Computing: Architecture, Technologies and Intelligent Applications. *IEEE Access*, *6*, 19774–19783. https://doi.org/10.1109/ACCESS.2018.2791469

Chen, M., Zhou, P., Wu, D., Hu, L., Hassan, M. M., & Alamri, A. (2020). AI-Skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework. *Information Fusion*, *54*(January 2019), 1–9. https://doi.org/10.1016/j.inffus.2019.06.005

Chen, X., Zhao, Y., & Li, Y. (2019). QoE-Aware wireless video communications for emotion-aware intelligent systems: A multi-layered collaboration approach. *Information Fusion*, *47*, 1–9. https://doi.org/https://doi.org/10.1016/j.inffus.2018.06.007

D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, *47*(3). https://doi.org/10.1145/2682899

Duc, Beno. (1997). *Fusion of audio and video information for multi modal person authentication Elizabeth Saers Bigiin b , c , Josef Bigiin d , Gilbert Maitre*. *18*, 835–843.

Duc, Benoît, Bigün, E. S., Bigün, J., Maître, G., & Fischer, S. (1997). Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters*, *18*(9), 835–843. https://doi.org/https://doi.org/10.1016/S0167-8655(97)00071-8

Fayaz, F. A. (2021). *Real time data evaluation with wearable devices : An Impact of Artifact Calibration Method on Emotion Recognition*. 2021–2024. https://doi.org/10.1109/ICCS54944.2021.00038

Fayaz, F. A., Malik, A., Batra, I., Gardezi, A. A., Ansarullah, S. I., Ahmad, S., Alqahtani, M., & Shafiq, M. (2023). *Impediments of Cognitive System Engineering in Machine-Human Modeling*. https://doi.org/10.32604/cmc.2023.032998

Fayaz, F. A., Malik, A., & Yatoo, A. A. (2021). Cognitive Internet of things (CIoT) a success for data collection. *2021 Sixth International Conference on Image Information Processing (ICIIP)*, *6*, 284–287. https://doi.org/10.1109/ICIIP53038.2021.9702706

Fayaz, F. A., Mohi-Ud-Din, S., Batool, I., Kaur, S., & Rashid, M. (2019). Novel Face

Recognition Based Examinee Authentication System using Python D-Lib. *2019 Fifth International Conference on Image Information Processing (ICIIP)*, 480–485. https://doi.org/10.1109/ICIIP47207.2019.8985983

Fortino, G., Galzarano, S., Gravina, R., & Li, W. (2014). A framework for collaborative computing and multi-sensor data fusion in body sensor networks. *INFORMATION FUSION*. https://doi.org/10.1016/j.inffus.2014.03.005

Fortino, G., Giannantonio, R., Gravina, R., Kuryloski, P., & Jafari, R. (2013). Enabling Effective Programming and Flexible Management of Efficient Body Sensor Network Applications. *IEEE Transactions on Human-Machine Systems*, *43*(1), 115–133. https://doi.org/10.1109/TSMCC.2012.2215852

Fortino, G., Parisi, D., Pirrone, V., & Di Fatta, G. (2014). BodyCloud: A SaaS approach for community Body Sensor Networks. *Future Generation Computer Systems*, *35*, 62–79. https://doi.org/https://doi.org/10.1016/j.future.2013.12.015

Gomez-Donoso, F., Orts-Escolano, S., & Cazorla, M. (2019). Large-scale multiview 3D hand pose dataset. *Image and Vision Computing*, *81*, 25–33. https://doi.org/10.1016/j.imavis.2018.12.001

Gravina, R., & Li, Q. (2019). Emotion-relevant activity recognition based on smart cushion using multi-sensor fusion. *Information Fusion*, *48*, 1–10. https://doi.org/10.1016/j.inffus.2018.08.001

Imani, M., & Montazer, G. A. (2019). A survey of emotion recognition methods with emphasis on E-Learning environments. *Journal of Network and Computer Applications*, *147*(August), 102423. https://doi.org/10.1016/j.jnca.2019.102423

Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. *Pattern Recognition Letters*, *115*, 101–106. https://doi.org/10.1016/j.patrec.2018.04.010

Jiang, Y., Li, W., Hossain, M. S., Chen, M., Alelaiwi, A., & Al-Hammadi, M. (2020). A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. *Information Fusion*, *53*(February 2019), 209–221. https://doi.org/10.1016/j.inffus.2019.06.019

Katona, J. (2021). A review of human–computer interaction and virtual reality research fields in cognitive infocommunications. *Applied Sciences (Switzerland)*, *11*(6). https://doi.org/10.3390/app11062646

Khezri, M., Firoozabadi, M., & Sharafat, A. R. (2015). Ac ce p te d cr t. *Computer Methods and Programs in Biomedicine*. https://doi.org/10.1016/j.cmpb.2015.07.006

Lipponen, J. A., & Tarvainen, M. P. (2019). A robust algorithm for heart rate variability time series artefact correction using novel beat classification. *Journal of Medical Engineering and Technology*, *43*(3), 173–181.

https://doi.org/10.1080/03091902.2019.1640306

Mehedi, M., Rabiul, G., Uddin, Z., Huda, S., Almogren, A., & Fortino, G. (2019). Human emotion recognition using deep belief network architecture. *Information Fusion*, *51*(November 2018), 10–18. https://doi.org/10.1016/j.inffus.2018.10.009

Muhammad, G., & Alhamid, M. F. (2017). User emotion recognition from a larger pool of social network data using active learning. *Multimedia Tools and Applications*, *76*(8), 10881–10892. https://doi.org/10.1007/s11042-016-3912-2

Pirouz, B., Haghshenas, S. S., Haghshenas, S. S., & Piro, P. (2020). Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of Coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability (United States)*, *12*(6). https://doi.org/10.3390/su12062427

Qian, Y., Zhang, Y., Ma, X., Yu, H., & Peng, L. (2019). EARS: Emotion-aware recommender system based on hybrid information fusion. *Information Fusion*, *46*(February 2018), 141–146. https://doi.org/10.1016/j.inffus.2018.06.004

Ringeval, F., Valstar, M., Marchi, E., Lalanne, D., & Cowie, R. (2015). *AV + EC 2015 – The First Affect Recognition Challenge Bridging Across Audio , Video , and Physiological Data*. 3–8.

Santos, M. A. G., Munoz, R., Olivares, R., Filho, P. P. R., Ser, J. Del, & de Albuquerque, V. H. C. (2020). Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. *Information Fusion*, *53*(June 2019), 222–239. https://doi.org/10.1016/j.inffus.2019.06.004

Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, *3*(2), 1–20. https://doi.org/10.1007/s42979-022-01043-x

Shirke, B., Wong, J., Libut, J. C., George, K., & Oh, S. J. (2020). Brain-IoT based Emotion Recognition System. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 991–995. https://doi.org/10.1109/CCWC47524.2020.9031124

Soundararajan, R., & Biswas, S. (2019). Machine vision quality assessment for robust face detection. *Signal Processing: Image Communication*, *72*(December 2018), 92–104. https://doi.org/10.1016/j.image.2018.12.012

Wang, Z., Li, H., Yin, Z., & Xiao, H. (2017). A new, direct approach toward modeling thermo-coupled fatigue failure behavior of metals and alloys. *Acta Mechanica Solida Sinica*. https://doi.org/10.1016/j.camss.2016.10.001

Yang, J. U. N., & Hu, L. (2018). Wearable Affective Robot. *IEEE Access*, *6*, 64766–64776. https://doi.org/10.1109/ACCESS.2018.2877919

Yin, Z., Zhao, M., Wang, Y., Yang, J., & Zhang, J. (2017). *Computer Methods and Programs in Biomedicine Recognition of emotions using multimodal physiological signals and an ensemble deep learning model*. *140*, 93–110. https://doi.org/10.1016/j.cmpb.2016.12.005

Al Khatib, I., Bertozzi, D., Poletti, F., Benini, L., Jantsch, A., Bechara, M., Khalifeh, H., Hajjar, M., Nabiev, R., & Jonsson, S. (2007). Hardware/software architecture for real-time ECG monitoring and analysis leveraging MPSoC technology. *Transactions on High-Performance Embedded Architectures and Compilers I*, 239–258.

Ayata, D., Yaslan, Y., & Kamaşak, M. (2017). Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods. *IU-Journal of Electrical & Electronics Engineering*, *17*(1), 3147–3156.

Critchley, H. D. (2002). Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, *8*(2), 132–142. https://doi.org/10.1177/107385840200800209

Feidakis, M., Daradoumis, T., & Caballé, S. (2011). Endowing e-learning systems with emotion awareness. *2011 Third International Conference on Intelligent Networking and Collaborative Systems*, 68–75.

Gonçalves, V. P., Giancristofaro, G. T., Filho, G. P. R., Johnson, T., Carvalho, V., Pessin, G., Neris, V. P. de A., & Ueyama, J. (2017). Assessing users' emotion at interaction time: a multimodal approach with multiple sensors. *Soft Computing*, *21*(18), 5309–5323. https://doi.org/10.1007/s00500-016-2115-0

Goshvarpour, A., & Abbasi, A. (2017). An emotion recognition approach based on wavelet transform and second-order difference plot of ECG. *Journal of AI and Data Mining*, *5*(2), 211–221.

Kaklauskas, A., Zavadskas, E. K., Bardauskiene, D., Cerkauskas, J., Ubarte, I., Seniut, M., Dzemyda, G., Kaklauskaite, M., Vinogradova, I., & Velykorusova, A. (2019). An Affect-Based Built Environment Video Analytics. *Automation in Construction*, *106*, 102888. https://doi.org/https://doi.org/10.1016/j.autcon.2019.102888

Lin, C.-J., Lin, C.-H., Wang, S.-H., & Wu, C.-H. (2019). Multiple Convolutional Neural Networks Fusion Using Improved Fuzzy Integral for Facial Emotion Recognition. *Applied Sciences*, *9*(13). https://doi.org/10.3390/app9132593

Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, *59*(1), 55–64. https://doi.org/https://doi.org/10.1016/S1071-5819(03)00052-1

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125. https://doi.org/https://doi.org/10.1016/j.inffus.2017.02.003

Takahashi, K. (2004). Remarks on emotion recognition from bio-potential signals. *2nd International Conference on Autonomous Robots and Agents*, *3*, 1148–1153.

Wallbott, H. G., & Scherer, K. R. (1989). Assessing emotion by questionnaire. In *The measurement of emotions* (pp. 55–82). Elsevier.

Wu, G., Liu, G., & Hao, M. (2010). The analysis of emotion recognition from GSR based on PSO. *2010 International Symposium on Intelligence Information Processing and Trusted Computing*, 360–363.

Zucco, C., Calabrese, B., & Cannataro, M. (2017). Sentiment analysis and affective computing for depression monitoring. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1988–1995.

Ali, O., Ishak, M. K., & Bhatti, M. K. L. (2021). Emerging IoT domains, current standings and open research challenges: a review. *PeerJ Computer Science*, *7*, e659. https://link.gale.com/apps/doc/A672203859/AONE?u=anon~97979f08&sid=googleScholar&xid=ae221450

Chen, M., Herrera, F., & Hwang, K. (2018). Cognitive Computing: Architecture, Technologies and Intelligent Applications. *IEEE Access*, *6*, 19774–19783. https://doi.org/10.1109/ACCESS.2018.2791469

Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., López, M. T., Zangróniz, R., Latorre, J. M., & Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics*, *64*, 55–73. https://doi.org/10.1016/j.jbi.2016.09.015

Jiang, Y., Li, W., Hossain, M. S., Chen, M., Alelaiwi, A., & Al-Hammadi, M. (2020). A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. *Information Fusion*, *53*(February 2019), 209–221. https://doi.org/10.1016/j.inffus.2019.06.019

Li, X., & Zhang, X. (2021). Multi-Task Allocation under Time Constraints in Mobile Crowdsensing. *IEEE Transactions on Mobile Computing*, *20*(4), 1494–1510. https://doi.org/10.1109/TMC.2019.2962457

Li, Z., Liu, Y., Ma, M., Liu, A., Zhang, X., & Luo, G. (2018). MSDG: A novel green data gathering scheme for wireless sensor networks. *Computer Networks*, *142*, 223–239. https://doi.org/https://doi.org/10.1016/j.comnet.2018.06.012

Liu, X., Dong, M., Ota, K., Hung, P., & Liu, A. (2016). Service Pricing Decision in Cyber-Physical Systems: Insights from Game Theory. *IEEE Trans. Serv. Comput.*, *9*(2), 186–198. https://doi.org/10.1109/TSC.2015.2449314

Liu, Y., Liu, A., Guo, S., Li, Z., Choi, Y.-J., & Sekiya, H. (2020). Context-aware collect data with energy efficient in Cyber–physical cloud systems. *Future Generation Computer Systems*, *105*, 932–947. https://doi.org/10.1016/j.future.2017.05.029

Luo, T. (2015). Quality of Contributed Service and Market Equilibrium for Participatory Sensing (TMC). *IEEE Transactions on Mobile Computing*, *14*, 829–842. https://doi.org/10.1109/TMC.2014.2330302

Ma, F., Liu, X., Liu, A., Zhao, M., Huang, C., Wang, T., & Ning, Z. (2018). A Time and Location Correlation Incentive Scheme for Deep Data Gathering in Crowdsourcing Networks. *Wirel. Commun. Mob. Comput.*, *2018*. https://doi.org/10.1155/2018/8052620

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141–162. https://doi.org/10.1016/j.cognition.2015.06.010

Park, J., Salim, M. M., Jo, J. H., Sicato, J. C. S., Rathore, S., & Park, J. H. (2019). CIoT-Net: a scalable cognitive IoT based smart city network architecture. *Human-Centric Computing and Information Sciences*, *9*(1), 29. https://doi.org/10.1186/s13673-019-0190-9

Qi, Y., Du, C. D., Liu, T., Zhao, X., & Dong, C. (2020). Experts' conservative judgment and containment of COVID-19 in early outbreak. *Journal of Chinese Governance*, *5*(2), 140–159. https://doi.org/10.1080/23812346.2020.1741240

Qian, Y., Zhang, Y., Ma, X., Yu, H., & Peng, L. (2019). EARS: Emotion-aware recommender system based on hybrid information fusion. *Information Fusion*, *46*(February 2018), 141–146. https://doi.org/10.1016/j.inffus.2018.06.004

Rashid, M. M., Kamruzzaman, J., Hassan, M. M., Imam, T., & Gordon, S. (2020). Cyberattacks Detection in IoT-Based Smart City Applications Using Machine Learning Techniques. *International Journal of Environmental Research and*

Public Health, *17*(24). https://doi.org/10.3390/ijerph17249347

Reddy, S., Estrin, D., & Srivastava, M. (2010). *Recruitment Framework for Participatory Sensing Data Collections BT - Pervasive Computing* (P. Floréen, A. Krüger, & M. Spasojevic (eds.); pp. 138–155). Springer Berlin Heidelberg.

Ren, Y., Zeng, Z., Wang, T., Zhang, S., & Zhi, G. (2020). A trust-based minimum cost and quality aware data collection scheme in P2P network. *Peer-to-Peer Networking and Applications*, *13*(6), 2300–2323. https://doi.org/10.1007/s12083-020-00898-2

Spinelli, B., Celis, L. E., & Thiran, P. (n.d.). *A General Framework for Sensor Placement in Source Localization*.

Teng, H., Liu, Y., Liu, A., Xiong, N. N., Cai, Z., Wang, T., & Liu, X. (2019). A Novel Code Data Dissemination Scheme for Internet of Things through Mobile Vehicle of Smart Cities. *Future Gener. Comput. Syst.*, *94*(C), 351–367. https://doi.org/10.1016/j.future.2018.11.039

Thiagarajan, A., Ravindranath, L., Lacurts, K., & Eriksson, J. (2009). *VTrack : Accurate , energy-aware road traffic VTrack : Accurate , Energy-aware Road Traffic Delay Estimation Using Mobile Phones*.

Thoits, Peggy, A. (2011). Mechanisms Linking Social Ties and Support to Psysical and Mental Health. *Journal of Health and Social Behavior*, *52*(2), 145–161. http://dx.doi.org/10.1016/j.tics.2011.08.003%0Ahttp://jn.physiology.org/cgi/doi/10.1152/jn.00838.2004%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/3204199%0Ahttp://bmjopen.bmj.com/%0Ahttp://www.nature.com/articles/s41398-019-0678-0%0Ahttp://www.brain.oxfordjourna

Yang, D., Xue, G., Fang, X., & Tang, J. (2012). Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing. *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, 173–184. https://doi.org/10.1145/2348543.2348567

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*(September 2017), 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Hagarty, I., & Morgan, G. (2020). Social-emotional learning for children with learning disabilities: a systematic review. *Educational Psychology in Practice*, *36*(2), 208–222. https://doi.org/10.1080/02667363.2020.1742096

Ko, Y. C., & Fujita, H. (2013). Knowledge-based system acquisition for competitiveness predictions in decision making. *SoMeT 2013 - 12th IEEE International Conference on Intelligent Software Methodologies, Tools and Techniques, Proceedings*, 265–270. https://doi.org/10.1109/SoMeT.2013.6645673

Molnar, B., & Mattyasovszky-Philipp, D. (2019). An architectural approach to Cognitive Information System. *10th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2019 - Proceedings*, 459–462. https://doi.org/10.1109/CogInfoCom47531.2019.9089899

Patel, C. I., Garg, S., Zaveri, T., Banerjee, A., & Patel, R. (2018). Human action recognition using fusion of features for unconstrained video sequences. *Computers and Electrical Engineering*, *70*, 284–301. https://doi.org/10.1016/j.compeleceng.2016.06.004

Thoits, Peggy, A. (2011). Mechanisms Linking Social Ties and Support to Psysical and Mental Health. *Journal of Health and Social Behavior*, *52*(2), 145–161. http://dx.doi.org/10.1016/j.tics.2011.08.003%0Ahttp://jn.physiology.org/cgi/doi/10.1152/jn.00838.2004%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/3204199%0Ahttp://bmjopen.bmj.com/%0Ahttp://www.nature.com/articles/s41398-019-0678-0%0Ahttp://www.brain.oxfordjourna

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). 2005_Tomasello_et.al_UnderstandingAndSharing. *Behavioral and Brain Sciences*, *28*, 675–735.

Wang, Y. (2007). *5-IJCINI-1101-CI Framework.pdf*. *1*(March), 1–27.

Wolfert, P., Robinson, N., & Belpaeme, T. (2022). A Review of Evaluation Practices of Gesture Generation in Embodied Conversational Agents. *IEEE Transactions on Human-Machine Systems*, *0*(0), 1–11. https://doi.org/10.1109/THMS.2022.3149173

Wright, J. J., & Liley, D. T.

J. (1996). Dynamics of the brain at global and microscopic scales: Neural networks and the EEG. *Behavioral and Brain Sciences*, *19*(2), 285–320. https://doi.org/10.1017/s0140525x00042679

Zhu, Y., Chen, X., Chu, K., Wang, X., Hu, Z., & Su, H. (2022). Carbon Black/PDMS Based Flexible Capacitive Tactile Sensor for Multi-Directional Force Sensing. *Sensors*, *22*(2). https://doi.org/10.3390/s22020628

Cheng, J., Chen, M., Li, C., Liu, Y., Song, R., Liu, A., & Chen, X. (2021). Emotion Recognition From Multi-Channel EEG via Deep Forest. *IEEE Journal of Biomedical and Health Informatics*, *25*(2), 453–464. https://doi.org/10.1109/JBHI.2020.2995767

Moin, A., Aadil, F., Ali, Z., & Kang, D. (2023). Emotion recognition framework using multiple modalities for an effective human–computer interaction. *Journal of Supercomputing*, *79*(8), 9320–9349. https://doi.org/10.1007/s11227-022-05026-w

Spinelli, B., Celis, L. E., & Thiran, P. (n.d.). *A General Framework for Sensor Placement in Source Localization*.

Wang, F., Wu, S., Zhang, W., Xu, Z., Zhang, Y., Wu, C., & Coleman, S. (2020). Emotion recognition with convolutional neural network and EEG-based EFDMs. *Neuropsychologia*, *146*(May), 107506. https://doi.org/10.1016/j.neuropsychologia.2020.107506

Yang, D., Alsadoon, A., Prasad, P. W. C., Singh, A. K., & Elchouemi, A. (2018). An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment. *Procedia Computer Science*, *125*(2009), 2–10. https://doi.org/10.1016/j.procs.2017.12.003

Publications:

International Conferences

1. Fayaz, F. A. (2021). Real time data evaluation with wearable devices: An Impact of Artifact Calibration Method on Emotion Recognition. 2021–2024. https://doi.org/10.1109/ICCS54944.2021.00038
2. F. A. Fayaz, A. Malik and A. A. Yatoo, "Cognitive Internet of things (CIoT) a success for data collection," 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 2021, pp. 284-287, doi: 10.1109/ICIIP53038.2021.9702706. https://ieeexplore.ieee.org/document/9702706
3. "Novel Hybrid Emotion Recognition Framework (NHERF) for Virtual Learning Assessment Model a Review" http://www.solidstatetechnology.us/index.php/JSST/article/view/1851
4. F. A. Fayaz, S. Mohi-Ud-Din, I. Batool, S. Kaur and M. Rashid, "Novel Face Recognition Based Examinee Authentication System using Python D-Lib," 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 2019, pp. 480-485, doi:10.1109/ICIIP47207.2019.8985983 https://ieeexplore.ieee.org/document/8985983

International Conference paper accepted for oral presentation

Scheduled on 16th and 17th of February 2014

1. NICEDT-2024: NIELIT's International Conference on Communication, Electronics and Digital Technologies, Dates: 16-17 February 2024, Guwahati India, Paper titled "Designing a Robust Concealer for Emotion Detection using Various Paradigms for Machine Human Interaction"

International Journals:

1. Fayaz, F. A., Malik, A., Batra, I., Gardezi, A. A., Ansarullah, S. I., Ahmad, S., Alqahtani, M., & Shafiq, M. (2023). *Impediments of Cognitive System Engineering in Machine-Human Modeling*. https://doi.org/10.32604/cmc.2023.032998

International Journal Paper under revision:

1. **Springer Nature** "A Resilient Overlay for Human Emotion Recognition Using Mixed Frameworks in Machine-Human interactions" SNCS-D-23-05533R1. https://www2.cloud.editorialmanager.com/sncs/default2.aspx

Patients:

1. **Patient-Id 4872**
   In progress (Utsukhta Maap Saadhan for online teaching)
   Eagerness measuring Technique for online teaching

Book Chapter:

2. Concerns over Ensuring Security on 6G Networks in CRC, Taylor and Francis, U.K Book titled "Network Security and Data Privacy in 6G Environment: Impacts and Challenges"

Appendices

<u>Appendix-I</u> (Journal SCIE/ Scopus indexed Paper "*Impediments of Cognitive System Engineering in Machine-Human Modeling*"

Computers, Materials & Continua
DOI: 10.32604/cmc.2023.032998
Article

Tech Science Press
check for updates

Impediments of Cognitive System Engineering in Machine-Human Modeling

Fayaz Ahmad Fayaz[1,2], Arun Malik[2], Isha Batra[2], Akber Abid Gardezi[3], Syed Immamul Ansarullah[4], Shafiq Ahmad[5], Mejdal Alqahtani[5] and Muhammad Shafiq[6,*]

[1]National Institute of Electronics and Information Technology (NIELIT), J&K, 191132, India
[2]Department of Computer Science & Engineering, Lovely Professional University (LPU), Punjab, 144402, India
[3]Department of Computer Science, COMSATS University Islamabad, Islamabad, 45550, Pakistan
[4]Department of Computer Science, Govt. Degree College Sumbal Kashmir, 193501, India
[5]Industrial Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh, 11421, Saudi Arabia
[6]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Korea
*Corresponding Author: Muhammad Shafiq. Email: shafiq@ynu.ac.kr
Received: 03 June 2022; Accepted: 19 August 2022

Abstract: A comprehensive understanding of human intelligence is still an ongoing process, i.e., human and information security are not yet perfectly matched. By understanding cognitive processes, designers can design humanized cognitive information systems (CIS). The need for this research is justified because today's business decision makers are faced with questions they cannot answer in a given amount of time without the use of cognitive information systems. The researchers aim to better strengthen cognitive information systems with more pronounced cognitive thresholds by demonstrating the resilience of cognitive resonant frequencies to reveal possible responses to improve the efficiency of human-computer interaction (HCI). A practice-oriented research approach included research analysis and a review of existing articles to pursue a comparative research model; thereafter, a model development paradigm was used to observe and monitor the progression of CIS during HCI. The scope of our research provides a broader perspective on how different disciplines affect HCI and how human cognitive models can be enhanced to enrich complements. We have identified a significant gap in the current literature on mental processing resulting from a wide range of theory and practice.

Keywords: Cognitive-IoT; human-computer interaction; decision making

1 Introduction

In Cognitive Information Systems (CIS) and machine models, the exchange of information in the domain of human cognition and human-computer interaction leads to many problems. To develop goal-oriented strategies, independent research requires organizational and environmental analysis to understand human thinking and cognition. Starting from human cognition, our goal and proposal is to discover potential links in Human-Computer Interaction (HCI) that can improve relationship

Appendix-II (Research IEEE/Scopus Conference Paper "Real time data evaluation with wearable devices: An Impact of Artifact Calibration Method on Emotion Recognition"

2021 International Conference on Computing Sciences (ICCS)

Real time data evaluation with wearable devices: An Impact of Artifact Calibration Method on Emotion Recognition

Fayaz Ahmad Fayaz
School of Computer Science and Engineering
Lovely Professional University (LPU), Jalindhar India &
Scientist-C at NIELIT J&K, Srinagar
erfayaz009@gmail.com

Dr. Arun Malik
Associate Professor
School of Computer Science and Engineering
Lovely Professional University Jalindhar India
arunmalikhisar@gmail.com

2021 International Conference on Computing Sciences (ICCS) | 978-1-6654-9445-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCS54944.2021.00038

Abstract- ***Smartwatch technology is transforming the environment of transmission and monitoring for stakeholders and research participants who want to provide real-time data for evaluation. A range of sensors are available in smartwatches for gathering physical activity and location data. Here, combining all of these elements allows the collected data to be sent to a remote computer, allowing for real-time monitoring of physical and perhaps emotional development. Photoplethysmography is an easy and economical optical sensing technology that is commonly used to assess heartbeats. PPG is a non-invasive device that measures the volumetric fluctuations of blood flow using a light source and a sensor at the top layer of skin. Models concerning HRV (Heart Rate Variability) analysis are being studied in various domains, including human emotion recognition (HER). Smartwatches as sensor-based devices play an essential role as photoplethysmographic (PPG) data are frequently evaluated for this assessment. However, the nature of these waves (in terms of additional interruptions) may not always be flawless, even though they are susceptible to many factors, such as motion artifacts, light sources, stress distribution, ethnic background, or circumstances. Here techniques for antique rectification play a significant role &, as a response, impact the outcome. This research proposes a novel data distortions mitigation strategy for improving emotion detection classification efficiency using photoplethysmography waves throughout auditory invitation & a Support Vector Machine (SVM) model. Compared to previously undertaken data using a conventional toolset, i.e.,48.81, the presented scheme provides an improved categorization in trigger sensing, i.e., 68.75 percent. An alternative indicator, such as electroencephalographic activity, could be used in conjunction with PPG for further improvement.***

Keywords – Smartwatch as Sensors, HRV, PPG, human emotion, support vector machine (SVM).

I. Introduction

Human emotional responses are related to temperament, disposition, personality, and passion towards an event or a specific task. Biological and physiological behaviors can be represented in various ways, including facial expressions, vocals, text, gestures, and bio-signals.

Despite advances in human-computer interaction (HCI), It's still challenging to spot human emotional responses, making it challenging for robots to give the idea that machines understand people. [1][2]. High cholesterol, cognitive and emotional anxiety, hypertension, diabetes, maternity, and emotional responses can all be assessed using Heart Rate Variability (HRV) measurement, which evaluates the physiological fluctuation of the pulse rate due to sympathetic transmission.[3][4]. Physiological parameters such as electroencephalography (EEG), electromyograms (EMGs), photoplethysmography (PPGs), inhalation patterns (RSP), an electrocardiogram (ECG) are crucial criteria for emotion detection within sensory data because they are impulsive actions[5][6]. Previous research has used HRV analysis to promote understanding caused by emotional sounds, particularly in terms of diversity in assessing inter-beat-intervals (IBIs) [7]. Because PPG is prone to dynamic disturbances, using an appropriate artifacts reduction approach is critical. For this outcome we may have basic three basic action plans to be identified: removal, interpolation (using various methods such as nearest neighbours, cube smoothing, or piece-wise cubic Hermite), and intelligence efforts. The rectification of incorrect IBI readings can be done by taking into account the surrounding IBIs over a short period [8],[9],[10]. The PPG sensors can be placed in the fingers, ears, forehead, and wrist to evaluate human parameters. It's challenging to estimate human parameters effectively since the wrist is particularly vulnerable to body motions that distort the PPG signal, which determines the performance[11][12]. The overall goal of this study is to increase the efficiency of sentiment identification systems that use features retrieved from HRV analysis to perform categorization. As a result, this research proposes a new method for rectifying aberrations in IBI time series that involves substituting the overall average of the beats preceding and following the stated variables for absent sounds. During studies generating emotions using audio samples from the IADS-2 databases, IBI signals from a PPG sensor (Empatica E4) were taken into account. The impact of the artifact calibration method on the data was then assessed by comparing it to the Kubios toolbox's artifact repair strategy [13][14]. The Support Vector Machine (SVM) algorithm had been used to detect the presence or absence of stimuli, using as input parameters those found with statistically significant using a baseline t-test. A forearm, multimodal smartwatch (Empatica E4) has been utilized for the investigation to capture a range of mental parameters from which features are extracted for feeding to detect an audio stimulus. Fig. 1 illustrates the planned execution process.

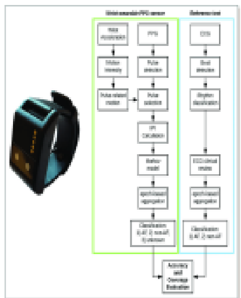


Fig. 1: illustrates the execution process of PPG sensing flow.

978-1-6654-9445-8/21/$31.00 ©2021 IEEE
DOI 10.1109/ICCS54944.2021.00038

Appendix-III (Research IEEE/Scopus Conference Paper “Cognitive Internet of things (CIoT) a success for data collection”

2021 Sixth International Conference on Image Information Processing (ICIIP)

Cognitive Internet of things (CIoT) a success for data collection

Fayaz Ahmad Fayaz
National Institute of Electronics & Information Technology, Lovely Professional University (LPU), India
erfayaz009@gmail.com

Arun Malik
Lovely Professional University, India
arunmalikhisar@gmail.com

Arshad Ahmad Yatoo Lovely Professional University,
Jalandhar India
arshadyatoo@gmail.com

2021 Sixth International Conference on Image Information Processing (ICIIP) | 978-1-6654-3361-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIIP53038.2021.9702706

Abstract-In conjunction with data generated by intelligent machines, cognitive IoT uses cognitive computing technology and the actions these devices can accomplish. The Cognitive Internet of Things (CIoT) is seen as the new IoT is combined with mental and mutual frameworks to facilitate success and intelligence. This leading research area has recently emerged as intelligent sensing. Researchers examine the sensing data performance problems with Smarter technologies, in which people usually use smart gadgets, contribute training datasets towards the Cognitive Internet of things collected by sensors. Moreover, Cognitive Intent of Things (CIOT), shortcomings in the scope of sensing data, contribute to the loss of human life and civil instability. To answer this problem, we propose a new metric in this article, called the Quality of Information Coverage (QIC), which will personify information distribution and data sensing incentives to leverage the QIC. In addition, a market-based compensation system is being developed to pledge the QIC. To produce optimum kickbacks for CIoT and news outlets, we evaluate the optimal business solution and examine an acceptable representation. Then, by detailed computations, the results of a competition reward system are studied. The findings suggest that the way the method of reward management hits the balance point with a greater QIC than most current systems. The QIC told a system in this work guarantees that, relative to existing algorithms, the sample variance number obtained datasets for specific regions decreases by approximately less than 40 to 55 percent since these data sets are calibrated. Compared to these non-QIC-aware algorithms, the average sale price is Sensing proposed should be less than 17 to 18 percent.

Keywords - CIOT, AI, Ubiquitous Computing (UC), Quality Information, Information Sensing Application (ISA).

I. INTRODUCTION

The cognitive Internet of Things application uses technologies for computation originating from brain psychology, artificial intelligence, and knowledge, in association with information produced by the linked devices and the behaviour that such devices execute [1]. The pervasiveness of touch screen devices [2][3], CIoT systems collect data at a cheap cost and provide a new framework for tackling complex sensing applications that require cognitive computing, such as autonomous vehicles, communications networks, and environmental and weather tracking systems [4][5]. To update the IoT ecosystems, cloud storage and big data are essential. Innovative features such as cognitive processing, artificial learning, Information extraction, analytical thinking, and processing of natural languages can imitate or improve human competence [6]. Consequently, one of the relevant works of literature is to make this form of machine intelligence, through integrating enough "artificial intelligence "as it should, to provide us with the best service of the hour. Waze [7], which emits ubiquitous network traffic, Weather-Lah [8], which produces so well reality on the ground, and Noise-Tube [9], which creates noise charts, are standard information frameworks. The handling of the vast amount of data produced by cognitive Internet of things sensors &equipment has also been widely analyzed among advanced systems as its heroes' helpful knowledge. Because the costs of implementing and servicing sensing mechanisms in apps for intelligence are high due to the broader spectrum of sense, these systems frequently employ a list of figures, part-time, region-price, and it is possible to generate meaningful data. After the media outlet of collected data samples by Intelligence Wearable Electronics & an extensive range of data sampling. There is massive potential for static and mobile sensing systems [10] to produce applications that are much more capable than Systems that are present by understanding and injecting more human intelligence into technologies and ecosystems. CIoT increases the precision and efficacy of sophisticated, sensor-driven solutions. Intelligence Sensor is a significant challenge for CIoT production to reach a high Level of Information Collection (QIC). Waze[7], for instance, is a community-based traffic and navigation technology that helps vehicles exchange traffic congestion data in real-time—saving travel time prices in a town with all other cars.
Knowledge sensing instruments in CIoT mean details transmitting knowledge from the immediate neighborhood, then the same amount is payable to Information Sensing Applications (ISA) to expenses with reporters. However, the intensity of perception intelligence In CIoT, devices become inconsistent. In the region where the information is in CIoT, sensing instruments are densely spaced, collecting more data to disseminate more data to the ISA. Nevertheless, intelligence detection devices in CIoT are more widely

Appendix-IV (Research Scopus Paper "Novel Hybrid Emotion Recognition Framework (NHERF) for Virtual Learning Assessment Model a Review"

Solid State Technology

Home Current Aims and Scope For Authors Archives Ethics & Policies About Search

Home / Archives / Vol. 63 No. 2s (2020) / Articles

Novel Hybrid Emotion Recognition Framework (NHERF) for Virtual Learning Assessment Model a Review.

Fayaz Ahmad Fayaz, Dr. Arun Malik

PDF

Issue
Vol. 63 No. 2s (2020)

Section
Articles

Abstract

Covid-19 has taught us how to learn from home via with the advancement in ICT, the concept of virtual learning and its applications like virtual classroom is a shared online space where the learners and the tutor work together simultaneously and participants take the advantage to attend

Appendix-V (Research IEEE/Scopus Conference Paper "Novel Face Recognition Based Examinee Authentication System using Python D-Lib"

2019 Fifth International Conference on Image Information Processing (ICIIP)

Novel Face Recognition Based Examinee Authentication System using Python D-Lib

Fayaz Ahmad Fayaz
Deputy Director (Scientist-C) &
Research Scholar, LPU Jalindhar, India.
National Institute of Electronics &
Information Technology (NIELIT)
Srinagar/Jammu , Govt. of India.
fayaz@nielit.gov.in

Shakir Mohi-ud-din
Project Student, National Institute of
Electronics & Information Technology
(NIELIT),Srinagar/Jammu, Govt. of
India.
shakirmattoo@gmail.com

Irtiza Batool
Project Student, National Institute of
Electronics & Information Technology
(NIELIT) Srinagar/Jammu, Govt of
India.
irtizabatool@gmail.com

Satinder Kaur
Project Student, National Institute of
Electronics & Information Technology
(NIELIT) Srinagar/Jammu, Govt. of
India.
satinder.kaur3291@gmail.com

Mamoon Rashid
Assistant Professor, School of Computer
Science & Engineering, Lovely
Professional University, Jalandhar, India.
mamoon873@gmail.com

***Abstract* – *In the fast growing world of Information & Communication Technology (ICT),* authentication of human faces with minimum human intervention has a greater demand in multiple sectors. As HFR (Human Face Recognition is a well-known technique for authenticating users or stake holders. Biometric as a dominant branch for verification, HFR has its applications in video monitoring / surveillance systems, HCI. This paper is proposed to present a method for recognizing an Examinee using the technique called face recognition wherein Python programming D-Lib & other associated libraries / dependencies are used. This model Biometric system which requires a validation machine.**

Keywords— Face-Detection; Face Recognition; Examinee Authentication; Blink Detection; Depth Sensor; Infrared Beam

I. INTRODUCTION

The objective of this article is to present a model which furnishes an easier & user friendly human-machine interaction to authenticate the Examinee by his facial features which are extracted using the photo which he/she submitted during the registration process of the exams. A machine can detect and recognize a person's face; resulting in the authentication of the candidate. A customized login screen having the potential to scrutinize user's access will be developed for Admin module with the features of facial recognition.

The goal of this work is to present a set of Programs that can be later packaged in an easily portable framework amongst the different processor architectures, which we see in machines (computers) today. It has become more difficult for face detection because of some unpredictable attributes. For example, with changing features like eyeglasses and a beard will have influence for detecting effectiveness [1]. Furthermore, distinct angles of lighting will track face generates dissimilar brightness on the face, which will hold the identification process. To generate a code for correct & genuine facial recognition to have efficient use of hardware, a well inbuilt libraries & a comprehensive study of D-Lib platform was formulated [2]. The project, in general, has a lot of applications in various varying ways using face recognition to authenticate, however, our scope of research is focused on using the face recognition to authentication of the Examinee with minimal human interaction. This project focuses on the interaction of screen frames obtained from the live video feed and determining whether the captured frame has a human face and then face-recognition is triggered after human face detection. This Project is tested on photo Gallery as well as a live camera feed.

II. BACKGROUND

For evaluation of mass education, web based examination model is quite beneficial [3]. The model was used for the evaluation and authentication of basic skills of computer science. Authentication of students is one of the major challenge for examinations which are taken in online mode [4]. This research has used profile based authentication framework along with credentials of username and password for solving the challenges of virtual learning authentication. For user authentication in online examinations, three main questions in terms of system, authentication methods and threats are discussed [5]. This research provides classification of authentication methods for online examination systems. Secure authentication scheme for face images is proposed with the use of hash functions [6]. This research discusses about the design issues in face images in terms of security, scalability and collision-freeness. A challenge of authenticity of users

Appendix-VI (ICMCER-2020 Conference Certificate)

CERTIFICATE

OF PARTICIPATION

In Association with

IFERP

connecting engineers... developing research

Organized by

1ST INTERNATIONAL CONFERENCE ON

MULTIDISCIPLINARY AND CURRENT EDUCATIONAL RESEARCH

(ICMCER-2020)

30TH - 31ST JULY 2020 | VIRTUAL CONFERENCE

This is to certify that Fayaz Ahmad Fayaz of NIELIT J&K, Govt. of India & Lovely Professional University (LPU) Jalandhar Punjab India presented his/her research paper titled "Novel Hybrid Emotion Recognition Framework (NHERF) for Virtual Learning Assessment Model a Review" in the "1st International Conference on Multidisciplinary and Current Educational Research (Virtual Conference)" organized by Mahachulalongkornrajavidyalaya University, Ayutthaya, Thailand held on 30th - 31st July 2020.

ICMCER 2020

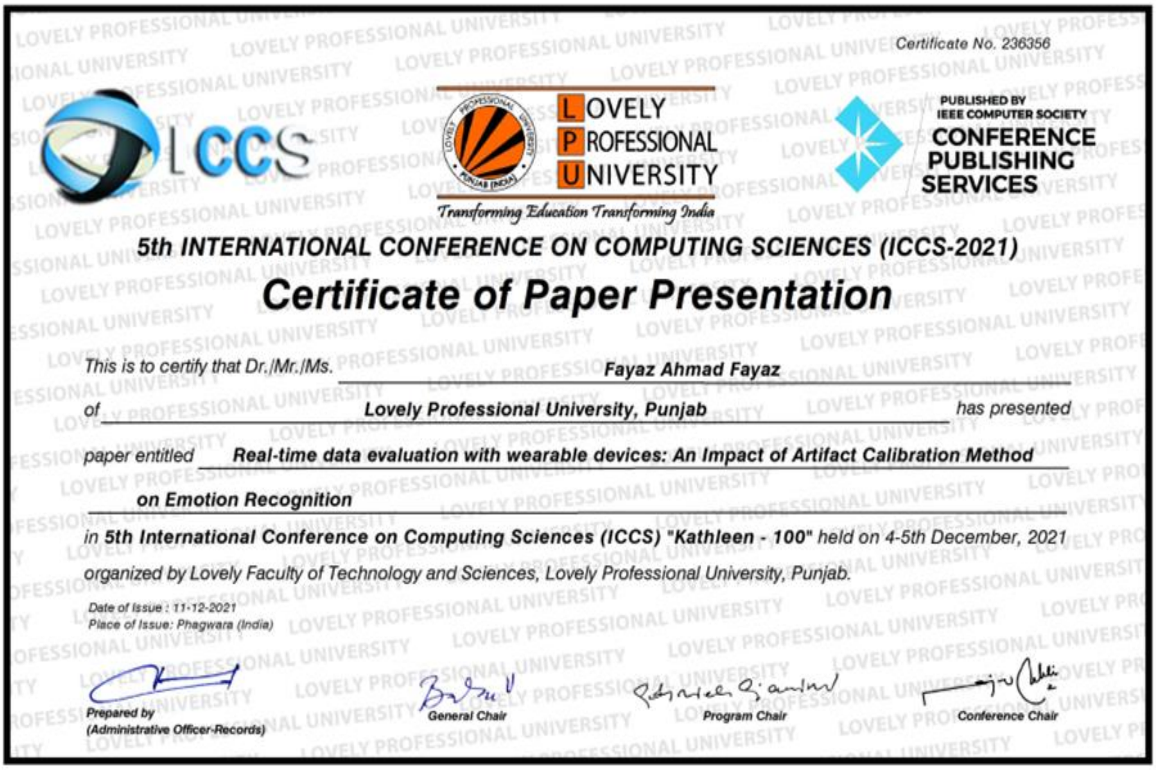
Dr. Phraratsutaporn
Associative Professor & Dean
Faculty of Education
Mahachulalongkornrajavidyalaya University
Thailand

Dr. Lampong Klomkul
Acting Director for Research, Information and Academic Services, ASEAN Studies Centre
Mahachulalongkornrajavidyalaya University
Thailand

IFERP

Mr. Rudra Bhanu Satpathy
CEO & Founder
IFERP

Appendix-VII (ICCS-2021 Conference Certificate)



Certificate No. 236356

ICCS

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

PUBLISHED BY IEEE COMPUTER SOCIETY

CONFERENCE PUBLISHING SERVICES

5th INTERNATIONAL CONFERENCE ON COMPUTING SCIENCES (ICCS-2021)

Certificate of Paper Presentation

This is to certify that Dr./Mr./Ms. ***Fayaz Ahmad Fayaz***

of ***Lovely Professional University, Punjab*** *has presented*

paper entitled ***Real-time data evaluation with wearable devices: An Impact of Artifact Calibration Method on Emotion Recognition***

in ***5th International Conference on Computing Sciences (ICCS) "Kathleen - 100"*** *held on 4-5th December, 2021*

organized by Lovely Faculty of Technology and Sciences, Lovely Professional University, Punjab.

Date of Issue : 11-12-2021

Place of Issue: Phagwara (India)

Prepared by (Administrative Officer-Records) | General Chair | Program Chair | Conference Chair

Appendix-VIII (ICIIP-2021 Conference Certificate)

ICIIP2021/IEEE/CRN-53038/CPP/Paper ID-1570764179

juit

iciip

IEEE

Certificate of Participation

This is to certify that Prof. / Dr. / Mr. / Ms. Fayaz Ahmad Fayaz

has participated and presented a paper entitled :

Cognitive Internet of Things (CIoT) a Success for Data Collection

in **2021 Sixth International Conference on Image Information Processing (ICIIP 2021)** organized by the Department of Computer Science & Engineering and Information Technology at Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, INDIA, during 26 - 28 November, 2021.

Dr. Aman Sharma **Conference Co-chair** **ICIIP-2021**	**Dr. Vipin Tyagi** **Conference Chair** **ICIIP-2021**	**Dr. Ekta Gandotra** **Executive General Chair** **ICIIP-2021**	**Dr. P.K. Gupta** **Principal General Chair** **ICIIP-2021**