# Modelling and Application of Biomarkers for Affective State Mining Using Soft Computing Techniques

A THESIS
SUBMITTED TO DELHI TECHNOLOGICAL UNIVERSITY
FOR THE AWARD OF THE DEGREE OF

## DOCTOR OF PHILOSOPHY

**In**
**Computer Science & Engineering**

**by**
**Aditi Sharma**
**(2K18/PhDCO/505)**

**Under the Supervision of**

**Prof. Kapil Sharma**
**Delhi Technological University, India**

**&**

**Dr. Akshi Kumar**
**Manchester Metropolitan University, UK**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**DELHI TECHNOLOGICAL UNIVERSITY**
**(Formerly Delhi College of Engineering)**
**DELHI- 110042 (INDIA)**
**2023**

**COPYRIGHT NOTICE**
© DELHI TECHNOLOGICAL UNIVERSITY, DELHI 2023
**ALL RIGHTS RESERVED**

# CANDIDATE'S DECLARATION

I, hereby certify that the thesis titled **"Modelling and Application of Biomarkers for Affective State Mining Using Soft Computing Techniques",** submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy' is an authentic record of my research work carried out under the guidance of Prof. (Dr.) Kapil Sharma, Delhi Technological University, India & Dr. Akshi Kumar, Manchester Metropolitan University, UK. Any material borrowed or referred to is duly acknowledged.

The matter presented in this thesis has not been submitted elsewhere in part or entirely to any other university or Institute for award of any degree.

**ADITI SHARMA**
2K18/PHDCO/505
Department of Computer Science and Engineering
Delhi Technological University

# DELHI TECHNOLOGICAL UNIVERSITY

(Govt. of National Capital Territory of Delhi)

BAWANA ROAD, DELHI-110042

# CERTIFICATE

Date: 29 March 2023

This is to certify that the work embodied in the thesis titled "*Modelling and Application of Biomarkers for Affective State Mining Using Soft Computing Techniques*" has been completed by Aditi Sharma, Roll No. 2k18/PHDCO/505 under our supervision and guidance towards fulfilment of the requirements for the degree of Doctor of Philosophy of Delhi Technological University, New Delhi, India. This work is based on original research and has not been submitted in full or in part for any other diploma or degree of any university.

Supervisors

| | |
|---|---|
| Prof. Kapil Sharma | Dr. Akshi Kumar |
| Professor | Senior Lecturer |
| Dept. of Information Technology | Dept. of Computing & Mathematics |
| Delhi Technological University | Manchester Metropolitan University |

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my Ph.D. supervisors, Prof. Kapil Sharma and Dr, Akshi Kumar for constantly encouraging me and giving me time and unconditional support during my research work. Their expertise was invaluable in formulating the research questions and methodology. Their insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I will always be indebted to them, for their wise counsel.

I am very grateful to Prof. Rajni Jindal, Chairperson (DRC) and Prof. Vinod Kumar, Head, Department of Computer Science & Engineering, DTU, for their constant encouragement and support to accomplish this task.

I am thankful to all the faculty members of the Department of Computer Science & Engineering, and the members of SRC for the motivation and right guidance. My heartfelt gratitude to Dr. MPS Bhatia, the SRC(Expert) for his suggestions and right direction for the successful completion of my dissertation.

I am thankful to my family for their love and support, all of whom have made my accomplishments possible. I am also thankful to all my colleagues, who had supported my work, by giving genuine feedbacks.

**Aditi Sharma**
2K18/PhDCO/505
Department of Computer Science & Engineering,
Delhi Technological University
New Delhi-110042
*E-mail:* aditisharma_phdco2k18@dtu.ac.in

# ABSTRACT

The craving to succeed in this fast-paced life is impacting the stress levels of everyone. Stress is a psychological condition in which a person feels overwhelmed with pressure. The body reacts to these changes with physical, mental, and emotional responses, these changes, the feelings that a person experience is their affective state. These changes can be observable or non-observable from the outside view. The measures of these changes are biomarkers, which can be of physiological, visual, electrochemical, or psycholinguistic in nature. Medical practitioners can measure and understand these biomarkers and can analyze the psychological and emotional state of a person. Analyzing the biomarkers manually is a tedious and time-consuming task for trained medical practitioners, delaying early identification and timely intervention. With the availability of IoT based sensors for healthcare, these biomarkers can be monitored using various wearable devices, implants, and cameras. Motivated by the need to design a model for affective state recognition using sensor-based bio-signals, this research proffers multi-model deep learning-based models for affective state recognition for identifying psychological as well as emotional state, using different fusion strategies to combine different modalities of biomarkers.

Upsurge in IoMT devices has made the remote healthcare a reality. The bio-signals of patients can be monitored from remote location by medical professionals, making it possible for everyone to have access to healthcare. Though the benefits offered are unparalleled and promise useful decision support information, but all this is challenged by a lot of noise created owing to the large volume and variety of information sent at almost light speed. One of the most significant threats that the IoMT poses is of data security & privacy. As these devices capture and transmit data in real-time with no standard data protocols and data ownership regulations, it makes it highly susceptible to hacks and frauds. To resolve these issues, two novel methods has been proposed to reduce the size of the data using genetically optimized fuzzy c-means clustering, and Federated transfer learning approach for proving privacy to the subjects, by training the model in decentralized environment. The preliminary results have shown promising results for both resolution techniques. This research also establishes the relation between cause and its effect on human affective state, by proposing a causal affective theory, and validating it with the help of a case study on Indian students during COVID-19.

## CONTENTS


**Page No.**

| | |
|---|---|
| **Declaration** | i |
| **Certificate** | ii |
| **Acknowledgements** | iii |
| **Abstract** | iv |
| **List of Tables** | ix |
| **List of Figures** | x |
| **List of Abbreviations** | xii |

**CHAPTER 1: INTRODUCTION** **1-15**

| | |
|---|---|
| 1.1 MOTIVATION & SCOPE | 4 |
| 1.2 AFFECTIVE STATE MINING | 5 |
| 1.3 BIOMARKERS | 7 |
| 1.4 SOFT COMPUTING TECHNIQUES | 9 |
| 1.5 STATEMENT OF RESEARCH QUESTION & RESEARCH OBJECTIVES | 10 |
| 1.6 ORGANIZATION OF THESIS | 14 |
| 1.7 CHAPTER SUMMARY | 15 |

**CHAPTER 2: LITERATURE REVIEW** **16- 37**

| | |
|---|---|
| 2.1 AFFECTIVE PSYCHOLOGICAL STATE | 17 |
| 2.2 AFFECTIVE EMOTIONAL STATE | 22 |
| 2.2.1. EMOTION THEORIES | 23 |
| 2.2.2. DISCRETE & DIMENSIONAL PERSPECTIVE OF EMOTIONS | 24 |
| 2.2.3. DATASETS | 31 |
| 2.3 AFFECTIVE BIOMARKERS | 33 |
| 2.4 MAJOR FINDINGS | 35 |
| 2.5 CHAPTER SUMMARY | 37 |

**CHAPTER 3: MULTI-MODAL AFFECTIVE STATE RECOGNITION 38-74**


3.1 AFFECTIVE PSYCHOLOGICAL STATE 38

3.1.1. HIERARCHAL DEEP NEURAL NETWORK MODEL- PHYSIOLOGICAL BIOMARKERS 39

3.1.1.1. METHODOLOGY 40

3.1.1.2. DATASET 40

3.1.1.3. PRE-PROCESSING 41

3.1.1.4. FINDINGS 41

3.1.2. ANXIOUS DEPRESSION- PSYCHOLINGUISTIC BIOMARKERS 43

3.1.2.1. METHODOLOGY 43

3.1.2.2. DATASET 44

3.1.2.3. PRE-PROCESSING 46

3.1.2.4. FINDINGS 47

3.2 AFFECTIVE EMOTIONAL STATE 48

3.2.1. EmoHD: EMOTION HEALTH DETECTION – AUDIO, VIDEO & LINGUISTIC MODALITIES 49

3.2.1.1. METHODOLOGY 49

3.2.1.2. DATASET 50

3.2.1.3. VISUAL MODALITY 51

3.2.1.4. AUDIO MODALITY 53

3.2.1.5. TEXT MODALITY 54

3.2.1.6. WEIGHTED DECISION FUSION 55

3.2.1.7. FINDINGS 57

3.2.2. MEmoR: MULTI-MODAL EMOTION RECOGNIITON – VIDEO AND PHYSIOLOGICAL MODALITIES 60

3.2.2.1. METHODOLOGY 60

3.2.2.2. DATASET 61

3.2.2.3. PHYSIOLOGICAL MODALITY 61

3.2.2.4. FINDINGS 62

3.2.3. DREAM: DEEP LEARNING BASED RECOGNITION OF

EMOTIONS 65
3.2.3.1. METHODOLOGY 65
3.2.3.2. DATASET 66
3.2.3.3. AUDIO MODALITY 68
3.2.3.4. VEDIO MODALITY 69
3.2.3.5. PHYSIOLOGICAL MODALITY 69
3.2.3.6. AVERAGE DECISION FUSION 70
3.2.3.7. FINDINGS 71
3.3. CHAPTER SUMMARY 74

**CHAPTER 4 : ENHANCED MULTI-MODAL AFFECTIVE STATE RECOGNITION 75-93**

4.1. GENETICALLY OPTIMIZED FUZZY C-MEANS DATA CLUSTERING 76
4.1.1. METHODOLOGY 76
4.1.1.1. FUZZY C-MEANS CLUSTERING 77
4.1.1.2. GENETIC OPTIMIZATION 80
4.1.1.3. SUMMARIZATION 81
4.1.1.4. DEEP HIERARCHAL MODELLING 81
4.1.2. DATASET 83
4.1.3. ALGORITHM 83
4.1.4. FINDINGS 84
4.2 FTL-EMO: FEDERATED TRANSFER LEARNING FOR EMOTION RECOGNITION 88
4.2.1. METHODOLOGY 88
4.2.2. FEDERATED TRANSFR LEARNING 89
4.2.3. DATASET 91
4.2.4. PRE-PROCESSING 91
4.2.5. FINDINGS 92
4.3. CHAPTER SUMMARY 93

**CHAPTER 5: CAUSAL THEORY OF AFFECTIVE EXPERIENCE 94-108**

5.1. CAUSAL THEORY 94
5.2. CASE STUDY 96
5.2.1. METHODOLOGY 96
5.2.2. PARTICIPANTS AND C-19 MHQ 97
5.2.3. METHODOLOGY 102
5.2.4. FINDINGS 104
5.3. CHAPTER SUMMARY 108

**CHAPTER 6: CONCLUSION AND FUTURE WORK 109-111**

6.1. FUTURE SCOPE 111

**References 112-118**

**List of publications 119-120**

**Curriculum Vitae 121-127**

# LIST OF TABLE (S)

**Page No.**

| | |
|---|---|
| **Table 1.1.** Mapping of Ros with Publications | **12** |
| **Table 2.1.** Affective Psychological State Detection Models | **17** |
| **Table 2.2.** Affective Emotional State Detection Models | **28** |
| **Table 2.3.** Emotion Recognition Datasets | **31** |
| **Table 3.1.** Results of Proposed Model on WESAD data | **41** |
| **Table 3.2.** Accuracy of different classifiers over WESAD | **42** |
| **Table 3.3.** Lexicon for Anxiety Detection | **45** |
| **Table 3.4.** Classification accuracy of classifiers | **48** |
| **Table 3.5**. Emotion-wise data distribution | **51** |
| **Table 3.6.** F1-Score for varied emotions for IEMOCAP | **57** |
| **Table 3.7.** F1- Score for MELD dataset | **58** |
| **Table 3.8.** Comparison of Models | **59** |
| **Table 3.9**. Valence- Arousal Prediction Results | **63** |
| **Table 3.10.** Discrete Emotion Prediction Accuracy | **63** |
| **Table 3.11.** Accuracy of Emotions | **64** |
| **Table 3.12.** Model Performance for Dimensional States | **72** |
| **Table 3.13.** Model Performance for Discrete Emotional states | **72** |
| **Table 4.1.** Performance of Proposed Model | **86** |
| **Table 4.2.** Performance of FTL-Emo | **92** |
| **Table 4.3.** Comparison of FTL-Emo | **92** |
| **Table 5.1.** CTAE Examples | **95** |
| **Table 5.2**. Characteristics of the Participants | **98** |
| **Table 5.3.** Analysis of job offer to students | **99** |
| **Table 5.4.** Analysis of the Remote Classes | **100** |
| **Table 5.5.** Analysis of the Online Exam | **101** |
| **Table 5.6.** % distribution of participants' affective psychological state | **106** |
| **Table 5.7.** Performance results of the predictive models | **107** |

# LIST OF FIGURE (S)

**Page No.**


**Fig. 1.1.** Categories of psychological disorder **2**

**Fig. 1.2.** IoT based sensors in Health Care 4.0 **4**

**Fig. 1.3.** Sample Anxious Depression tweet **5**

**Fig. 1.4.** Tripartite Model (ABC) of attitude formation in social psychology **6**

**Fig. 1.5.** Scherer's typology of affective states **7**

**Fig. 1.6.** Biomarking using IoMT sensors **8**

**Fig. 1.7.** Soft Computing Techniques **10**

**Fig. 2.1**. Discrete and Dimensional Emotion Models **25**

**Fig. 2.2.** Plutchik's Model and Russell' Model **26**

**Fig. 2.3.** Types of Affective Biomarkers **34**

**Fig. 3.1.** Biomarkers for Stress detection **38**

**Fig. 3.2.** Architecture of Hierarchal Deep Neural Network **39**

**Fig. 3.3.** Accuracy of Subjects **43**

**Fig. 3.4.** Architecture of Anxious-Depression detection model **44**

**Fig. 3.5.** Accuracy of classifiers **47**

**Fig. 3.6.** Performance of proposed AD prediction model **48**

**Fig. 3.7.** EmoHD Architecture **49**

**Fig. 3.8.** Face extraction using MTCNN **52**

**Fig. 3.9.** MLP Network **53**

**Fig. 3.10.** Performance of various models on MELD and IEMOCAP **58**

**Fig. 3.11.** Proposed Model for Discrete Emotions **59**

**Fig. 3.12.** MEmoR Architecture **60**

**Fig. 3.13.** Accuracy of Discrete Emotion vs Valence-Arousal **64**

**Fig. 3.14**. Confusion matrix of the proposed model for discrete emotions **64**

**Fig. 3.15.** Comparison of Models for Discrete Emotions **65**

**Fig. 3.16.** Architecture of DREAM **66**

**Fig. 3.17.** Ablation Study **73**

**Fig. 4.1.** Architecture of Genetically Optimized Fuzzy C-Means Data **78**

Clustering model
**Fig. 4.2.** Architecture of Deep Hierarchical Model **82**

**Fig. 4.3.** Execution time of different clusters **85**

**Fig. 4.4.** Accuracy of different clusters **86**

**Fig. 4.5.** Accuracy of Summarized vs. Non-Summarized Data **87**

**Fig. 4.6.** Execution Time of Summarized vs Non-Summarized Data **87**

**Fig. 4.7.** Architecture of FTL-Emo **89**

**Fig. 4.8.** ROC curve of U7 & U23 **93**

**Fig. 5.1.** The Causal theory of affective experience in triggered situations **95**

**Fig. 5.2.** COVID-19 triggered common mental health issues in students **97**

**Fig. 5.3**. Affective Mental State Prediction Model **103**

**Fig. 5.4.** Random Forest with ‘K’ decision trees **103**

**Fig. 5.5.** Support Received during Pandemic from a social group **104**

**Fig. 5.6.** Activities missed most during the COVID-19 **105**

**Fig. 5.7.** Technologies students want to learn during nationwide lockdown **105**

**Fig. 5.8.** Issues faced by the participants **106**

**Fig. 5.9.** Confusion matrix of artificial neural network **107**

**Fig. 5.10**. Confusion matrix of random forest **107**

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ACC | Three Axis Acceleration |
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| AUC | Area Under Curve |
| BERT | Bidirectional Encoder Representations from Transformers |
| CNN | Convolutional Neural Network |
| DL | Deep Learning |
| DT | Decision Tree |
| IoT | Internet of things |
| KNN | K Nearest Neighbor |
| LSTM | Long Short Term Memory |
| MLP | Multi-Layer Perceptron |
| ML | Machine Learning |
| NLP | Natural Language Processing |
| NLTK | Natural Language Toolkit |
| NN | Neural Networks |
| NB | Naïve Bayesian |
| P | Precision |
| RBF | Radial Basis Function |
| RF | Random Forest |
| RNN | Recurrent Neural Network |
| ROC | Receiver Operator Characteristic |
| RQ | Research Question |
| R | Recall |
| SC | Soft Computing |
| ECG | Electrocardiogram |
| EDA | Electrodermal Activity |

| | |
|---|---|
| RESP | Respiration Rate |
| Temp | Body Temperature |
| BVP | Blood Value Pressure |
| FL | Federated Learning |
| TL | Transfer Learning |
| EMG | Electromyogram |
| WESAD | Wearable Stress and Affects Detection Dataset |
| 1D-CNN | 1 Dimensional Convolution Neural Network |
| EHR | Electronic Health Record |
| IoMT | Internet of Medical Things |
| EEG | Electroencephalogram |

# CHAPTER 1

# INTRODUCTION

The craving to succeed in this fast-paced life takes away the time to overhaul oneself. An individual encounters constant pressure to excel at everything, job stress, nagging by parents and peer pressure leading to increased risk of developing mental health problems. Undeniably, stress is a common problem in modern life psychology. However, mental health is not visible to anyone and it is hard to identify the real mental health status of a person. In contrast, physical health whenever deteriorates has observable signs & symptoms and we seek immediate diagnosis, advice and treatment for it from the healthcare professionals. Moreover, some of the mental health conditions like depression, acute stress syndrome, anxiety and insomnia have common symptoms, so classifying the correct type becomes challenging. For example, stress and anxiety both have symptoms like a dry mouth, sudden sweating, and increased breathing rate. Therefore to classify them as stress or anxiety, the subject needs to be evaluated for an extended period as anxiety attack lasts only for a few minutes and is often triggered by external stimuli.

The physical and psychological impacts can be cyclically linked: emotional distress and poor mental health can trigger or flare a physical health problem and, as a result, cause further distress. Likewise, poor physical health can lead to an increased risk of developing mental health problemzs. A mild amount of stress can be favourable, as it has been observed that a person gives near-optimal work performance under mild-stress. Eustress or beneficial stress [1] is often related to a positive challenge as compared to distress which has negative implications. However, prolonged and chronic stress can severely impact a person's health, affect the whole body and increase the risk of developing certain illnesses. It can have several physical or psychological symptoms, which can make functioning on a daily basis more challenging.

Feeling of uncertainty and panic are natural when a traumatic or gruelling event occurs. But for some people the struggle becomes overwhelming when worries and fears start interfering with relationships and daily life. Formally, mental illnesses are health conditions involving changes in emotion, thinking or behaviour (or a combination of these) [2]. The general cognitive function is hindered to an extent that it can trigger inappropriate responses because those responses are based upon inaccurate thoughts. That is, the person finds it difficult to stay focused, process information, store it in memory, and accurately respond. Mental illness is conceptualized as a clinically significant behavioural dysfunction or psychological syndrome. There many different categories of mental/ psychological disorders defined in the ICD-10, 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO) known as mental and behavioural disorders, ICD codes F00 to F99. Fig.1.1 shows the broad categories of psychological disorders [3].

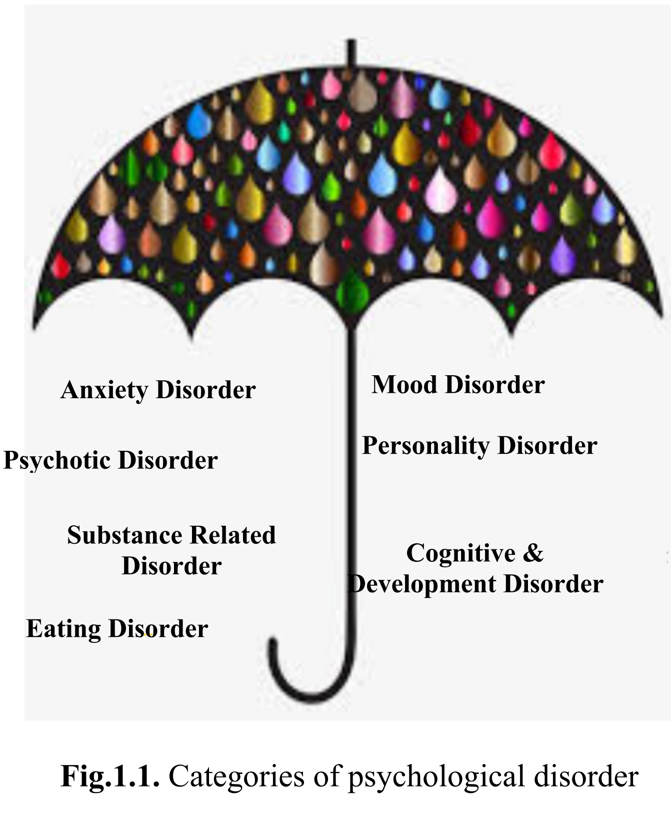



**Fig.1.1.** Categories of psychological disorder

A typical mental status examination (MSE) done by a professional is a standardized procedure used to evaluate the person's mental and emotional functioning. It involves a precise series of observations as well as some specific questions. The complete diagnostic picture is analyzed using seven key evaluation parameters described as:

- Appearance, psychomotor behaviour, and attitude
- Characteristics of speech
- Affect and mood
- Thought content, thought form (Delusions, Illusions, Hallucinations), and concentration
- Orientation
- Memory (long-term, short-term)
- General intellectual level
- Insight and judgement

Events cause emotional reactions which can influence attitudes and behaviours. Affect is the visible reaction a person displays toward events. This characterizes the affective event theory where an emotional episode consists of emotional experiences from an event that affect the mood and emotions cycles. Every individual acts differently when encountered with the same situation and affective state is not dependent on a single attribute. That is, it works differently for different individuals. Therefore, to analyse the affective state of an individual without prior medical history is a difficult task [4]. Various biomarkers can be used to track the affective state of an individual like sleep pattern, level of cortisol and adrenaline hormones, walking pattern, outdoor activities, size of eye pupil, heartbeat rate while performing physical activities and while in the resting period.

The current generation, healthcare 4.0 improves clinical treatment such that medical practitioners can monitor personal health information shared through sensors to be more watchful and connected with the patients proactively. The smart IoT based devices available in the market have helped patient management by remotely monitoring health conditions and timely alerting the hospital about any irregularities using biomarkers on a daily basis [5]. Smart healthcare, as shown in figure 1.2, works both on clinical and non-clinical data.

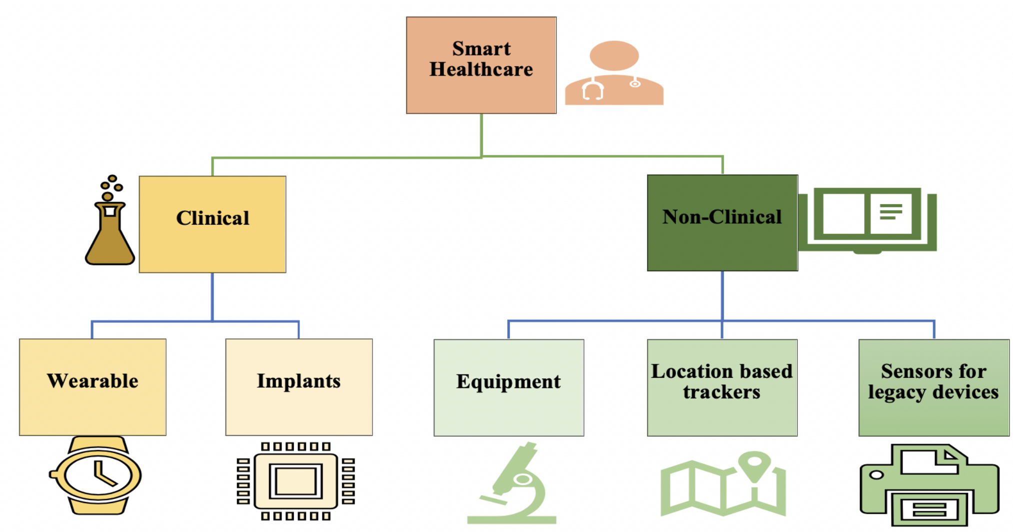



**Fig. 1.2.** IoT based sensors in Health Care 4.0

Clinical trials for any disease require the subject to visit hospital and always be available physically for examination. With the help of IoT-based sensors, the health condition of the user can be tracked remotely using wearable IoT such as a wristwatch, or with the implantation sensor in the subject's body like a pacemaker. The benefits of using IoT in healthcare include, but are not limited to:

- Higher patient engagement
- Better patient outcomes
- Decrease in errors
- Enhanced patient experience

## 1.1. Motivation & Scope

As per the report of WHO, depression is one of the leading causes of disability. Suicide is the second leading cause of death among 15-29-year-olds. People with severe mental health conditions die prematurely – as much as two decades early – due to preventable physical conditions. Clinical trials for any mental well-being require the subject to visit hospital and always be available physically for examination. With the help of IoT-based sensors, the affective state of the user can be tracked remotely using wearable IoT such as a wristwatch, or with the implantation sensor in the subject's body like a pacemaker. In the non-clinical collection of data, the bio-signals of the subject can be traced with the help of their smart devices such as mobile phones, the daily/ monthly activities like walking, running, and sitting to track the health of the user. Indeed, this health data fetched from IoT devices can allow caregivers to make informed decisions and therefore deliver better outcomes.

Simultaneously, social media is omnipresent and allows people to self-express, stay connected and in touch with friends and acquaintances across the globe. Though feelings are hard to articulate but online self-expression provides a means to convey a mental condition into a physical form. Social Media can facilitate pre-diagnosis of a clinical mental health condition related to anxiety, depression or anxious depression in active extroverts who verbalize and share their internal restlessness. The following figure 1.3 depicts a sample anxious depressive twitter post.

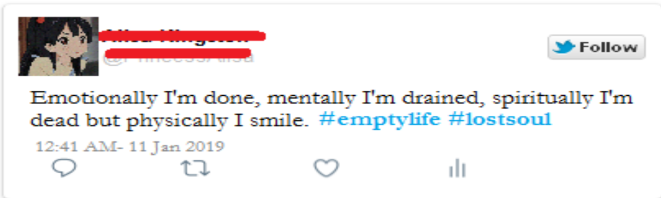



**Fig. 1.3.** Sample Anxious Depression tweet

A growing number of studies examine mental health within social media contexts, linking social media use and behavioural patterns with stress, anxiety, depression, suicidal ideation, and other mental illnesses. Undoubtedly, early depression or disturbed affective state can be detected analysing user's online activity and posted thoughts.

## 1.2. Affective State Mining

Psychology is the scientific study of how people behave, think and feel**.** Social psychology is based on the three basic and interrelated human capacities defined by the ABCs of affect, behaviour, and cognition. That is, in order to effectively maintain and enhance our own lives through successful interaction with others, we rely on these three basic and interrelated human capacities:

- *Affect (feelings)* - the feelings we experience as part of our everyday lives.
- *Behaviour (interactions)* – the way we act and present ourselves in interactions.
- *Cognition (thought)* – what we think and the we way connect our thinking in social world (our cognitive patterns)

Attitudes are a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour. The following figure 4 depicts the Tripartite Model (ABC) of attitude formation in social psychology.

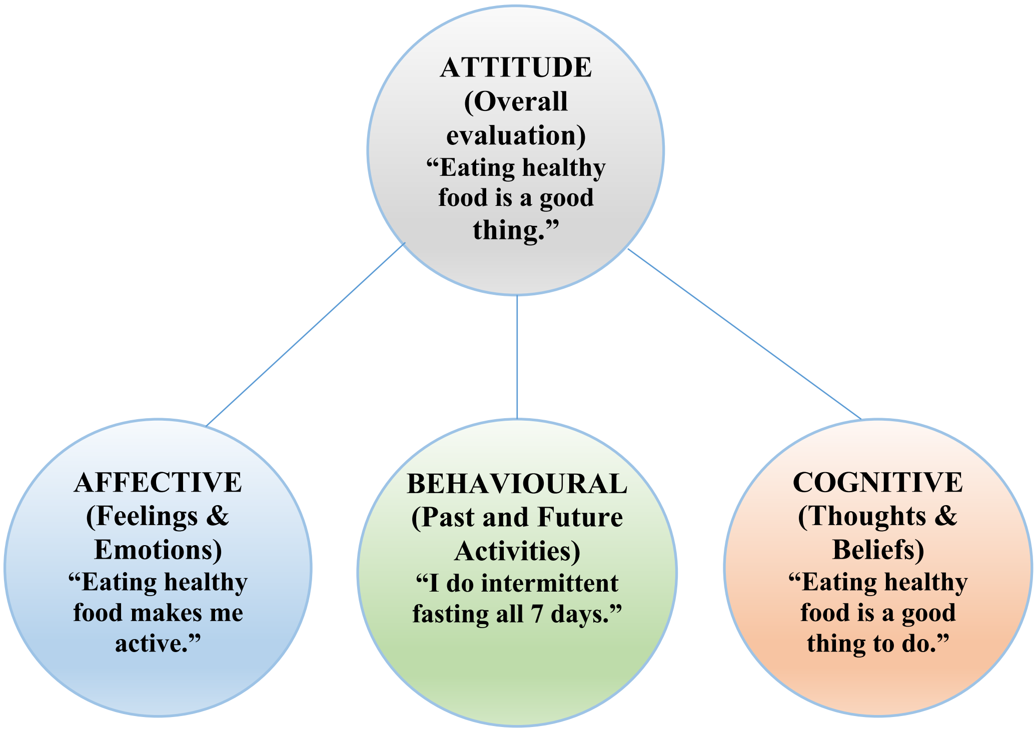



**Fig.1.4.** Tripartite Model (ABC) of attitude formation in social psychology

Cognition is our thoughts, beliefs and ideas about something whereas behaviour involves person's intention to do something that is to perform an action in response to feelings. Affect defines our feelings which we experience in the form of mood and emotions. Emotion is a reaction experienced usually toward a specific object and typically accompanied by physiological and behavioural changes in the body. It is the outward expression of our feelings [7]. Typically, an external event or situation may trigger thoughts and interpretations. This perceived situation (cognition) may further generate affective responses in the form of emotion reactions and bodily state (physiological sensations, facial, speech or body cues), consequently prompting behavioural actions. For example, in a situation where *"family member is late and unreachable on phone"* may trigger thoughts about a mis-happening or accident which may cause physical sensations such as increased heart rate, sweating, stimulate emotions of anxiousness, worry and prompt a frantic calling behaviour.

The Scherer's typology of affective states [8] postulates a generic framework to cognize sentiments. Figure 1.5 outlines the five types of affective states as defined by Scherer et al.

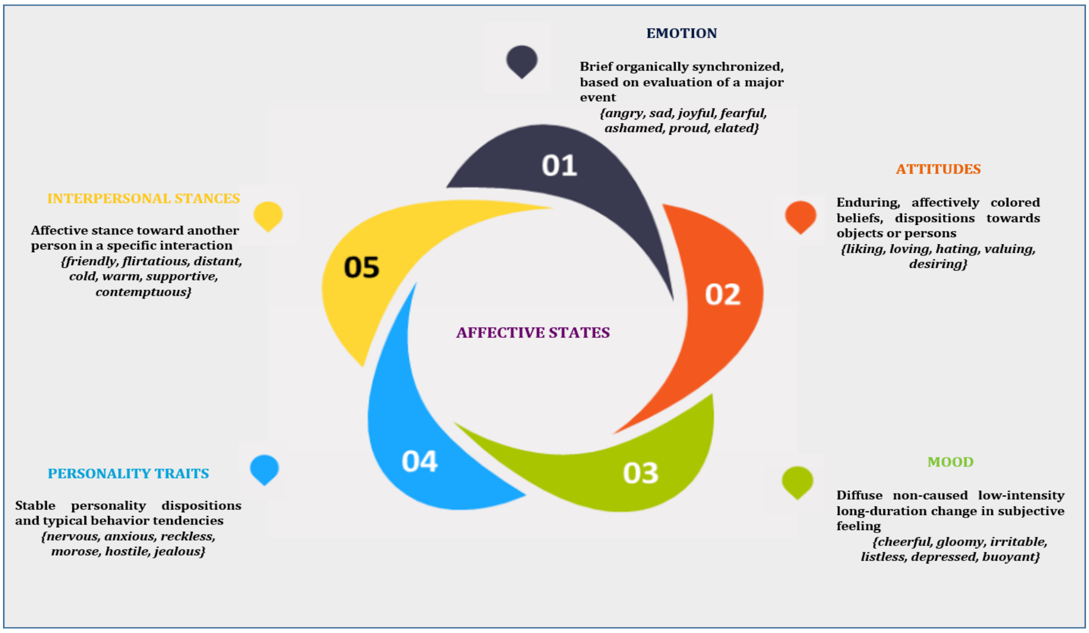



**Fig. 1.5.** Scherer's typology of affective states

Affective computing is the study and development of systems and devices that can recognize, interpret, process, simulate and mediate human affects. Affective state or emotion is a psychophysiological process, triggered by an external (event, object) or internal (memory) stimuli and captured through verbal (voice, tone, text) and non-verbal (facial expressions, body language, physiological indicators) observed manifestations. These modes of expressions or observed manifestations are referred to as biomarkers.

## 1.3. Biomarkers

Digital biomarkers can be derived from various sources such as natural interactions with digital games, social media data and physiological data. The digital biomarkers are revolutionizing the healthcare industry providing endless applications within the structure, such as: timely intervention and diagnosis, improved accuracy proactive treatments, better treatment outcomes. An affective & emotional biomarker quantifies the affective state of user using physiological signals, behavioural signals, speech signals, eye gaze and fixation data, and sentiment analysis of text data. A refined use of biomarkers might be beneficial for predicting the presence of an early disorder that is not yet clinically evident. Some Biomarkers are invasive in nature, some are non-invasive as shown in figure 1.6.

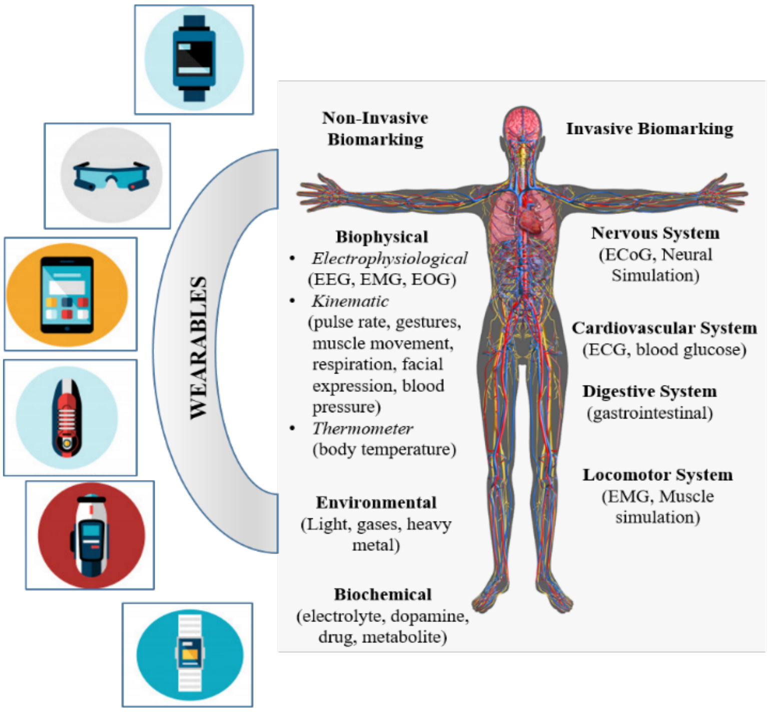



**Fig.1.6.** Biomarking using IoMT sensors

At a facile level, a biomarker (biological marker) is a characteristic of our body that can be measured and evaluated. It acts as an indicator of wellness of body and is a metric that can be derived from various bodily sources such as, changes in the electrical activity of the brain, levels of protein in our body, temperature of the skin and body weight. Some of the biomarkers that can be used for identifying the mental affective state of an individual are as follows:

- **Electrocardiogram (ECG)**: ECG provides the frequency of cardiac cycles. It is sensed using photo detectors, so is not able to be detected by wrist-wearable device.
- **Electromyogram (EMG):** EMG is used to detect musculoskeletal movements. These signals can detect face and hand gestures.
- **Body Temperature (TEMP):** Skin temperature of the subject is measured using thermistor sensor. Body temperature is negatively correlated with stress.

- **Respiration (RESP):** RESP gives the person inhalation and exhalation rate. The slowed respiration rate shows the level of stress in a user.
- **Blood Volume Pressure (BVP):** BVP is the amount of blood in blood tissue during a certain time period. BVP also provides pulse rate and blood flow volume, as it is obtained by photoplethysmography.
- **Electrodermal Activity (EDA):** EDA gives the flow of electricity through the skin. The changes arise in skin when brain sends the signal due to different emotion activation. Skin conduction increases when a person is un- der stress.
- **Three Axis Acceleration (ACC):** ACC gives an indication of different activities like lying, sitting, standing, walking, running and cycling by recording the human movement in all the three dimensions. Fast hand movement over short time depicts sign of mental stress.

With the advancement in healthcare and information technology, these biomarkers can now be measured by having sensors in wearable devices like fitness watch, or a chest wearable or with the help of implants such as pacemaker. But biomarkers in psychology cannot be restricted to electrophysiological signals. Behavioural signals (e.g, posture), speech signals, social interactions and social data can also be used as observable traits that could be indicative of mental health or health decline [10]. Social behaviour is a significant indicator in numerous mental health disorders and therefore imperative for health modelling. Gauging the valence of attitude in written text or spoken words and other cues of social interaction can be used as a critical digital biomarker mental health modelling. Thus, combining these observable traits which are collected and analysed from diverse sources such as smartphone data (e.g, geolocation, accelerometer), social media (e.g, Reddit, Twitter, and Facebook), or physiology from integrated trackers (e.g, Apple watch, Fitbit) may allow accurate predictive modelling of affective states.

## 1.4. Soft Computing Techniques

Soft Computing (SC) techniques are the group of techniques that manifests an emerging approach to computing by providing robust and low cost solutions for modelling the complex real world problems [9]. These are broadly categorized into following five types as shown in figure 7: (i) Machine Learning, (ii) Neural Networks,

(iii) Evolutionary Computation, (iv) Fuzzy Logic, (v) Probabilistic Reasoning. These techniques provide a foundation for intelligently mining this huge gamut of unstructured data available online across various social media. It is so because they tend to explore and exploit the human knowledge such as cognition, recognition, understanding and learning into the fields of computing [10]. This ensued the possibility of building intelligent systems that are autonomous and self-tuned designed systems. SC, thus, provides an opportunity to represent ambiguity in human thinking and deals with the uncertainty in real life by providing most optimal solution. The guiding principle of SC is to exploit the tolerance for imprecision, uncertainty and partial truth to achieve robustness, low-cost solutions [10]. The following figure 1.7 depicts the various SC techniques and their categorization [9].

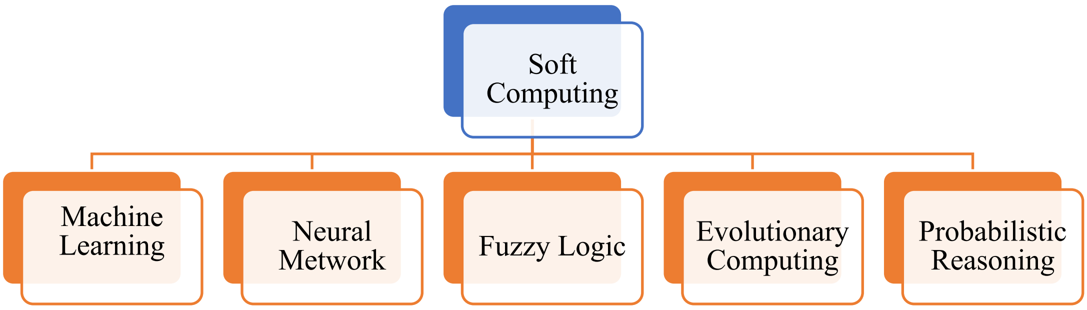



**Fig. 1.7.** Soft Computing Techniques

## 1.5. Statement of Research Question & Research Objectives

This research plans to comprehend the feasibility, scope and relevance of the alliance of using soft computing techniques for affective state mining using digital biomarkers. The goal of this work is to:

- Provide an absolute coverage of the past studies for affective state detection, current trends, public available datasets, and interesting insights into emotional models and theories.
- Develop soft computing-based models for effectively detecting affective emotional and psychological state from different modalities.

The statement of research question is defined as:

***"Which observed manifestations of human affective states can be modelled for mental well-being detection and how soft computing techniques can be applied to detect emotional distress or psychological disorder?"***

In response to the recognized need to exploit new computational models to detect human affective state in real-time, this unifying research question can be broken down into the following four sub-questions, each of which will be addressed by this research:

- What different human physiological changes can be characterized as affective biomarkers, to model the human affective state?
- How can different affective biomarkers be combined to accomplish the task of identifying human affective state to detect emotional distress?
- How can the soft computing techniques be used for real-time detection of psychological disorder?
- Can emotion triggers be associated with human affective psychological state?

This statement of research characterizes three key research objectives as follows:

- **Research Objective 1:** To model multi-dimensional human affective states for emotional intelligence.
- **Research Objective 2**: To propose novel techniques for human emotional and psychological state mining using affective biomarkers.
- **Research Objective 3:** To study casual theory of emotional processing which comprehends relationship between affective state, its causes and effects.

This research primarily aims to identify different biomarkers that can be used to detect the affective state of a person. In this direction, we have identified different biomarkers and categorized them into four types. We have used multi-faceted conceptual modals for recognizing the affective psychological state as well as affective emotional state of a person. Four different multi-modals have been developed during this work on different types of datasets, containing physiological, visual, audio and textual data. To deploy these models in real-time environment to be used for smart healthcare system, we identified two major issues, and provided their solutions as well. First is to reduce time latency due to overload of data, it was resolved by reducing the size of the data, by clustering the data with similar signals in the same time frame, and reducing the size of the data, by replacing each cluster with one summarized data entity. Second major issue identified in making this a real-time application is privacy of the users, to make this model secure for the end users, we proposed the use of federated learning while fine tuning the model, ensuring no data of an individual is shared and stored at a centralized server. Both of these models, help to generate a user-centric, secure and fast real time application that can remotely monitor the affective state of a person, accurately with the help of right sensors at the user end. In addition to this work, A Causal Theory of Affective Experience was proposed that establish a relationship between the event and the affective reaction of a person, the theory was supported with the findings of a questionnaire based survey conducted on Indian students during the Covid pandemic. In this reference, table 1.1 represents the mapping between research objectives and research publication fulfilling the requirement of corresponding aim and query.

**Table 1.1**. Mapping of ROs to Publications

| **Research Objectives (ROs)** | **Publications(s)** |
|---|---|
| **RO1** | • A. Kumar, K. Sharma, **A. Sharma**, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers". *Pattern Recognition Letters*, Elsevier, ISSN: 0167-8655, Vol. 145(2021), pp: 81-87, DOI: 10.1016/j.patrec.2021.01.030 **[SCI, Impact Factor: 4.757].**<br>• A. Kumar, K. Sharma, **A. Sharma**, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers". *Pattern Recognition Letters*, Elsevier, ISSN: 0167-8655, Vol. 145(2021), pp: 81-87, DOI: 10.1016/j.patrec.2021.01.030 **[SCI, Impact Factor: 4.757].**<br>• A. Kumar, K. Sharma, **A. Sharma**, "Hierarchical deep neural network for mental stress state detection using IoT based |

| | |
|---|---|
| | biomarkers". *Pattern Recognition Letters*, Elsevier, ISSN: 0167-8655, Vol. 145(2021), pp: 81-87, DOI: 10.1016/j.patrec.2021.01.030 **[SCI, Impact Factor: 4.757].**<br>• A. Kumar, **A. Sharma**, A. Arora, "Anxious Depression Prediction in Real-time Social Data" in proceedings of *International Conference on Advances in Engineering Science Management & Technology 2019*, Uttaranchal University, Dehradun, arXiv preprint arXiv:1903.10222. **[SSRN].**<br>• A. Kumar, **A. Sharma**, "DREAM: Deep Learning-based Recognition of Emotions from Multiple Affective Modalities using consumer-grade body sensors and video cameras". Communicated in IEEE Transactions on Consumer Electronics, [SCI, Impact Factor: 4.414]. |
| **RO2** | • A. Kumar, K. Sharma, **A. Sharma**, "Genetically optimized Fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics". *Applied Soft Computing*, Elsevier ISSN: 1568-4946, Vol. 109 (2021), pp: 107525, DOI: 10.1016/j.asoc.2021.107525 **[SCIE, Impact Factor: 8.263]**.<br>• A. Kumar, **A. Sharma**, L. Han, R. Ranjan, "FTL-Emo: Federated Transfer Learning for Privacy Preserved Biomarker based automatic Emotion Recognition", Accepted at 4th International Conference on Data Analytics & Management (ICDAM-2023), London Metropolitan University London, 23rd – 24th June 2023. |
| **RO3** | • A. Kumar, K. Sharma, **A. Sharma**, "Empirical Analysis of Psychological Well-Being of Students during the Pandemic with Rebooted Remote Learning Mode", presented in International Conference on Data Analytics and Management (ICDAM-2022), Springer, 2022, Poland, Europe. DOI: 10.1007/978-981-19-7615-5 **[WoS, Scopus].** |

## 1.6. Organization of Thesis

This thesis shows how soft computing techniques can help identifying the affective state of a person using their biomarkers' information. The organization of thesis is follows:

*Chapter 1* will briefly introduce the research that was undertaken. It will discuss the basic concepts and issues of Affective State Recognition. It will cover the motivation and purpose of the outlined research topic. It will also contain the main idea for the development of the thesis.

*Chapter 2* will comprise a detailed state-of-art literature survey required to explore and analyze the existing work on the psychological and emotional affective state, along with the coverage of emotion models and theories. This chapter gives insights into existing datasets that are available for affective state mining, and report gaps and future directions in the said research area.

*Chapter 3* will describe the preliminaries to comprehend the affective state of a human, and discuss the affective biomarkers applicable for identifying the same. This chapter contains the model to identify the affective psychological state of a human from their affective biomarkers- physiological as well as psycholinguistic. This chapter also discusses three different models developed for emotion recognition using deep learning techniques for combination of different modalities in detail with different fusion strategies. The methodologies and findings of research objective 1 will be presented in this chapter.

*Chapter 4* will describe the challenges faced in real-world scenario in detecting affective state. This chapter contains solutions to two different such issues, one to large production of data in real-time, second to ensuring privacy of the users. This chapter discuss both the models in detail. The findings of research objective 2 is primarily covered in this chapter.

*Chapter 5* will describe the causes for change in human affective state. An affective causal theory will be presented in this chapter, that highlight the relation between an event and its effect on human affective state. The theory is supported with the help of a case study conducted during COVID times. The findings of research objective 3 will be presented in this chapter.

*Chapter 6* will present the conclusive results and detailed analysis of the models proposed in chapter 3 (psychological state models), chapter 4 (emotion state models), chapter 5 (Resolving issues of Affective State Models), and chapter 6 (Causal relationship between event and affect) respectively. A brief summary of all the ideas, observations and contributions of the resultants obtained in each research objective will be given. Also, the future directions will be highlighted in this chapter.

## 1.7. Chapter Summary

This chapter is a preface to the research work undertaken for this thesis. It presents the preliminaries of the research statement, the research objectives and the need to have effective methods for identifying affective state in real-time. It contains a brief overview of the terminologies used. It summarizes the organization of the thesis.

***

# CHAPTER 2

# LITERATURE REVIEW

Affective state of a human comprises of multiple factors like their emotional state, stress factor/ psychological state, their attitude, personality type, and mood. Although emotional state, and psychological state majorly indicate the affective state of a person. The psychological state also helps to determine the mental health, whereas emotional state helps in identifying the emotional health of the person. Stress is the body's reaction to any change that requires an adjustment or response. The body reacts to these changes with physical, mental, and emotional responses. Only trained medical practitioners can measure such indicators, which can be tedious and time-consuming, thus delaying early identification and timely intervention. Persistent and chronic stress can lead to a long-term behavioural dysfunction or psychological syndrome. It is thus imperative to design and develop intelligent models to mine affective states and emotional experiences that can support clinicians pro-actively. Biomarkers play an important role in quantifying both mental health and emotional health. Affective state recognition from wearable biosensors and social media complement mood stabilization, stress and depression management, especially for mental well-being.

Various studies has been put forward by many researchers to develop and design artificial intelligence based automated tools that can recognize affective state of a human, both in terms of their emotions, and psychological state. To understand the issues in developing such models, and the limitations of already developed models we have conducted an extensive literature review of studies published in the past decade. As both psychological state and emotional state of a person are different, we have divided our survey into two categories as well, firstly we discuss the studies that has focused on the psychological state, and later we will discuss the studies that has focussed on discussing the emotional state.

## 2.1. Affective Psychological State

Mental health is as vital as physical health. Most of the organizations try to arrange motivational sessions or activity sessions to ensure their employees get a 'mental vacation' and feel relaxed. This is important as a stressed or depressed person will always find it challenging to focus and so will always take extra time to pursue the same work in comparison to a mentally healthy and relaxed person. As the early stages of mental illness have invisible symptoms, it is not always possible to notice the change until the symptoms are persistent, increase in frequency and severity and interfere with life activities and roles. By this time, it becomes too late and in worst cases, untreated mental illnesses can lead to loss of life an average of 25 years early [5]. Thus, early identification and intervention are necessary to recover and reclaim lives. Various artificial intelligence based techniques have been reported to supplement clinical practice in various mental healthcare studies.

To analyse the behaviour of a person under stress, researchers have proposed machine learning-based techniques; however, the availability of dataset has always been an issue. Some of the work conducted by researchers for identifying the affective psychological state of an individual is summarized in table 2.1.

**Table 2.1.** Affective Psychological State Detection Models

| Sr. No. | Author & Year | Dataset | Techniques Used | Result | Description |
|---|---|---|---|---|---|
| 1 | Priya et al. 2020 [11] | Questionnaire DASS 21 | Decision Tree, Random Forest, Naïve Bayes, SVM, K-NN | Accuracy:<br>DT: Anxiety: 0.733<br>Depression: 0.778<br>Stress: 0.628<br>RF:<br>Anxiety: 0.714,<br>Depression: 0.798, Stress: 0.723<br>NB:<br>Anxiety: 0.733,<br>Depression: 0.855<br>Stress: 0.742 | • Data was collected through questionnaire from people of different cultures, financial status, job status.<br>• Anxiety, Stress and Depression was predicted using 5 different machine learning algorithms.<br>• Data was imbalanced.<br>• Random forest classifier has achieved best F1 score. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | SVM: Anxiety: 0.678 Depression: 0.803, Stress: 0.667 | |
| 2 | Halim et al. 2020 [12] | 50 Automotive drivers | SVM, Neural Network, Random Forest | SVM: Accuracy: 97.95% Precision: 89.23% Sensitivity: 88.83% Specificity: 94.92% | • The brain activity was recorded through EEG signal at the time of driving to find a link between brain activity and psychological state.<br>• Psychological state was self-reported by the driver.<br>• SVM, Neural Network and Random forest was applied and evaluated.<br>• SVM performed better to distinguish the rest state from stress state. |
| 3 | Ihmig et al. 2020 [13] | EDA, ECG and RSP bio-signals of 80 Spider-fearful individuals while watching spider videos | K-NN, SVM, Naïve Bayes, Bagged Trees, Decision Trees | Bagged Trees: Accuracy of 2-levels of anxiety: 89.8% And Accuracy of 3-levels of Anxiety: 74.4 | • 80 people having spider fear (Arachnophobia) were shown with 16 1-minute video clips of spiders each with 5 minute resting period.<br>• EDA, ECG and RSP bio-signals were measured through wearable device.<br>• Five different machine learning models were used to train and classify the signals into different levels of anxiety.<br>• Bagged Trees have shown the best promising results. |
| 4 | Penchina et al. 2020 [14] | EEG signal of 13 participants, out of which 8 were Autistic | SVM, Multiclass 2-layer LSTM RNN classifier | SVM: Accuracy: 87.88 LSTM-RNN Model: Accuracy: 93.27% | • Anxiety in individuals suffering from Autism spectrum Disorder(ASD)<br>• Participants were made to sit in a room with dim light, while their EEG signal was being recorded.<br>• A new hybrid deep learning model was proposed to evaluate the EEG signal using EEGNet. |
| 5 | Elgendi et al. | SRAD: 6 Bio-signals | Transitional and Longitu | Accuracy: 75.02% | • Anxiety and stress detection during driving. |

| | | | | | |
|---|---|---|---|---|---|
| | 2020 [15] | of 17 drivers | dinal Analysis : INTENSE( Interaction Ensemble) | | • Evaluated from two different perspective: Transitional and Longitudinal.<br>• For longitudinal: a new model INTENSE proposed.<br>• Compared with PCA, K-Means |
| 6 | Kumar et al. 2020 [16] | DASS 21, DASS 42 | Naïve Bayes, K-NN, MLPNN, RBFN, J48, Random Forest | RBFN: Accuracy: Anxiety: 97.48, Depression: 96.02, Stress: 96.17 | • Eight different machine learning models to find the severity levels of stress, depression and anxiety.<br>• Imbalanced dataset.<br>• Random forest resulted in accuracy to detect anxiety as 100%.<br>• RBFN performed better in all states. |
| 7 | Flesia et al. 2020 [17] | 2053 people data collected through questionnaire | Logistic, SVM, Naïve Bayes, Random Forest | F- Measure: Logistic: High Stress: 0.478, Low Stress: 0.767, SVM: High Stress: 0.465, Low Stress: 0.757, Naïve Bayes: High Stress: 0.479, Low Stress: 0.790, Random Forest: High Stress: 0.480, Low Stress: 0.858, | • Study on 2053 Italian citizens, by collecting the data through questionnaires to know the stress levels induced in the individuals. Four different machine learning algorithm was used.<br>• Random forest among them resulted in highest F-Measure for both High Perceived Stress class, and Low- Perceived Stress class. |
| 8 | Ritcher et al. 2020 [18] | Biomarkers of 125 individuals under controlled environment | Machine Learning | Sensitivity: 71.44%, Specificity: 70.78%, | • Questionnaire survey of 400 individuals, out of which 59 were individuals were of high anxiety.<br>• The 125 individuals were divided into 4 category based upon the anxiety level, and then the different cause and relation were identified for the anxiety levels. |

| | | | | | |
|---|---|---|---|---|---|
| 9 | Can et al. 2019 [19] | 21 individuals data for 9 days collected through Empatica E4. | K-NN, Logistic Regression, Random Forest, Multilayer Perceptron, PCA+LDA, PCA+SVM | Accuracy: PCA+LDA: 82.35 PCA+SVM: 82.35 K-NN: 80.39, Logistic Regression: 90.19, Random Forest: 86.27, Multilayer Perceptron: 92.15 | • A 3-class stress detection system was developed using 6 different type of classifier on data collected from 21 participants over the interval of 9 days through Emperica E4.<br>• Heart Rate, EDA and ACC was captured.<br>• Multilayer perceptron generated the best accuracy of 92.15. |
| 10 | Giannaakis et al. 2019 [20] | SRD'15 : 24 Participants recorded bio-signals | DW1Net1D: 1-Dimensional Deep Wide Convolution Neural Network | Average Accuracy: 89.8 | • A deep- learning multi-kernel architecture is used for recognizing the affective stress states using 6-fold cross validation on the heart rate of 24 individuals through ECG signals. |
| 11 | Islam et al. 2018 [21] | Facebook Users' comments extracted through NCapture | Decision Tree, K-NN, SVM, Boosting | F-Measure: Decision Tree: 0.73, K-NN: 0.67, SVM: 0.73, Boosting: 0.72 | • As social media is used to express your feelings, it can be used for identifying the depression, and stress also.<br>• The level of stress of an individual is detected through the user's comment on Facebook.<br>• Data was analyzed using Emotional Process, Linguistic style and Temporal Process. |
| 12 | Giannaakis et al. 2017 [22] | 23 participants were shown 12 videos of ½ minute to 2 minutes. Stress level at each duration was self- | K-NN, Naïve Bayes, AdaBoost, SVM | Average Accuracy of Emotion Recall: SVM: 65.82, Naïve Bayes: 73.26, AdaBoost: 81.03, K-NN: 88.70 | • The study investigates the use of facial signs for predicting the stress and anxiety levels in an individual.<br>• 23 participants were shown neutral, stressful, and relaxed videos while taking the pictures of the participants at the same time.<br>• The facial images were evaluated on 27 features. |

| | | | | | |
|---|---|---|---|---|---|
| | | declared by the user. | | | |
| 13 | Subhani et al. 2017 [23] | 42 participant study doing performing Montreal Imagined Stress Task(MIST) | Logistic Regression, SVM, Naïve Bayes | Naïve Bayes: Accuracy: 94.6%, SVM: Accuracy: 93.9%, Logistic Regression: 94.0% | • EEG signal of 22 participants were recorded out of 42 participants, while following MIST paradigm.<br>• The experiment was performed under three states: stress, control and rest.<br>• MIST paradigm was used as it can compare the stress condition with a controlled condition of similar nature. |
| 14 | Sau et al. 2017 [24] | Manually collected from Medical college and hospital of 630 Patients(520 special care) | Bayesian Network, Logistic regression, Naïve bayes, Random forest, Sequential random optimization | Random Forest provided best accuracy of 91% for non-special care patients and 89% for special care patients. | • Collected data from medical practitioners for the real patients of the hospital.<br>• Feature selection and classification was applied using Weka tool.<br>• Random Forest works better than other classifiers for both the scenarios. |
| 15 | Gjoreski et al. 2016 [25] | Under Lab observation bio-signals as well as real-life 55 days bio-signals of 5 participants. | Leave One User Out(LOSO) Evaluation, Context vs. No context | Lab Env.: LOSO( with 2 class) Accuracy: 83; LOSO(3 classes) Accuracy: 72. Real Life Stress Detection: No Context Accuracy: 76; With Context Accuracy: 92. | • Continuously monitored the participants in real-life by providing them with an activity recognizer(wrist watch) every 2 minutes.<br>• Participant themselves have to manually update their affective state every 20 minutes<br>• The study was recorded for 55 days.<br>• The evaluation was performed on both real life data as well as laboratory data. |

As, expressed in the table 2.1, different datasets has been used for identifying the psychological affective state of an individual. Different Biomarkers has been used by different researchers for identifying the affective state, but WESAD is considered to be the most suitable one, as it has 7 different physiological signals collected through both wrist-based and chest-based sensors. The latest work on stress detection using WESAD is proposed by [39], they used three classifiers, namely logistic regression, decision tree, and random forest, and rather than evaluating the result into three categories, they added one more output category as meditation. Also, rather than applying each classifier on the complete dataset, they applied it on individual subjects, resulting in an accuracy of 88% to 99% for the individual subject. WESAD is considered as the most recent benchmark dataset to analyse mental health of a person since it contains the maximum number of biomarkers on a single subject to determine the affective state of the subject. Previously, many researchers have created datasets to evaluate the stress level. Picard et al. [26] built a dataset containing physiological data of a single person depicting eight different emotions for 20 days. As different individuals can represent different behaviours for the same emotional feeling; therefore, the dataset collected from a single source cannot predict accurately for all the users. Healey et al. [27] also created a dataset for evaluating stress using ECG, Electrodermal Activity and Respiration, and Electromyogram data. However, this dataset was only used to evaluate the stress of a driver, so it did not apply to all the subjects performing different actions. In 2012, DEAP was published by Koelstra et al.; which contained the facial videos and EEG signals of the users to analyse the emotions using peripheral signals [2]. Although it used multiple subjects, the features used were limited, so DEAP was also unable to predict the emotion of all the users accurately. [29] used mobile phones to construct data for analysing the level of stress in a user. They used biomarkers like physical activity level, social interaction, and social activity along with the location of the user.

## 2.2. Affective Emotional State

Emotion recognition has vast scope in the field of medicine or psychology to identify the emotions of the patients with facial nerve paralysis. Emotion recognition can be used to label the plethora of videos in the internet based on the emotional content and these labels can further be used to auto suggest the videos for the end-user based on their preference or age. The emotion recognition can be applied to background music generation for videos based on the emotional content or for games based on the progress of the player. Though emotion recognition has a lot of potential to improve the quality of life, there is a long way to go for machines that can replicate the emotional intelligence of human beings. The early days of the emotion recognition field were dominated by detecting emotions from facial expressions. Slowly, there have been improvement in the field to recognize the emotions from multiple modalities. Extracting shared representation of different modalities, removing redundant features from different modalities and learning key features from each modality are crucial for multimodal emotion recognition. Table 2 summarizes the different studies conducted on exploring the emotion recognition.

### 2.2.1. Emotion Theories

Basic understanding of theories of emotion is a prerequisite for affective computing as it helps to understand emotional affective state of a person. Moors organized different emotion theories based on the causation of emotions and states that the theories differ in the component they associate the emotion with and the phenomenon they explain [30]. A brief overview of some of the emotion theories is given below:

- ***James' Theory :*** James' theory is one of the earliest emotional theories proposed by William James and Carl Lange [31]. This theory states that stimulus creates physiological responses and the experience of this physiological responses results in experience of emotion. It is one of the highly criticized theory of emotion and was refuted by Cannon [32] through his argument that artificial induction of arousal through injection of adrenalin did not produce any real emotion.
- ***Two-factor Theory :*** Two-factor theory was proposed by Stanley Schachter and Jerome Singer [33]. Two-factor theory states that stimulus creates physiological response and cognitive attribution of this physiological response results in emotion[30].This theory focuses on interaction between physical arousal and how the emotion is labelled cognitively.
- ***Emotion Appraisal Theories*** *:* Multiple appraisal theories have been proposed in the literature. In appraisal theories, the cognitive component is placed at the

onset of emotion and this differentiates it from two-factor theory. According to appraisal theories, the appraisal of stimulus causes an action tendency and this action tendency is manifested in physiological responses which results in behaviour. Emotion is considered as a totality of all the components – stimulus, action, physiological response and behaviour[30].

- ***Network Theories of Emotion** :* The network theory assumes that emotions are stored in memory as schema. A stimulus activates the stored emotional schema based on its proximity to an already experienced stimulus. The schema can be activated by stimulus and response as well [30]. The most influential network theory was proposed by Bower and it states that concepts, events, and emotions are all represented as nodes within the network [35].
- ***Affect Program Theory :*** Affect program theory proposes a hypothesis that each basic emotion has a unique neural circuit and these circuits are installed via evolution to serve specific adaptational functions. Affect program theory is of the view that basic emotions are building blocks of emotional life [30].
- ***Conceptual Act Theory*** **:** Conceptual act theory proposed by Barrett [36] links the emotional experience and emotional perception. This theory states "emotions emerge when physical sensations in the self and physical actions in others are meaningfully linked to situations during a process that can be called as both cognitive and perceptual (creating emotional experiences, and emotion perceptions, respectively)" [36]. According to Russell and Barrett [37] building blocks of emotional life is based on valence and arousal. The affective quality of the stimulus causes the person to be in "core affect" which has neurophysiological side and mental side. Russell and Barrett also state that the basic emotions are nothing but categorization of "core affect".

### 2.2.2. Discrete and Dimensional Perspective of Emotion

Emotional states are conceptualized and described by two different contending perspectives – discrete and dimensional perspective. One argument in favour of dimensional model is that vocabulary to identify all the emotion is limited and varies from language to language and culture to culture [39]. Barrett states that because of the individual differences in emotional state, one static, nomothetic theory may not be suitable for all individuals and concludes that dimensional model is suitable for individuals high in valence focus and low in arousal, while discrete model is suitable for individuals high in arousal and low in valence focus [40]. Both discrete and dimensional perspective of emotions is widely used in affective computing. Discrete and dimensional perspective of emotions can be reconciled to certain extent by representing each discrete dimension as a combination of multiple dimensions[41][42]. Figure 2.1 summarizes some of the discrete and dimensional emotion models.

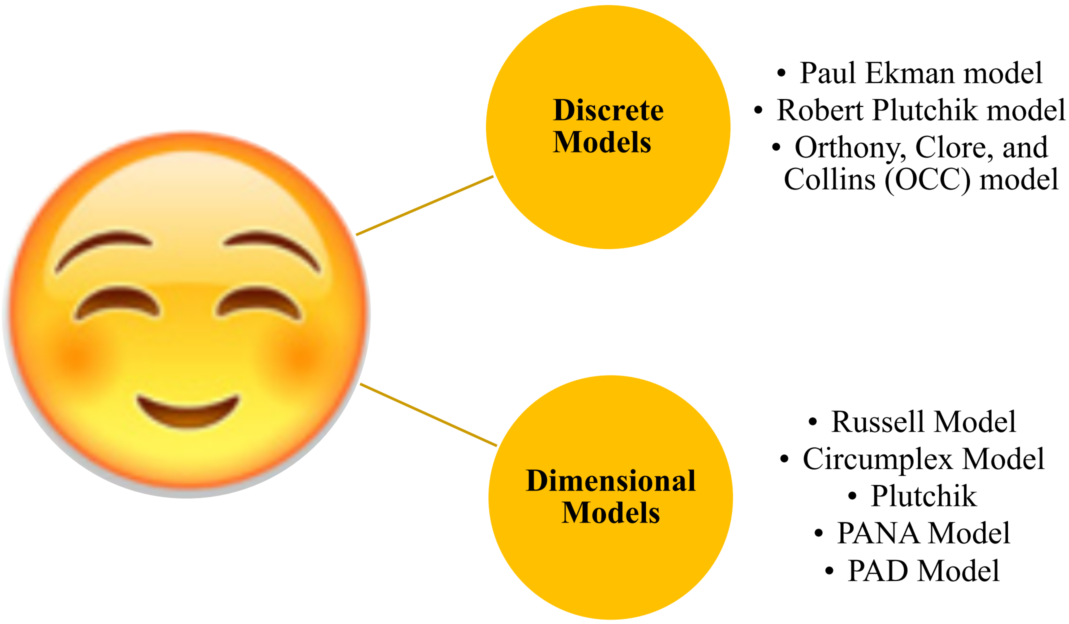



**Fig. 2.1**. Discrete and Dimensional Emotion Models

**Discrete Perspective of Emotion :** Discrete emotional perspective states that each emotion corresponds to unique experience, physiology and behaviour [43]. Paul Ekman [44] proposed six basic emotions: fear, anger, joy, sadness, disgust, and surprise while recent studies by Jack et.al., [45] have proposed that only four basic emotions namely happy, sad, fear/surprise, disgust/anger can be expressed in facial expression. Some of the discrete models are briefly discussed as follows

- ***Paul Ekman model:*** It distinguishes emotions based on six basic categories [42]. These fundamental emotions are happiness, sadness, anger, disgust, surprise, and fear. However, the synergy of these emotions could produce other complex emotions such as guilt, shame, pride, lust, greed, and so on.

- ***Robert Plutchik model:*** Putchik named eight of such fundamental emotions, that is, acceptance/trust and anticipation in addition to the six primary emotions posited by Ekman [49]. The eight emotions in opposite pairs are joy vs sadness, trust vs disgust, anger vs fear, and surprise vs anticipation.
- ***Orthony, Clore, and Collins (OCC) model*:** OCC dissented to the analogy of "basic emotions" as presented by Ekman and Plutchik [46]. They discretized emotions into 22, adding 16 emotions to the emotions Ekman posited as basic, thus spanning a much wider representation of emotions, with additional classes of relief, envy, reproach, self-reproach, appreciation, shame, pity, disappointment, admiration, hope, fears-confirmed, grief, gratification, gloating, like, and dislike.

**Dimensional Perspective of Emotions :** According to dimensional perspective of emotion, there are fundamental dimensions along which the emotional responses can be organized. There is no clear consensus on the dimensions that needs to be used, but all dimensional theories agree that emotions can be represented in few dimensions. Dimensional perspective of emotion suggest that emotions are same and vary only in pleasantness and intensity. Commonly used dimensions are valence, arousal and approach-avoidance.

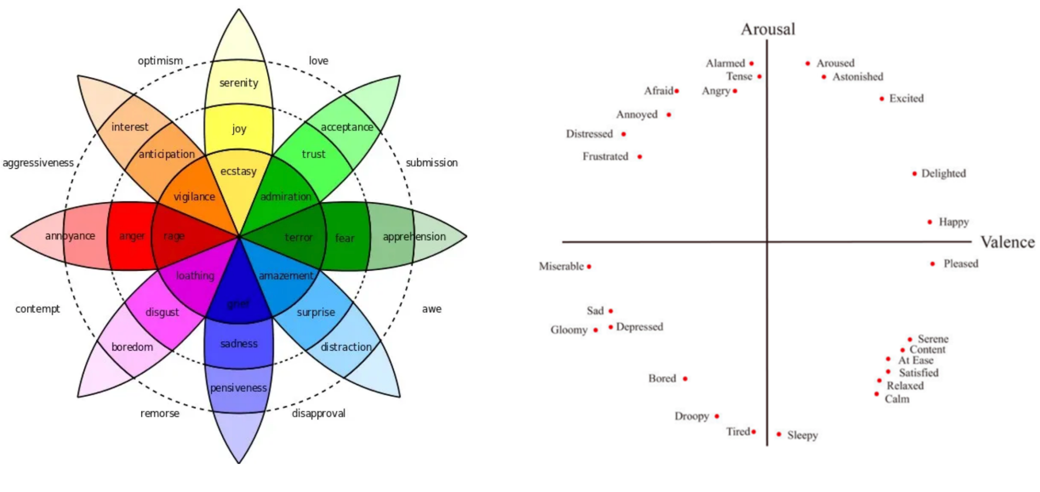



**Fig.2.2 .** Plutchik's Model and Russell' Model [48] [49]

Valence dimension scale contrasts quality of the emotion: pleasure and displeasure, while arousal scale contrasts quantity of the emotion: low arousal and high arousal [46]. One compelling reason to use dimensional model is that it has continuous scale and can represent any emotions of any intensity while discrete model is limited by the vocabulary. In this sub-section, a brief overview of few dominant dimensional models [47] namely – Russel's Model, Circumplex model, PANA model, Plutchik's model and PAD model are discussed. Structure of some of these models is shown in fig. 2.2.

- ***Russel's Model:*** Russell presents a circular two-dimensional model prominent in dimensional emotions representation called the circumplex of affect [48]. The model distinguishes emotions in the Arousal - Valence domains with Arousal differentiating emotions by Activations and Deactivations, whereas Valence differentiates emotions by Pleasantness and Unpleasantness. The inter-relationship can be represented in spatial model by a circle: pleasure (0°), excitement (45°), arousal (90°), distress (135°), dis- pleasure (180°), depression (225°), sleepiness (270°), and relaxation (315°). The Circumplex model of Affect establishes that emotions are not independent but related.
- ***PANA Model*** *:*Positive Affect and Negative Affect (PANA) model was developed by Watson and Tellegen [49]. PANA model states that Positive Affect and Negative Affect are the independent, uncorrelated dimensions of mood. The pleasantness-unpleasantness axis lies at 45°. Rotating this model by 45° makes it similar to circumplex model.
- ***Plutchik's Model*** *:* This model was developed by Plutchik [50]. It extends the circumplex model to third dimension – intensity of emotion. This model describes the relation among the emotions through colors of the color wheel. The vertical dimension of the cone represents the intensity while the circle represents the similarity among emotions. It has eight primary emotions: joy, trust, fear, surprise, sadness, disgust, anger and anticipation. Secondary and tertiary emotions are represented by combination of the primary emotions. As the distance from the centre of the circle increases, the intensity of the emotion decreases.
- ***PAD Model :*** Pleasure, Arousal and Dominance (PAD) model was developed by Russell and Mehrabian [51] and states that three dimensions: pleasure-displeasure, degree of arousal, and dominance-submissiveness are necessary and sufficient to adequately define emotional states. PAD model is widely used in non-verbal communication like body language, construction of animated agents

that express emotions and organizational studies to measure emotion towards specific product or entities.

**Table 2.2.** Affective Emotional State Detection Models

| Sr. No. | Author & Year | Dataset | Modality | Techniques Used | Result | Emotions |
|---|---|---|---|---|---|---|
| 1 | Fang et al. 2021 [52] | MAHNOB-HCI | EEG, Face | SVM | Accuracy: Arousal: 69.21%, Valance: 69.38% | Valance & Arousal |
| 2 | Zhang et al. 2021 [53] | DEAP | EEG, EMG, EDA, RESP | Regularized Deep Fusion of Kernel Machines | Accuracy: Arousal: 63.10%, Valance: 64.50% | Valance & Arousal |
| 3 | Panahi et al. 2021 [54] | ASCERTAIN | Face, Hand Gestures | SVM | Accuracy: Arousal: 75.24%, Valance: 76.81% | Valance & Arousal |
| 4 | Farhoudi et al. 2021 [55] | eNTERFACE'05 | Audio, Visual | CNN, 3DCNN | Accuracy: 81.7% | Angry, Disgust, Fear, Happy, Sad, Surprise |
| 5 | Nguyen et al. 2021 [56] | RECOLA | Audio, Visual | CNN, LSTM | RMSE: Arousal:0.474 Valance:0.187 | Valance & Arousal |
| 6 | Shi et al. 2021 [57] | IEMOCAP | Audio, Text | Spatio-Temporal Graph Convolution Network | Accuracy: Anger: 80%, Sad: 85%, Happy:80%, Neutral: 84% | Angry, Natural, Happy, Sad |
| 7 | Tzirakis et al. 2021 [58] | SEWA | Audio, Video, Text | DNN | Arousal:0.690 Valance: 0.783 | Low Arousal, High Arousal, Low Valance, High Valance |

| | | | | | | |
|---|---|---|---|---|---|---|
| 8 | Ayata et al. 2020 [59] | DEAP | RESP, PPG, FTT | Random Forest | Accuracy: Arousal: 73.08%, Valance: 72.18% | Valance & Arousal |
| 9 | Choi et al. 2020 [60] | MAHNOB-HCI | Video, EEG | LSTM | RMSE: 0.0366 | Valance & Arousal |
| 10 | Cimtay et al. 2020 [61] | DEAP | Face, EEG, GSR | CNN | Accuracy: 82.7% | Low Arousal, High Arousal, Low Valance, High Valance |
| 11 | Faridul et al. 2020 [62] | VIRI | Audio, Visible Image, Infrared Image | CNN | Accuracy: 86.36% | Happy, Sad, Anger, Surprise, Neutral |
| 12 | Ghaleb et al. 2020 [63] | CREMA-D, eNTERFACE'05 | Audio, Visual | CNN | Accuracy: 66.5%; Accuracy: 91.5% | Angry, Disgust, Fear, Happy, Surprise |
| 13 | Hunag et al. 2020 [64] | DEAP | GSR, RESP, ECG, EEG | CNN | Accuracy: 82.92% | Relax, Excitement, Fear, Depression |
| 14 | Tan et al. 2020 [65] | MAHNOB-HCI | Face, ECG, Temp, SCL, RESP, Mouth length, pupil size | Spiking Neural Network | Accuracy: 89.80% | Valance & Arousal |
| 15 | Wang et al. 2020 [66] | BioVid Emo, DEAP | ECG, SCL, tEMG | SVM | Accuracy: 89.80% (BioVid Emo) | Amusement, Anger, Disgust, Sad, Fear |
| 16 | Zhongmin et al. 2020 [67] | BioVid Emo | ECG, EMG | Multimodal Deep Belief Neural Network | Accuracy: 80.89% | Amusement, Anger, Disgust, Sad, Fear |

| | | | | | | |
|---|---|---|---|---|---|---|
| 17 | Gal et al. 2020 [68] | EMDeep | Respiration, Skin temp, heart rate, and galvanic skin response | Support vector Regression (SVR) | Accuracy: 74.0% | Angry, Disgust, Fear, Sadness, Surprise and Happiness |
| 18 | Chen et al. 2019 [69] | DECAF | ECG | Random Forest, SVM | Accuracy: Random Forest: 63.4, SVM: 64.5 | Arousal and Valance |
| 19 | Cai et al. 2019 [70] | IEMOCAP | Audio, Text | Multimodal CCN-Bi-LSTM-Attention-Bi-LSTM | Accuracy:71.25% | Angry, Excited, Neutral, Sad |
| 20 | Mehdi et al. 2019 [71] | DEAP | EDA, PPG, EMG | Deep Belief Network | Accuracy: 89.53% | Happy, Relaxed, Disgust, Sad, Neutral |
| 21 | Mostafa et al. 2018 [72] | BioVid Emo | ECG, EMG | SVM, KNN, Random Forest, RNN | Accuracy: 82% | Amusement, Anger, Disgust, Sad, Fear |
| 22 | Wiem et al. 2017 [73] | MANHOB-HCI | ECG, GSR, Resp, EEG | SVM | Arousal: 64.23(2 classes), 59.57%( 3 classes); Valence: 68.75 (2 classes), 57.44 (3 classes) | Valence, Arousal, Dominance |

From table 2.2, it was observed that most of the researchers have worked upon two main emotion models: Paul Ekman Model, and Russel Model. Multiple datasets are available for emotion recognition, as emotion recognition has many applications, like in human computer interaction. Some of these datasets are publicly available. Most commonly employed techniques for emotion recognition are SVM, and CNN. Although SVM only works well if the dataset is balanced, therefore CNN is the most employed general technique for emotion recognition from different modalities. In next section we discuss some of the available datasets briefly.

### 2.2.3. DATASETS

**Table 2.3.** Emotion Recognition Datasets

| Sr. No. | Dataset | Year | Modality | Emotion Model | Availability | Description |
|---|---|---|---|---|---|---|
| 1 | AMIGOS [74] | 2021 | Video, EEG, GSR, ECG | Discrete and Dimensional | On Request | Lab controlled study of 40 subjects |
| 2 | Multi-Person, Multimodal Board Games Affect and Interaction (MUMBAI) [75] | 2021 | Audio, Video | Discrete | On Request | Lab controlled study of 58 subjects |
| 3 | SEWA [76] | 2021 | Audio, Video | Dimensional | Public | 398 subjects data collected in open environment |
| 4 | K-EmoCon [77] | 2020 | Audio, Video, Biomarkers | Discrete and Dimensional | Public | 32 subjects data collected in controlled environment |
| 5 | MELD [78] | 2020 | Audio, Video, Text | Discrete | Public | Data recorded from TV series |
| 6 | IT Multimodal dataset for Emotion Recognition (IT-MDER) [79] | 2019 | ECG, EDA, BVP, RESP | Dimensional | On Request | 23 subjects data in controlled environment |
| 7 | ASCERTAIN [80] | 2018 | EEG,ECG,GSR, Facial Video | Dimensional | On Request | 58 subjects data in controlled environment |
| 8 | DREAMER [81] | 2018 | EEG, ECG | Dimensional | On Request | 23 subjects data in controlled environment |
| 9 | RAVDESS [82] | 2018 | Face, Speech | Discrete | Public | 24 subjects data recorded in Lab |
| 10 | BioVid Emo DB [83] | 2017 | Video, SCl, ECG, tEMG | Discrete | On Request | 94 subjects data recorded in Lab |

| | | | | | | |
|---|---|---|---|---|---|---|
| 11 | MSP-IMPROV [84] | 2017 | Audio, Video | Dimensional | On Request | 12 Subject data recorded in Lab |
| 12 | DECAF [85] | 2015 | MEG, ECG, tEMG | Dimensional | Public | 30 Subjects in controlled lab environment |
| 13 | CREMA-D [86] | 2014 | Audio, Video | Discrete | Public | 91 subjects data recorded in open environment |
| 14 | RECOLA [87] | 2013 | Audio, Video, ECG, EDA | Dimensional | Public | 46 subjects data in controlled environment |
| 15 | DEAP [88] | 2012 | Video, GSR, BVP, RESP, ST, EMG, EEG | Dimensional | On Request | 27 subjects data in controlled environment |
| 16 | MAHNOB-HCI [89] | 2012 | Audio, Video, Eye Gaze, EEG, ECG, GSR, RESP, ST | Discrete and Dimensional | Public | 220 subjects data in open environment |

Apart from above datasets, there are many more small datasets that have been developed for the emotion recognition. Table 2.3 contain datasets having multiple modalities. Some of datasets has been created for a specific modality to recognize emotions. For emotion recognition from Textual data only, there are many publicly available datasets such as ISEAR, EMOBANK, Daily Dialogue, Emotion Lines, SMILE dataset etc.

The latest dataset from Table 3, that incorporates most of the different modalities, and can detect emotion in both discrete and dimensional state is K-EmoCon. It's a multimodal publicly available dataset for emotion recognition in conversations provided by Park et al. in 2020 [26]. The dataset contains audio, video, and bio signals of 32 subjects, who participated in a debate task on a social issue in teams of 2, while wearing Physiological signal measuring devices, Emperica E4, NeuroSky, and Polar H7, along with Video cameras for recording the facial expressions and gestures of the participants. On average, 10 minutes of debate was conducted between each pair, for emotion recognition, audio and video was recorded of this debate, along with Bio-signals from 3 wearable devices. ACC (32Hz), BVP (64 Hz), EDA (4Hz), HR(1Hz), IBI, and Temp (4Hz) was measured through Emperica E4; Attention, Brainwave – EEG (125Hz) and Meditation through NeuroSky; and ECG (1 Hz) from Polar H7.

K-EmoCon is the first publicly available dataset that contains natural conversation between non-actors, while monitoring their physiological changes. Another unique feature in K-EmoCon is that each instance has been annotated by 7 people, one the person himself, second the opponent/partner in the debate, and lastly by 5 external observers. Although it makes processing the dataset difficult, but it provides different perspective for accurate emotion recognition. Park et al. have used 20 different types of emotions for annotation, including dimensional emotional model (Arousal and Valance), Basic Emotions (Cheerful, Happy, Angry, Nervous, and Sad), along with both common and less common BROMP (Baker Rodrigo Ocumpaugh Monitoring Protocol) Affective categories [77].

As expressed from table 2 and 3, the size of the datasets impacts the performance of the model. Due to difficulty in manual feature engineering, most of the researchers have proffered deep learning techniques such as CNN, LSTM over traditional machine learning models. And as evident from table 2, better accuracy has been achieved by the deep learning methodologies. These highlights observed from the existing literature has guided our path to choose right dataset and right methodology from the plethora of soft computing techniques.

## 2.3. Affective Biomarkers

From the pertinent literature, we have analyzed different biomarkers that is being used for affective state recognition. Majority of the studies use audio, visual signals, along with Bio signals. Although the biomarkers in psychology cannot be restricted to electro-physiological signals, as behavioural signals like posture, along with speech signals, social interactions and social media activities can also be used as observable traits that can indicate the status of mental health of an individual. Digital biomarkers can measure observable traits from diverse sources such as smartphone, social media, wearable devices to accurately predict the affective mental state. Some of changes in the body are not easily identifiable, they can be observed from their change in activities, like sleeping pattern, their change in writing style or cognitive thinking. On the basis of the measurement type of these changes we have broadly categorized the affective biomarkers into four type as shown in figure 2.3.

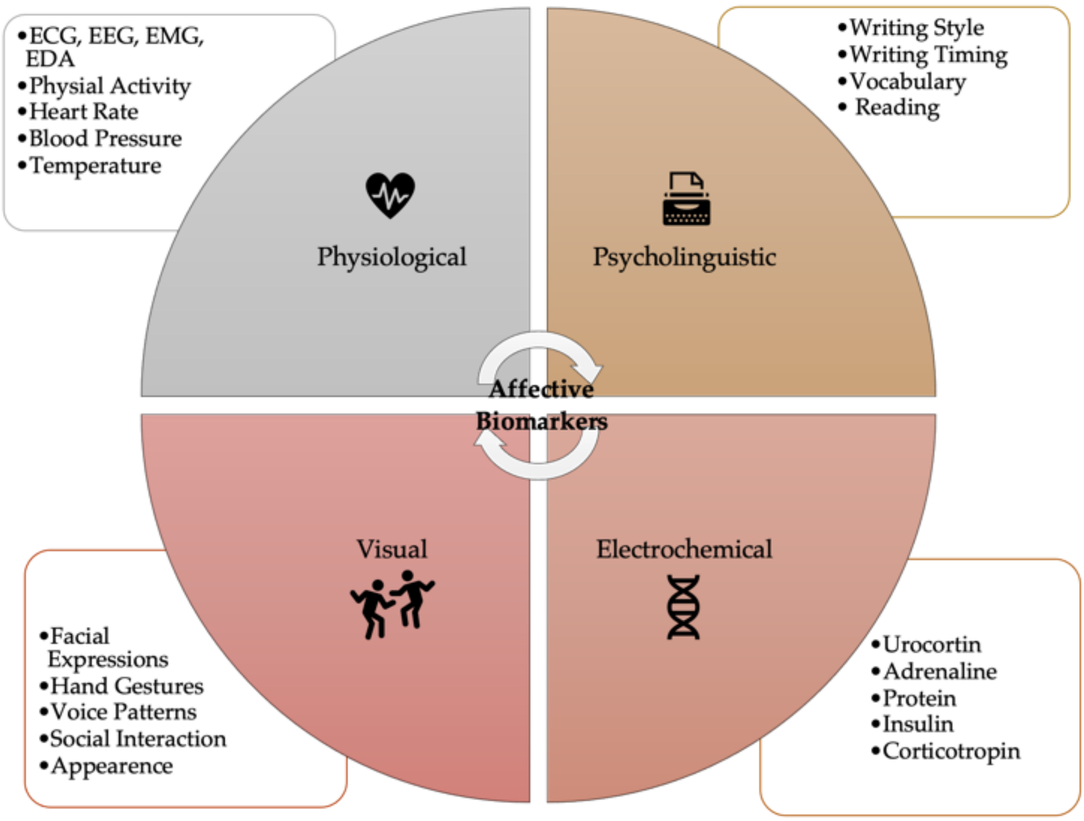



**Fig. 2.3.** Types of Affective Biomarkers

**Physiological Biomarkers:** Any measurable change happening inside a human body, also known physiological change is measured through Physiological Biomarkers. Such changes are observed through ECG, EEG, EMG, EDA etc. These biomarkers help in detecting both psychological as well as emotional affective state.

**Visual Biomarkers :** The visible reactions given by human body on experiencing a stimuli also helps in identifying the human affective state. The biomarkers used to measure these visible reactions fall under the category of visual biomarkers. Some of the visual markers are face expressions, hand gestures, sound pitch variations, appearance etc. These biomarkers are more prominently used in affective emotional state.

**Electrochemical Biomarkers:** The changes happening inside a human body that are not felt by the person fall into Electro-chemical biomarkers. These changes happen in biochemicals of the body which are regulated by the brain like Urocortin, Adrenaline, Insulin etc. Not much of the studies has been conducted on such biomarkers, but they are more prominently used in detecting affective psychological state.

**Psycholinguistic Biomarkers:** Any cue about the behavior of the person that can be articulated from their cognitive thinking through their linguistic activities fall under Psycholinguistic Biomarkers. These biomarkers include analysis of writing style, timing, vocabulary etc. These biomarkers are prominently used on a persons' social media activities to understand their affective psychological state.

## 2.4. Major Findings

It is evident from the related studies published within the domain in recent years that a variety of observable traits can be used for identifying the affective state of an individual and eventually indicate potential mental health and emotional health issues. Although the majority of work on Affective state recognition has been targeted for the healthcare, but their practical applications are prominent in other industries as well, like: Smart city, education, Enterprise Management applications, Automatic human computer interaction like chatbots and many more. The major findings from pertinent literature are:

- Affect recognition can be detected more accurately if different modalities can be used for this purpose.
- A person's environmental factors play an important role in managing affective state.
- The emergence of smartphone and wearable devices has begun to show promise for the assessment of real-time affective state.
- Several untapped sources of data, including social media data and multimodal data which combines psychophysiological activity, brain activity and facial expressions are open for research.
- There are multiple types of digital biomarkers which can be used for affective state mining.
- It is extremely difficult to distinguish which biomarkers are best suitable for identifying Affective psychological state, and affective emotional state.

- The affective event theory (AET) has not been studied generically to comprehend the cause-effect relationship between affective state, its causes and effects.
- Relation between emotion and mental (psychological) states has not been studied. As both are the part of affective state of humans, only one type of affective state may be identifiable using physiological signals.
- Majority of the physiological signals generated are under controlled lab environments, it is not certain that the individuals will act same in real-life scenario, with the availability of smart IoT based wearable sensors, the physiological signals should be collected in real-life situations.
- The evaluation criteria for measuring the accuracy of the proposed models are not standard. As while working on the affective state, it is important to have high recall as compared to the precision, so along with accuracy, the recall should also be verified.
- Majority of the researchers have used SVM, Random Forest, CNN, and LSTM models for classifying the affective state of the participants, although due to availability of multiple datasets, the transfer learning methodology can be employed to obtain more accurate results.
- No model has been designed to handle the real-time issues of affective state detection, like generation of volume of data in small intervals due to wearable sensors.
- Privacy of the subjects should be ensured. Decentralized data storage would be more efficient in data protection than centralized storage.
- Different fusion strategies are used to combine different modalities of affective state recognition.
- Each person responds to a stimuli differently, expressing different emotion, so generalized model could not result accurately for all. User specific models need to be developed.
- Datasets available are of very few subjects and are of limited variation. The cultural background of the subjects impacts their personality and affective state, to identify the impact of a particular stimuli on different people, a large dataset needs to be created, having people of different cultures, races etc.

- Context play an important role. The model should be developed that can incorporate the contextual information as well.
- How much each modality contributes to overall affective state recognition need to be explored.

## 2.5. Chapter Summary

This chapter presented an extensive literature review on the research work done in affective psychological state prediction and affective emotional state detection. We have identified the different emotional theories and models along with different biomarkers that can help understand the affective state of a person. The literature review facilitated identifying the research gaps within the domain.

***

# CHAPTER 3

# MULTI-MODAL AFFECTIVE STATE RECOGNITION

Automated human affective state detection has many promising real-world applications, and improvement in the same facilitates sound decision making. The human affective state comprises of a person's mental state and emotional state, although the both are used interchangeably sometimes, they are distinctly different. The objective of this thesis was to find different biomarkers that can be used to capture the change in human affective state, whether emotional or mental. To develop models that can handle different modalities of bio-markers to identify the affective state. It specifically aimed to develop novel techniques that can make it possible to have the real-world practical applications of human affective state detection.

## 3.1. Affective Psychological State

As every individual act differently on encountering the same situation, stress is not dependent on a single attribute, and it works differently for different individuals. Therefore, to evaluate the stress level of an individual without prior medical history is a difficult task [90]. Various biomarkers can be used to track the mental health of an individual like sleep pattern, level of cortisol and adrenaline hormones, walking pattern, outdoor activities, size of eye pupil, heartbeat rate while performing physical activities and while in the resting period as shown in figure 3.1.

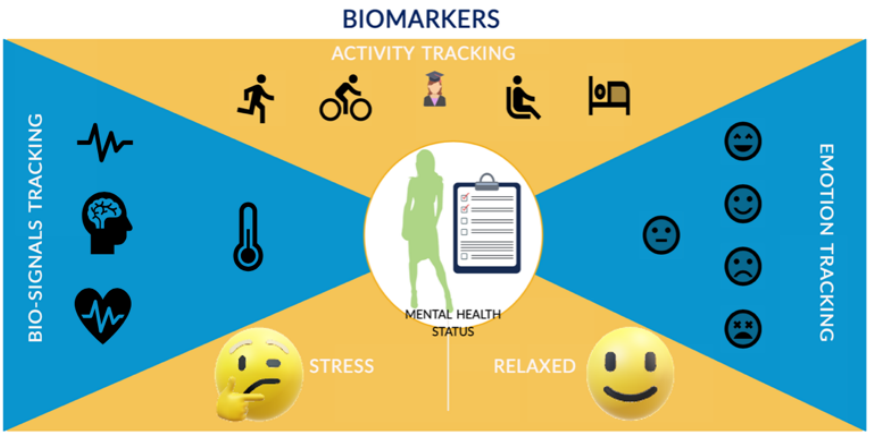



**Fig.3.1.** Biomarkers for Stress detection

Each psychological disorder has its own characteristic symptoms and some general warning signs to alert the need of professional help. An intelligent mental illness diagnostic can support clinicians with early detection. In this work, we have proposed a model to detect the metal stress at an early stage by evaluating the different biomarkers indicative of mental health.

### 3.1.1. Hierarchal Deep Neural Network Model- Physiological Biomarkers

The proposed hierarchal deep neural network that takes as input wearable stress and affects detection dataset (WESAD) that contains the bio-signals of 15 individuals collected from the wrist-wearable device (Empatica E4) and chest-worn device (RespiBAN) for a time-span of 2 hours. The different biomarkers that are taken into account to identify the stress in a user include ECG signals, TEMP, BVP, EDA, Resp, ACC, EMG.

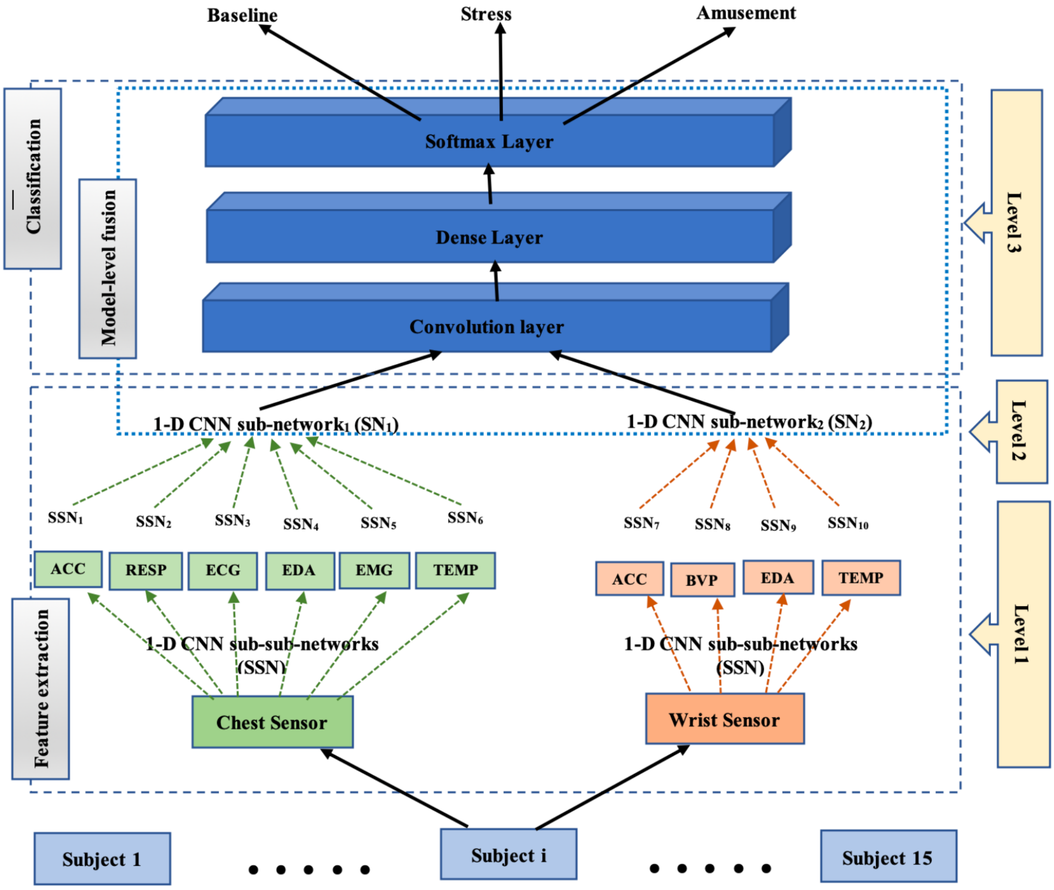



**Fig.3.2.** Architecture of Hierarchal Deep Neural Network

#### 3.1.1.1. Methodology

A hierarchal deep neural network was designed to detect psychological affective state, that takes as input wearable stress and affects detection dataset (WESAD) that contains the bio-signals of 15 individuals collected from the wrist-wearable device and chest-worn device for a timespan of 2 h. The model consists of three levels as shown in figure 3.2.

As WESAD contains bio-signals from 2 devices, so, to fetch the optimal values for every feature at each instance, at the first level, the sub-sub networks(SSN) for each bio-signal are used. 10 SSNs are a 1-dimensional convolution neural network (1D-CNN) containing two convolution layers with batch normalization and max-pooling and one dense layer. These SSNs generate a high-level representation of the respective wrist and chest-based biomarkers which are input to the respective sub-networks (SNs) at the second level. The SNs at the second level are 1D-CNN containing two convolution layers with batch normalization and max- pooling and one dense layer. This produces a combination of the high-level representation features of each device type biomarker. In next level, separately learned device type biomarkers are combined into one unified representation realizing a model-level fusion strategy. Thus, the shared representation is given to a convolution layer which generates the final feature vector and uses a denser layer to process the feature vector. The output layer generates a regression output with linear activation to finally detect the mental state. The network is compared with 5 machine learning based models (Decision tree and Random forest classifier, LDA, KNN, AdaBoost), and one deep learning model (CNN).

#### 3.1.1.2. Dataset

To identify the mental stress using behaviour biomarkers, we have used a sensor collected multimodal dataset which features psychological and motion data from both a wrist-worn device Empatica E4, and chest-worn device RespiBAN over 2 hours in a lab under controlled environment named WESAD. It is a publicly available data set for variable stress and affects detection collected by Schmidt et al. in a lab study on 15 subjects 12 being male and 3 been female. Different biomarkers that were used hey to monitor the stress level of a person are Blood Volume Pulse (BVP), Electrocardiogram (ECG), Electrodermal Activity (EDA), Electromyogram (EMG), Body Temperature (TEMP), Respiration (RESP) and Three-Axis Acceleration (ACC) motion. The data collected by RespiBAN was at 700Hz, whereas the data collected by the wrist device was at low resolution. All these biomarkers contribute to identifying the mental state of a person, whether he is amused, stressed, or is in normal state i.e., baseline. Each subject has 12 features, and the results were self-reported by the user. The dataset contains a total of 63000000 instances.

#### 3.1.1.3. Pre-processing

Since the different subjects have a different response to the different type of tests, they have variation in their signal values, to make the dataset ready for model, the pre-processing steps followed are:

- Min-Max normalization, so that all the subjects have test results in the same scale range.
- To evaluate a time series data, the continuous data is broken into short instances of one second, with sliding window of 0.25 seconds.
- The target class is converted to the numerical value, 1 for baseline, 2 for stress, and 3 for amusement.

#### 3.1.1.4. Findings

The proposed deep hierarchal neural network is trained with 65% of processed WESAD data once the continuous time series data is split over one second interval over the sliding window of 0.25 second. Model is tested over 35% of total data, i.e. 5 subjects are used for testing the models. Model is evaluated in terms of accuracy and f1-Score.

**Table 3.1.** Results of Proposed Model on WESAD data

| Subjects | Accuracy | Recall | Precision | F-1 Score |
|---|---|---|---|---|
| **S1** | 93.39 | 0.9973 | 0.901 | 0.9467 |
| **S2** | 96.98 | 0.9861 | 0.9474 | 0.998 |
| **S3** | 88.70 | 0.9193 | 0.826 | 0.612 |
| **S4** | 95.07 | 0.9693 | 0.9079 | 0.9292 |
| **S5** | 74.90 | 0.9051 | 0.5518 | 0.9685 |
| **S6** | 87.92 | 0.9593 | 0.7709 | 0.8548 |
| **S7** | 86.79 | 0.8877 | 0.8353 | 0.8601 |

| | | | | |
|---|---|---|---|---|
| **S8** | 72.19 | 0.5605 | 0.8761 | 0.6836 |
| **S9** | 87.24 | 0.8756 | 0.8145 | 0.798 |
| **S10** | 92.73 | 0.8831 | 0.9921 | 0.9348 |
| **S11** | 93.46 | 0.913 | 0.873 | 0.834 |
| **S12** | 87.24 | 0.8974 | 0.8222 | 0.8577 |
| **S13** | 88.72 | 0.961 | 0.92 | 0.93 |
| **S14** | 76.84 | 0.681 | 0.789 | 0.649 |
| **S15** | 93.56 | 0.925 | 0.95 | 0.917 |

The average Precision, Recall, Accuracy and F-1 score of each individual subject after 4-cross fold is shown in Table 3.1. Highest precision of 0.9921 has been attained for Subject 10, and Recall of 0.9973 has been obtained for Subject 1. Best average F-1 score of 0.998 and accuracy of 96.98% for Subject 2 was produced by Subject 2. The accuracy curve of the model is shown in figure 3.3 . It is observed that the accuracy of the model varies from 72% to 96%. The average accuracy achieved by the proposed model is better than the state-of-the-art results. The F-score of the model is 0.8325, and the average accuracy is 87.7% which is better than the state-of-the- art results provided by Schmidt et al. [91], Lin et al. [92] although F-score generated by CNN with late fusion is more than the pro- posed model. The comparison of the deep hierarchal model with results of other models is shown in Table 3.2.

**Table 3.2.** Accuracy of different classifiers over WESAD

| Classifier | Accuracy | F-Score |
|---|---|---|
| **Decision Tree** | 0.64 | 0.58 |
| **Random Forest** | 0.75 | 0.64 |
| **KNN** | 0.56 | 0.48 |
| **LDA** | 0.75 | 0.71 |
| **AdaBoost** | 0.79 | 0.69 |
| **CNN** | 0.85 | 0.86 |
| **Proposed Hierarchal Model** | 0.877 | 0.83 |

As validated from the results obtained, the accuracy of the model depends upon the subject's data, as different individuals personal characteristics of the subjects like medical history, traumas faced along with the continuous monitoring of the individual using IoT based wearable devices.

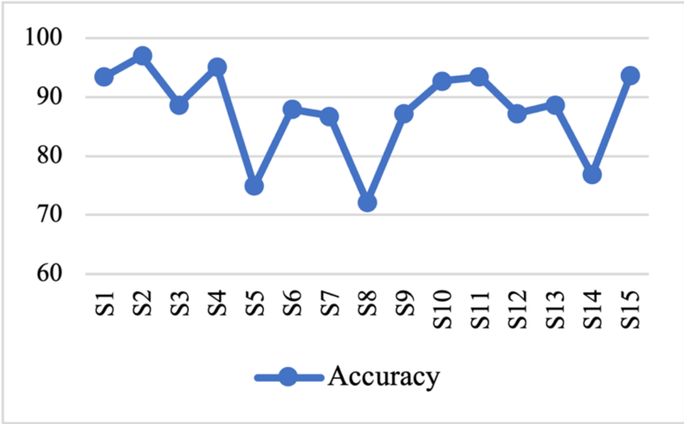



**Fig. 3.3.** Accuracy of Subjects

### 3.1.2. Anxious Depression- Psycholinguistic Biomarkers

Behavioral psychopathology relates anxiety and depression closely and anxious depression is defined as a mental state of individuals who are diagnosed with depression present in a manner that is more consistent with feeling anxious instead of sad. The pervasive social web can provide an apposite test-bed for understanding user behaviour and mental health. To predict the anxiety level of a person from their cognitive thinking process, we have taken into account their linguistic patterns through their social media posts. A predictive model was designed to detect this anxious depression disorder in online twitter posts of 100 sampled users. As the risk of anxiety and depression is higher in youngsters especially teenagers or students who are away from home, we considered the first 100 followers of MS India student forum.

#### 3.1.2.1. Methodology

The proposed prediction model is trained using hand-crafted features where the feature vector is a 5-tuple vector <w, t, f, s, c> . An anxiety lexicon base has been built as a part of this research which consists of 60 English words suggestive to anxiety. This seed list is grown using the WordNet. Tweets are analyzed using this lexicon base for the presence or absence of anxiety related word/words. Further, semantics of the tweet are determined using opinion polarity analytics by SentiWordNet. The users' posting patterns such as odd timing of posts and increased frequency of posts are also investigated to build the model for predicting anxious depression in real-time social data. An Ensemble Vote Classifier, a meta-classifier is used to combine the results of classifiers, namely the Multinomial Naïve Bayes, Gradient Boosting and Random Forest. The Ensemble Vote Classifier predicts the final class label via majority voting which is the class label that has been predicted most frequently by the classification models. The architecture of the proposed model is shown in figure 3.4.

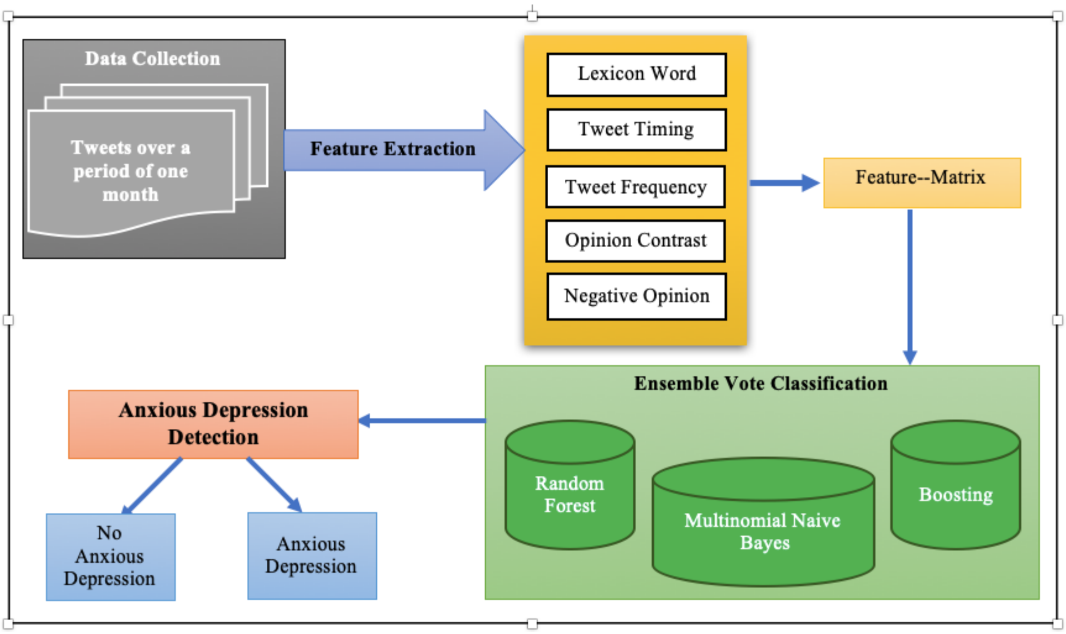



**Fig. 3.4.** Architecture of Anxious-Depression detection model

### 3.1.2.2. Dataset

The dataset with past one month tweets of 100 sampled users is scrapped using the Twitter API. The first 100 followers of MS India student forum are considered for this research. Each user's data consists of name, date of account creation, account verification status (verified or not), language, description, and tweet count. For each user, tweets are fetched, with date and time of post, number of re-tweets, hash tags, mentioned users.

The feature vector for building the learning model is trained using a 5-tuple vector <w, t, f, s, c>, where,

- <w:word>: Presence or absence of anxiety related word using the anxiety lexicon base
- <t:timing>: More than 2 posts during odd hours of night, specifically between 12am to 6am

- <f:frequency>: More than 3 posts in an hour during anytime of the day
- <s:sentiment>: More than 25% average posts in 30 days with negative polarity
- <c:contrast>: Presence of more than 25% polarity contrast in posts within the past 24 hours

The details of feature extraction and method of feature value assignment to generate feature vector are as follows:

*Anxiety Lexicon Base:* Depressive rumination is the compulsive focus of attention on thoughts that cause feelings of sadness, anxiety and distress. A person with anxious depression disorder is very likely to verbalize thoughts using specific anxiety related words. Therefore, an anxiety lexicon base with a seed list of 60 words is built with keywords that represent anxious depression in textual content. The seed list is eventually grown using WordNet. The processed tokens from the tweets are matched with the lexicon and the feature value is set to true ('1') if the word is present in the lexicon base else it is set to '0'. Table 3.3 presents the lexicon base of initial 60 words.

**Table 3.3.** Lexicon for Anxiety Detection

| **Anxious depression related words** |
|---|
| Fat, bad, weak, problem, tired, illusion, restless, bored, crap, shit, fuck, sad, escape, useless, meaningless, crying, reject, suffer, sleepless, never, bored, afraid, unhappy, ugly, upset, awful, torture, unsuccessful, helpless, suffer, fail, sorrow, nobody, blame, damaged, shatter. pathetic, insomnia, kill, panic, lonely, hate, depressed, frustrated, loser, suicidal, hurt, painful, disappoint, broke, abandon, worthless, regret, dissatisfied, lost, empty, destroy, ruin, die, sick. |

*Tweet Timing*: Chronic insomnia, which is sleeplessness, is one of the most common symptoms of anxious depression and stress. Users who are active through the midnight hours, i.e. from 12 am to 6 am, evidently show psychological disturbance in sleep pattern with increased restlessness and over-thinking. Therefore, tweet timing is an important feature and the value of feature is set to 1 if two or more than 2 tweets are posted after mid-night during odd hours of 12am to 6am, else it is set to 0.

*Tweet Frequency:* Though timing of tweets is primarily associated with odd-hour postings, generic tweet frequency within 24 hours also demonstrates user's restlessness and urge to share. The feature is set to true: '1' if the no. of tweets in any hour of the day is equal to or greater than 3, else it is set to false: '0'.

*Negative Polarity tweets:* Opinion polarity takes into account the amount of positive or negative terms that appear in a given text. In this research, to determine the polarity of tweets, SentiWordNet is used, it is a lexical resource which assigns the polarity to the words. The words of the WordNet are classified into the synset, and then each synset is assigned three values between the range of 0.0 and 1.0 representing the positive, negative and neutral polarity of the word. A single word can depict different sentiments in different scenarios, so as the word can have all three polarities of non-zero value. A cumulative value signifying the average of negative polarity tweets posted within the considered 30 days time frame is calculated. The feature is set to a value of '1' if 25% or more tweets posted have negative polarity; else it is set to '0'.

*Polarity Contrast:* The shift or contrast in polarity of posts from negative to positive or positive to negative is indicative of inconsistent mental state and restlessness. Typically described as a flip-flop behaviour, the person with anxious depression disorder often changes opinions and has confused thinking. Thus, to calculate this contrast *c*between opinion polarities of tweets, the following equation (1) is used:

$$c = \frac{(\delta \cdot PP + pw) - (\delta \cdot NP + nw)}{(\delta \cdot PP + pw) + (\delta \cdot NP + nw)} \quad (1)$$

where,

*pw* is the count of words with positive opinion polarity

*nw* is the count of words with negative opinion polarity

*PP* is the count of positive post

*NP* is the count of negative post

$\delta$ is the post co-efficient, the value of which is set to 3.

If a polarity contrast of >=0.25 magnitude is observed in tweets, then the feature value is set to '1'; else to '0'.

**3.1.2.3. Pre-processing**

Pre-processing is the process of cleaning and filtering the data to make it suitable for the feature extraction. The process includes:

- Removing numeric and empty texts, URLs, mentions, hashtags, non-ASCII characters, stop-words[1] and punctuations
- Tokenization of tweets is done using the TreebankWordTokenizer of Python Natural Language Toolkit (NLTK)[2]. The tokens are converted to lower case.
- Replacing slangs and emojis by their descriptive text using the *SMS Dictionary*[3]and *emojipedia*[4]respectively.
- Stemming to reduce the words to their root words using Porter's stemmer[6]. Stemming enhances the likelihood of matching to the lexicon.

### 3.1.2.4. Findings

Table 3.4 gives the accuracy of the individual classifiers and the ensemble vote classifier. The following Figure 3.5 illustrates the comparative performances of the classifiers graphically. The accuracy of the proposed AD prediction model is 85.09% with an F-score of 79.68%. The model is able to achieve motivating results and predicts users with anxious depression disorder. The following figure 3.66 depicts the performance results graphically.

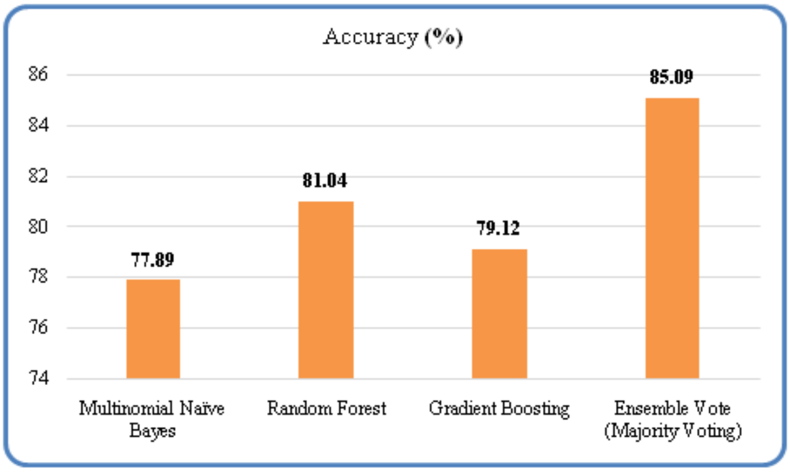



**Fig.3.5.** Accuracy of classifiers

[1] University of Glasgow Stop-word list
[2] https://www.nltk.org/
[3] SMS Dictionary. *Vodacom Messaging*. Retrieved 16 March 2012.
[4] https://emojipedia.org/
[6] Porter Stemmer: http://snowball.tartarus.org/

**Table 3.4.** Classification accuracy of classifiers

| CLASSIFIER | ACCURACY (%) |
|---|---|
| MULTINOMIAL NAÏVE BAYES | 77.89 |
| RANDOM FOREST | 81.04 |
| GRADIENT BOOSTING | 79.12 |
| ENSEMBLE VOTE CLASSIFIER | 85.09 |

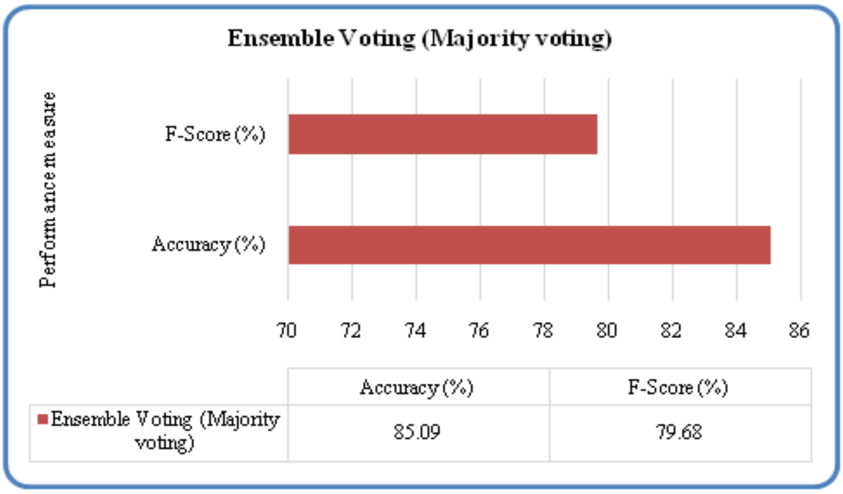



**Fig.3.6.** Performance of proposed AD prediction model

## 3.2. Affective Emotional State

Affective emotional state can be judged through various modalities - facial expression and body gestures (visual), speech (audio), spoken content (text), and physiological signals from sensors attached to the body. Most of the work conducted in the field is analyzing the human emotion state using unimodal data such as bio-signals, or only facial expressions. Deep learning architectures have achieved state-of-the-art results in computer vision with Convolutional Neural Networks (CNN), and in natural language processing (NLP) tasks with Transformer-based models Many datasets exist in the literature that contains different modalities, e.g. MELD contain audio, Video and text, whereas BioVid Emo contains a few biological biomarkers and Video signals. To understand the impact of each type of biomarker we have deployed the emotion recognition models on different datasets, having some variation in the models.

### 3.2.1. EmoHD: Emotional Health Detection- Audio, Video and Linguistic Modalities

EmoHD makes use of heterogeneous data from three modalities – visual (facial expressions), audio (tone) and text (linguistic), for classification of emotional state of the subject into one of seven categories - neutral, happiness, anger, sadness, fear, disgust and surprise. The EmoHD model is designed to emulate a real-world scenario. In real-time audio-visual input can be taken from video surveillance and separated into video and audio streams for individual processing. The video component is sampled to obtain image frames, from which the subject's face is extracted for facial expression recognition.

#### 3.2.1.1. Methodology

A fine-tuned ResNet50 pre-trained architecture is used for emotion classification. The audio signal is used for two components - speech emotion classification and automatic speech recognition (ASR). For classifying emotion based on audio, the audio clips are segmented into 960ms clips and converted to log mel spectrograms. A VGGish architecture pre-trained for audio classification is fine-tuned for this task. For automatic speech recognition, we use an out-of-the-box pre-trained model from nVIDIA's NeMo toolkit - QuartzNet which is trained for general human speech recognition. Transcription is compared with the available subtitles in the dataset using term frequency and cosine similarity.

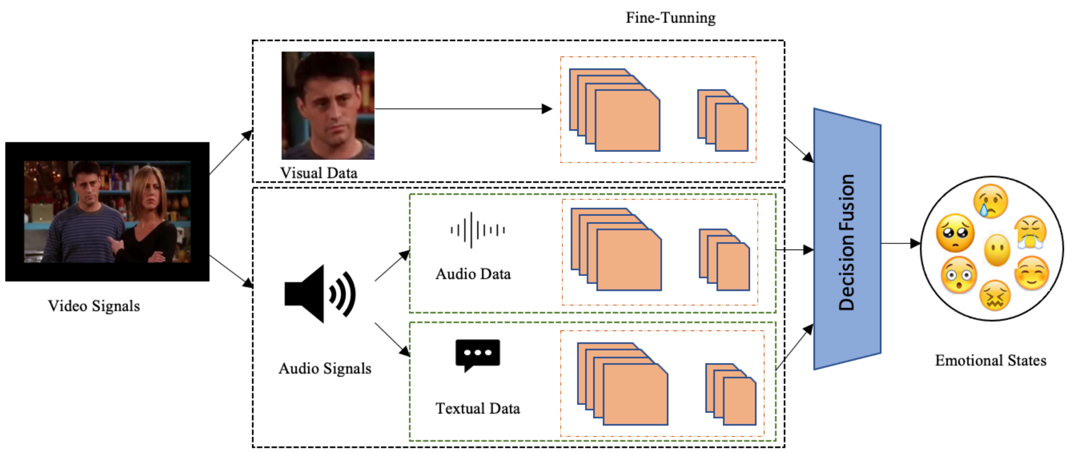



**Fig.3.7.** EmoHD Architecture

We find that the ASR is fairly accurate, hence the transcription is directly used instead of the subtitles for text emotion classification, in accordance with our objective of emulating a real-world scenario. For text classification, we fine-tune DistilBERT, a general- purpose language representation model which is condensed from the original BERT model resulting in reduced size and faster inference time while retaining almost all of the capabilities. For combining results from the three modalities, we choose a late fusion strategy which is justified both empirically and logically. The architecture of the model is shown in figure 3.7.

#### 3.2.1.2. Dataset

*MELD :* To fine-tune and evaluate our model, we use a portion of the MELD dataset, which is an extension of the EmotionLines dataset [78]. The latter only has text modality while the former includes video and audio to accompany the text. MELD contains 1433 dialogues taken from the 'Friends' TV series, where each dialogue encompasses one or more participants (actors) and emotions. An utterance is a part of the dialogue where only one participant is speaking and has one corresponding emotion. There are a total of 13708 utterances in MELD, with an average duration of 3.59 seconds. The utterances are categorized under our desired emotion states. For our experiments, we select a subset of MELD (henceforth referred to as 'MELD-sub') by leaving out the utterances where multiple participants are speaking, or multiple participants are facing the camera while one of them is speaking. The selected utterances have either a single participant, or only the speaker having their face completely visible on the camera.

*IEMOCAP :* Interactive emotional dyadic motion capture database was collected by SAIL laboratory at University of South California. Busso et. al. [93], they hired 10 actors to participate in a study. Total of 5 sessions were conducted, each having 2 actors having markers on their face, head, and hands to record their facial expressions as well as their hand movements. The study was conducted for a period of 12 hours. The sessions contain both scripted and spontaneous communication between the two actors. The 12-hour recording has been manually segmented into 10039 utterances, belonging to 9 different emotions- happiness, anger, surprise, sadness, fear, excitement, frustration, neutral, and others. For our study we have merged the frustration into anger and utterances of others has been dropped, converting the data with 7 emotions only. Table 3.5 highlights the emotion-wise distribution of the datasets.

**Table 3.5**. Emotion-wise data distribution

| EMOTION | CIFE | MELD | IEMOCAP |
|---|---|---|---|
| **ANGER** | 1785 | 1607 | 1103 |
| **DISGUST** | 266 | 361 | 471 |
| **FEAR** | 761 | 358 | 589 |
| **HAPPINESS** | 3636 | 2308 | 648 |
| **NEUTRAL** | 644 | 6436 | 1708 |
| **SADNESS** | 2485 | 1002 | 1084 |
| **SURPRISE** | 997 | 1636 | 988 |

*CIFE :*Candid images for facial expression (CIFE) is a dataset created by Li et al. [94] to construct an improved facial expression model for analyzing real time facial expression tasks. The CIFE dataset is produced through social media and the Web. Web crawling methods are employed to obtain natural expressions in the seven chosen categories of emotions. These categories are happy, anger, disgust, sad, surprise, fear and neutral. Utilizing related phrases of these expressions, a huge amount of pictures are accumulated corresponding to the seven classes of expressions. There were 14756 pictures in total for these seven expressions in which pictures of anger, disgust, fear, happiness, neutral, sadness and surprise, and some pictures were added manually to the dataset corresponding to the classes where data was unbalanced. Viola face detector was availed to uncover images of faces with these seven expressions [95]. CIFE dataset is a freely available public dataset.

**3.2.1.3. Visual Modality** : A video can be considered to be a spatio-temporally connected stack of images. Since our goal is real-time evaluation and our data also consists of short utterances, we choose to focus on facial expressions in sampled images from the video clips. We do not consider temporal features which would be effective in evaluating longer videos.

*Pre-processing* :The video clips from MELD-sub and IEMOCAP are sampled at 1 frame per second. Since the utterances are short (3-4 seconds on average), we avoid picking a peak frame for analysis. Instead, we adopt a majority voting strategy after individual emotion classification of the frames tagged to a video clip. As the sampled frame depicts a scene, but for emotion analysis, only facial expressions are required, multi-task cascaded convolution network (MTCNN) is used to extract faces. Frames with no faces detected are discarded. The extracted face images are then used to fine-tune the pre-trained model. The sampled frames from MELD-sub before and after pre-processing are shown in figure 3.8.

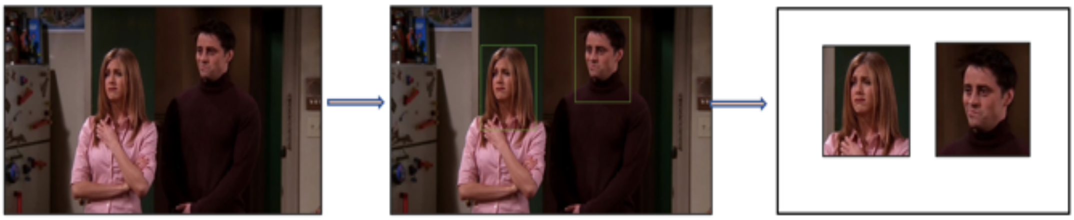


**Fig.3.8.** Face extraction using MTCNN

*Fine-tuning :* We start with a ResNet50 model loaded with weights pre-trained on the VGGFace2 dataset as our base model. VGGFace2 consists of face images with significant variations in pose, illumination, ethnicity and age of the subject. This acts as a suitable precursor to the data we will use downstream for emotion classification. ResNet50 is a convolution-based architecture which achieved state-of-the-art results on ImageNet and several other image classification tasks. ResNet50 model has been implemented in Keras for emotion detection from face recognition by pre-training the model on VGGFace2 dataset for face identification and then using CIFE to map the faces with different emotions, and at last this pre-trained model detects the affective emotion state of an individual from MELD-sub and IEMOCAP's visual signals.

CNN-architectures learn general features in the lower layers and more task-specific features in the higher layers. We utilized this property to fine-tune the base model by unfreezing some layers from the top and retraining the network on the CIFE dataset and converge the architecture towards our target task of emotion classification. The CIFE dataset also contains images with varying illumination, age and ethnicity, making it suitable to act as a bridge between VGGFace2 and both the datasets. The fully connected output layer of the base model, softmax classifier for 8631 categories is removed. The top 10 layers (average pooling layer and 9 convolutional layers) of the model are unfrozen and a fully connected layer with 5 nodes is added at the top for our target emotion categories for the second round of pre-training on CIFE, to minimize categorical cross-entropy loss using the Adam optimizer.

*Feature extraction and classification :* After fine-tuning on CIFE, we freeze all the layers of the model and remove the top layer. The MELD-sub and IEMOCAP images are passed through the fine-tuned model to obtain vectors representing the extracted features. We use these feature vectors as input to the multi layer perceptron (MLP) network for emotion classification. The configuration of the MLP network is shown in figure 3.9.

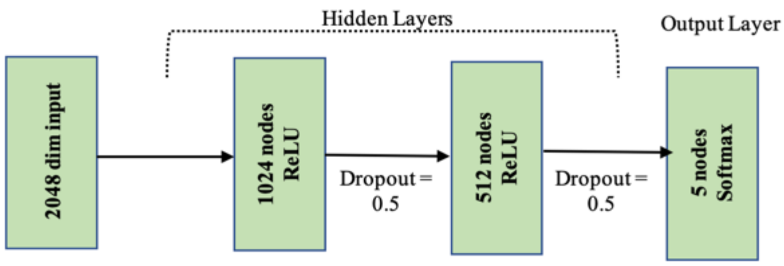



**Fig.3.9.** MLP Network

### 3.2.1.4. Audio Modality

To avoid training to fit a dataset on a particular domain, we adopt a transfer learning approach. We extract the audio from video clips using the *MoviePy* library in *wav* format and *librosa* library is used for audio pre-processing.

*Pre-processing :* The audio clips are divided into 1 second frames to match the image sampling. They are resampled to 16 kHz and converted to a spectrogram using short-time Fourier transform. The window size and window hop are set to 25ms and 10ms respectively. A mel spectrogram is then computed by mapping the spectrogram into 64 bins within a range of 125 to 7500 Hz. Finally, the log mel spectrogram is computed with an offset of 0.01.

*Fine-tuning and classification :* We utilize the VGGish model for the audio component. VGGish is pre-trained on a large YouTube dataset and provides a 128-dimension embedding. Since the model is trained for general audio categorization, we fine-tune it for the emotion classification task. The pre-processing is done in accordance with the audio features required as inputs to VGGish. The pre-trained weights are frozen, except for the fully connected layers, which are trainable. This Pre-trained model is then fine-tunned with MELD-sub and IEMOCAP datasets by taking 65% of both the datasets with seven classes. The final layer is modified to have 7 nodes for each emotion category. Additionally, the activation function for the final layer is changed from sigmoid to softmax since each audio sample has only one associated emotion label. The log mel-spectrogram features are passed in as input. Since the weights of convolutional layers are frozen, the top 3 fully-connected layers converge the model towards emotion classification.

#### 3.2.1.5. Text Modality

We analyze the spoken content by converting extracted audio clips to text using ASR. This analysis is important because there may be cases where the person is not expressive enough or speaks in monotone. While these are important factors in determining the emotional state of the subject, they would not be reliable in such cases. Another factor to consider is that facial features of a person might be similar for different emotions, for instance, furrowed eyebrows to show anger as well as surprise. In such a case audio features and spoken content will be the determining factors.

*Audio transcription :* NVIDIA's NeMo toolkit is used for transcription. The model architecture used for ASR is QuartzNet which is comparatively smaller than other existing models. In particular, the model we use is QuartzNet15x5NR-En which has been pre-trained on the LibriSpeech corpus and fine-tuned with Room Impulse Responses (RIR) and noise augmentation to make it robust to noise. The pre-trained model is used out-of-the-box since it has state-of-the-art performance on general speech. The audio transcription is done for both the dataset. We extract the output in two forms - lowercase alphabetic tokens for similarity check, and complete sentences formed by applying the DistilBERT-punctuation scheme for our input to the masked language model.

*Similarity check :* For Real-time applications, where the video-surveillance will be provided as input, and the transcripts will not be available, the ASR will provide the adequate help in detecting the emotions of an individual accurately. To gauge the performance of our ASR module, we are interested only in word-level similarity and not the context or meaning. Ideally, we want the transcription to match the subtitles word to word. For MELD-sub we perform the similarity check with the EmotionLines dataset [78], and for IEMOCAP, its textual component which is provided along with the dataset. We also used Term Frequency (TF) and Cosine Similarity (CS) for checking their alignment. The ASR module has shown the average accuracy of 93.07% for both the datasets.

*Fine-tuning and classification :* Transformer-based architectures and BERT-inspired training schemes dominate the state-of-the-art in NLP tasks. With each jump in accuracy, the models keep getting larger with the number (in millions) of parameters being - Google's BERT-base (110) and BERT-large (340), Facebook's RoBERTa (355), OpenAI's GPT-2 (1500) and nVIDIA's MegatronLM (8300) among others. However, with increasing size the models keep becoming less suitable for small devices such as smartphones, which are the backbone of smart services. We utilize DistilBERT, a model condensed from BERT-base using a student - teacher training method. It is reported to have 60% faster inference time at 97% of the performance of BERT-base while having only 66 million parameters, which is a 40% reduction in size.

We use HuggingFace's transformers library to fine-tune a pre-trained DistilBERT model, specifically Distilbert-base-uncased for complete sentence generation. These transcriptions generated from the ASR module are tokenized and padded to align with the maximum-length utterance. The fine-tunned model on Distilbert-base-uncased is then again fine-tunned, this time for mapping emotions to the sentences, it is fine-tunned on 65% of MELD-sub and IEMOCAP datasets separately. A 'CLS' token (Special classification token) is added at the beginning for the classification task. The output of the final transformer layer for the CLS token will be used as the input features that will be fed into an MLP classifier.

### 3.2.1.6. Weighted Decision Fusion

In general, multimodal data fusion can be done in three ways, early fusion, late fusion and joint fusion [96]. To enable real-time evaluation, we select a decision-level fusion after evaluating the three components parallel ensuring that they are unaffected by the quality of data in another modality. This makes it robust to data quality in either modality, in situations where the subject is visible in the camera, but the audio is unclear or vice-versa. The content of the subject's utterance is also evaluated to judge their emotional state. To process the results of the 3 parallel models (Visual, Audio, and Text Models) the weighted decision fusion technique has been used. Every model has been fitted for varied modality, each having results dependent on subjects not just the cause. To identify the impact of each modality the model has been trained separately for the three signals. A thorough review of impact of different fusion studies has been shown by [52]. In the proposed model, we have used the weighted late fusion framework provided by Tsanousa et al. [53]. The framework assigns weights on the basis of detection ratio rather than F-scores. Detection Ratio (DR) is shown in equation 6.

$$DR= TP/(TP+TN+FP+FN) \tag{6}$$

where, TP represent true positive, TN represents True Negative, FP represents False Positive, and FN is False Negative. DT is calculated for each class, as the model has been executed for 7 Discrete Emotions, the number of classes are 7. The weight, W of each output class is calculated as:

$$W= 1-DR \tag{7}$$

The weight of each class is then multiplied with the probability vector, P belonging to each model to find the predictive score of the class.

$$S=W*P \tag{8}$$

After calculating the weight of each class for individual model, the score of the model is calculated by adding the scores of each class. The final decision is opted trough the maximum function, i.e., the model having highest predictive score for the test case is chosen as the output level. As the proposed architecture contains 3 models for 3 modalities, the final output class is provided as:

$$Output\ Class= Max\ (S_{Visual}, S_{Audio}, S_{Text}) \tag{9}$$

This late weighted fusion strategy helps to choose the output class from the model best suitable for the output class.

#### 3.2.1.7. Findings

Initially the model is tested for each modality. In the first model (Visual), after fetching the facial expressions using MTCNN, the facial expressions of each frame are provided as an input to fine-tuned transferred network using ResNet50. The model provided an F1-score of 54.93 for IEMOCAP and 51.96 for MELD dataset. Both the datasets were also evaluated for the audio model (VGGish) and the textual model (BERT) and report an F1-score of 53.39 and 54.19 for IEMOCAP and 50.85 and 51.48 for the MELD dataset respectively. As video signals are further broken down into visual and audio signals, so same input is required for both, therefore the model is evaluated for these modalities, providing a better F1-score for both the datasets as compared to when tested on a single modality. Similarly, as textual data is fetched form audio signals using ASR, the two modalities are also combined to understand the impact of two modalities rather than one. But since textual data cannot be directly generated from the video signals, the model is directly evaluated on the three modalities to detect the final class using weighted late fusion. The performance of each implementation is shown in Table 3.6 and 3.7 for the IEMOCAP and the MELD datasets respectively. The performance of the models is shown for each emotion as well. As observed from the table 3.6 and 3.7 the model performs better for the 'happiness' emotion followed by 'sadness' emotion. This variation in accuracy of detection of each emotion has happened because of variations in the training datasets as well as certain emotions are expressed differently by different individuals. To have an effective real-time emotion detection, we need to fine tune the models on even larger datasets.

**Table 3.6.** F1-Score for varied emotions for IEMOCAP

| Modalities / Emotions | Visual | Audio | Text | Visual +Audio | Audio + Text | Visual + Audio + Text |
|---|---|---|---|---|---|---|
| **Anger** | 55.18 | 52.1 | 54.62 | 58.91 | 58.14 | 65.07 |
| **Disgust** | 52.61 | 51.66 | 47.3 | 54.03 | 54.1 | 58.41 |
| **Fear** | 57.06 | 53.08 | 51.62 | 59.73 | 56.32 | 64.81 |
| **Happiness** | 58.18 | 55.34 | 61.48 | 64.05 | 61.28 | 71.23 |
| **Neutral** | 54.29 | 56.65 | 52.53 | 59.24 | 52.1 | 65.51 |
| **Sadness** | 52.4 | 49.58 | 60.02 | 56.11 | 62.7 | 69.78 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Surprise** | 54.83 | 55.34 | 51.8 | 60.72 | 53.03 | 66.41 |
| **Total** | **54.93** | **53.39** | **54.19** | **58.97** | **56.81** | **65.88** |

**Table 3.7.** F1- Score for MELD dataset

| **Modalities / Emotions** | **Visual** | **Audio** | **Text** | **Visual +Audio** | **Audio + Text** | **Visual + Audio + Text** |
|---|---|---|---|---|---|---|
| **Anger** | 53.12 | 49.23 | 52.41 | 56.18 | 53.25 | 59.95 |
| **Disgust** | 49.14 | 47.62 | 43.11 | 48.91 | 49.21 | 54.82 |
| **Fear** | 57.81 | 50.21 | 48.93 | 55.69 | 54.31 | 59.31 |
| **Happiness** | 52.91 | 52.47 | 57.29 | 57.83 | 56.21 | 65.36 |
| **Neutral** | 51.23 | 53.78 | 49.71 | 55.78 | 49.78 | 61.46 |
| **Sadness** | 48.19 | 46.89 | 57.21 | 52.34 | 54.78 | 66.71 |
| **Surprise** | 51.36 | 55.78 | 51.72 | 56.53 | 56.45 | 61.29 |
| **Total** | **51.96** | **50.85** | **51.48** | **54.75** | **53.42** | **61.27** |

The proposed EmoHD model which combines three modalities, namely visual, audio and textual, surpasses the results with late weighted fusion in comparison to separate modalities or subset of modalities. But at the same time, it can also be observed from the figure 3.10 and 3.11, that although the model and execution environment is same, but the performance varies widely for two datasets.

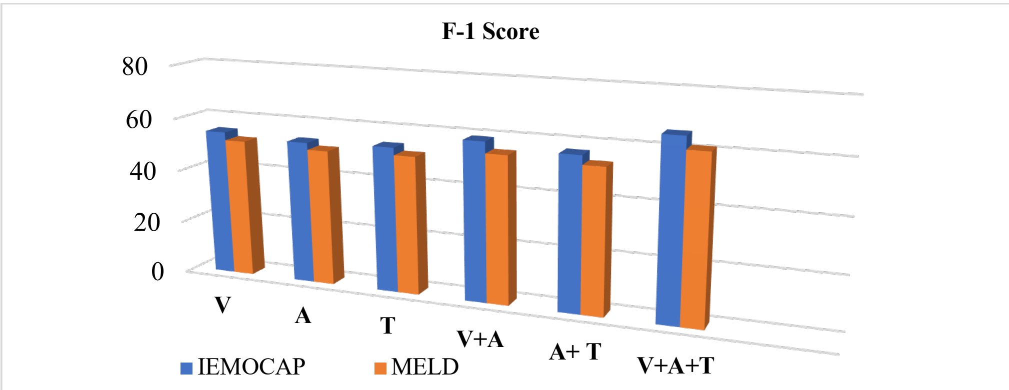



**Fig.3.10.** Performance of various models on MELD and IEMOCAP

The variation in the performance is impacted by a lot of factors, the most promptly, each individual expresses the emotions differently, the actors for both the datasets were different.

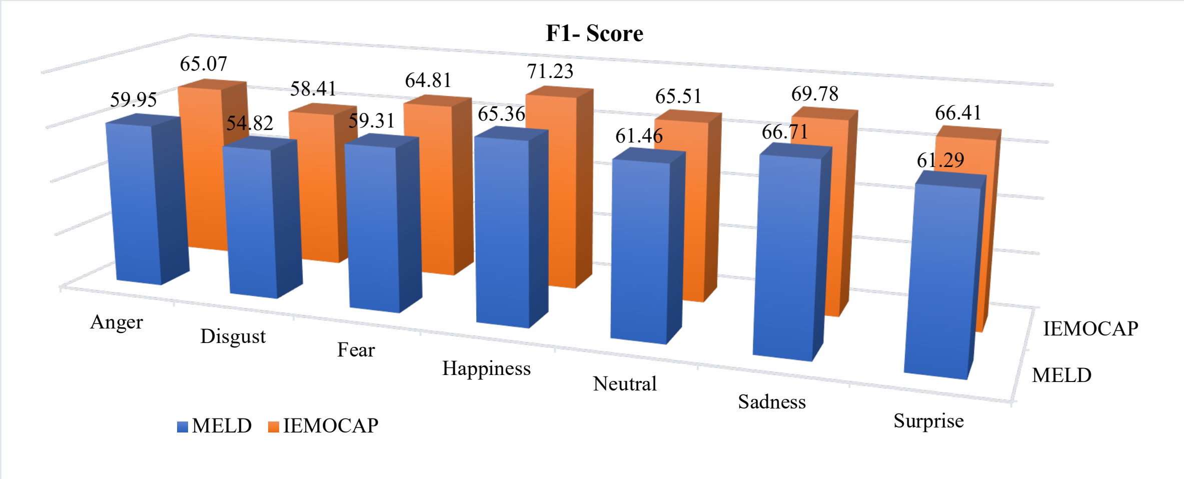



**Fig. 3.11.** Proposed Model for Discrete Emotions

The proposed deep transfer learning weighted late-fusion model has identified the discrete emotion more precisely in comparison to the existing models. The performance of the proposed EmoHD model is compared with existing state-of-the-art [97,98]. Table 3.8, shows the comparison of EmoHD model for both the datasets, with the results provided by Acheampong et al. for MELD [98] and Hazarika et al. for IEMOCAP [97]. As observable from Table 3.8, authors in [97] and [98] have performed empirical analysis on MELD and IEMOCAP datasets respectively. The deep convolution model shows better results on MELD, whereas modified recurrent neural network performs better on IEMOCAP. The proposed model has performed subpar than SOTA, as it takes individual modalities into account, and have used the existing data to pre-train the models which eliminates the impact of small dataset on the accuracy. Although the variation in results on the two datasets over same model indicates the accuracy of the model will depend upon the data (subjects). For accurate detection, a pre-trained model should be fine-tunned per user specific for accurate estimation of emotional state of an individual.

**Table 3.8.** Comparison of Models

| MODELS | IEMOCAP | MELD |
|---|---|---|
| **CNN** | 48.1 | 55.02 |
| **C-LSTM** | 54.9 | 56.44 |
| **C-LSTM + ATT** | 56.1 | - |
| **DIALOGUERNN** | 59.8 | 57.03 |
| **TL-ERC** | 58.85 | - |

| | | |
|---|---|---|
| CONGCN | - | 59.40 |
| **PROPOSED MODEL** | **65.88** | **61.27** |

As indicated from results, the weighted late fusion model can act as a promising model for merging the three modalities of visual, audio and text for accessing the affective emotional state of an individual in real-time using video surveillance.

### 3.2.2. MEmoR: Multi-modal Emotion Recognition- Video and Physiological Modalities

MEmoR model considers two modalities, namely, visual and physiological for the classification of emotional state of the subject (employee in case of industry) into one of five discrete categories - anger, sadness, fear, disgust, and amusement.

#### 3.2.2.1. Methodology

The proposed model is quite similar to EmoHD, the major difference comes in the modalities used. This model is studied on the combination on physiological biomarkers and visual biomarkers. The architecture of the proposed model is shown in figure 3.12, containing four model, one for visual modality and 3 for physiological modalities. The visual input is taken and sampled to obtain image frames, from which the subject's face is extracted for facial expression recognition.

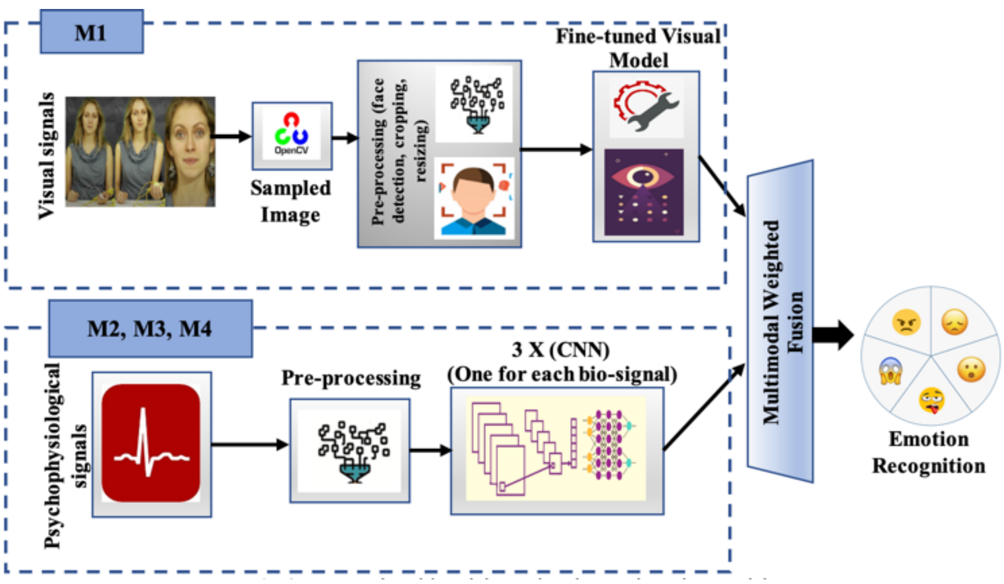



**Fig.3.12.** MEmoR Architecture

A ResNet50 architecture pre-trained for face recognition is fine-tuned for emotion classification. The physiological signal data is used to train a convolution neural network (CNN) and detect the emotional state of an individual. As the data modalities contain information that are correlated to each other at a higher level and have varied data dimensionality, for combining the results of the two modalities, we have used a late data fusion technique.

#### 3.2.2.2. Datasets

The BioVid Emo multimodal database [83] is used to evaluate and validate the proposed MEmoR model. It is a multimodal high-quality physiological database containing discrete fundamental feelings and is created to examine human feelings and to mine emotional affective states. Bio signals like SCL, ECG, EMG are recorded in a controlled environment to identify different emotional states like amusement, sadness, anger, disgust and fear. It contains the data of individuals that were shown various film clips and their evoked feelings and emotional states were recorded. For each individual the most dominant emotion is observed and provided for the dataset. In this dataset, 94 individuals have participated across three diverse age ranges of 18 to 35 years, 36 to 50 years, and 51 to 65 years of age. Of all the participants, no one had any affective or emotional problems. There were 50 female and 44 males participants. But only 86 participants' data is complete and thus available in the dataset and the rest of the data entries are discarded due to incomplete or defiled entries. 15 film clips were selected, and the affective states were documented by instigating emotions through these clips. For instance, to trigger the feeling of sadness, clips like "The Champ" were shown to the participants where "A boxer is seriously injured and dying while his son enters" [83]. BioVid Emo database is now used by researchers for further emotional or affective state mining.

#### 3.2.2.3. Physiological Modality

Psychophysiological signals measure any change a human body experiences, irrespective of any visual symptoms of the change. Digital affective biomarkers can help to track these physiological changes in real-time using wearable devices. The Department of Medical Psychology, University of Ulm, Germany has recorded three different such Psychophysiological signals using 6 Ag/AgCl electrodes fixed on the index and ring fingers, 2 on the upper right and lower left of the body, along with 2 on descending portion of the upper trapezius muscles for recording Skin Conductance Level (SCL), Electrocardiogram (ECG) and Trapezius Electromyogram (tEMG) respectively. Each bio-signal recorded acts as a good indicator of different affective emotion states of the individual. While experiencing sad emotions, the heart rate tends to decline, whereas while in an angry state, the skin temperature rises. To effectively identify the human affective emotional state, three separate deep learning models $M_2$, $M_3$ and $M_4$ were developed to identify the emotion state individually for each signal.

- CNN model for SCL $M_2$: The SCL signals measure the electrodermal activity of the sympathetic nervous system, it can measure the fluctuations in skin conductance shown due to eccrine sweat and gland activity.
- CNN model for ECG $M_3$: ECG signals can further help to derive the heart rate, interbeat interval and heart rate variability etc. These signals can help to identify fear, disgust, and amusement in individuals.
- CNN model for tEMG $M_4$: tEMG measures the muscle activities, an increase in muscle tone relates to increased activity of the sympathetic nervous system indicating anger, amusement, arousal, etc.

#### 3.2.2.4. Findings

The models were executed separately on the two types of output categorized dataset. Initially, the model was tested for two output classes Valence and Arousal on 7 different scenarios. First only CNN based $M_1$ model was used on SCL signals only resulting in 65.54 accuracy, second deep learning model was executed only on ECG signal resulting in 72.89 accuracy. Third the model was trained on only tEMG signals producing 69.81 accuracy, afterwards the three signals were fused at feature level, i.e., early fusion was applied on the three bio-signals and then the CNN model was used resulting in better accuracy of 77.92%.

The face detection was also tested individually for affective emotion state detection using ResNet50 and CNN, resulting in 74.32 % accurate prediction. The last cases were then combined with late fusion and model accuracy has improved a lot resulting in 79.81% accurate prediction. But the proposed model has surpassed the results with late weighted fusion resulting in 83.79%, implicating that each feature separately works best for a certain class but with class based late weighted fusion, the accuracy of the model can be improved a lot. The results are shown in table 3.9. The proposed model has performed better than the state of art Biomodal deep belief network (BDBN) by Zhongmin et al. for discrete emotions [99].

**Table 3.9**. Valence- Arousal Prediction Results

| MODELS | Accuracy | F1-score |
|---|---|---|
| **SCL** | 65.54 | 0.64 |
| **ECG** | 72.89 | 0.71 |
| **tEMG** | 69.81 | 0.70 |
| **Feature fusion of psychophysiological signals** | 77.92 | 0.73 |
| **Video signals** | 74.32 | 0.75 |
| **Feature fusion of psychophysiological & late Fusion with video signals** | 79.81 | 0.77 |
| **Proposed model - MEmoR** | 83.79 | 0.81 |

The same 7 models have also been applied on the discrete emotions and it has been observed there too that weighted late fusion model works best in comparison to early fusion or single features only. These results are highlighted in table 13 along with state of art result.

**Table 3.10.** Discrete Emotion Prediction Accuracy

| | Accuracy | F1-score |
|---|---|---|
| **SCL** | 63.78 | 0.68 |
| **ECG** | 69.56 | 0.64 |
| **tEMG** | 70.93 | 0.61 |
| **Feature fusion of psychophysiological signals** | 73.29 | 0.76 |
| **Video signals** | 68.51 | 0.65 |
| **Feature fusion of psychophysiological & late fusion with video signals** | 76.32 | 0.74 |
| **BDBN model [33]** | 80.89 | - |
| **Proposed model - MEmoR** | 81.54 | 0.79 |

Figure 3.13 shows the variation in the results for two varied outputs categorized BioVid Emo dataset on all the 7 models.

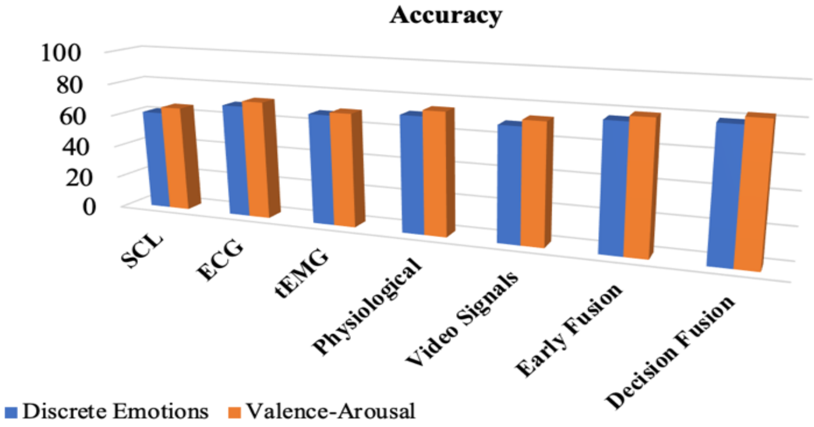



**Fig.3.13.** Accuracy of Discrete Emotion vs Valence-Arousal

The proposed deep learning weighted late fusion model has identified the discrete emotion as well as valence-arousal affective state effectively, figure 3.14 shows the confusion matrix of the proposed model discrete emotion dataset, with table 3.10 containing the accuracy of each emotional state for the proposed model with state-of-the-art results, highlighting that Amusement state has been identified better in comparison to other emotional states.

| | *Amusement* | *Fear* | *Disgust* | *Sad* | *Anger* |
|---|---|---|---|---|---|
| *Amusement* | 1798 | 20 | 14 | 43 | 37 |
| *Fear* | 47 | 1435 | 113 | 87 | 94 |
| *Disgust* | 38 | 51 | 1524 | 27 | 49 |
| *Sad* | 7 | 61 | 23 | 1704 | 95 |
| *Anger* | 12 | 69 | 57 | 112 | 1388 |

**Fig. 3.14**. Confusion matrix of the proposed model for discrete emotions

**Table 3.11.** Accuracy of Emotions

| Emotions | MEMOR MODEL | BDBM [33] |
|---|---|---|
| **Amusement** | **91.02** | 89.25 |
| **Anger** | 75.31 | 90.74 |
| **Fear** | 77.64 | 52.60 |
| **Disgust** | 80.38 | 86.45 |
| **Sad** | 79.92 | 85.42 |
| **Valence** | 84.16 | - |
| **Arousal** | 82.71 | - |

Figure 3.15 compares the most affective three models for the discrete emotions, that is, early fusion model of all three bio-signals, early fusion of three bio-signals late fused with visual signals, and lastly the proposed model.

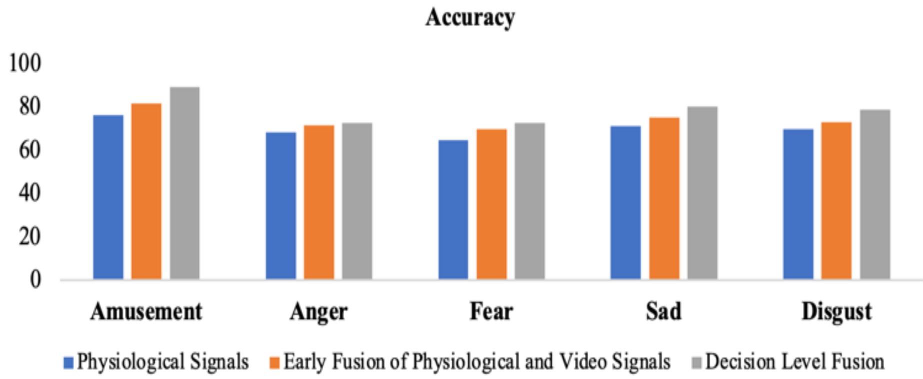



**Fig.3.15.** Comparison of Models for Discrete Emotions

### 3.2.3. DREAM: Deep Learning based Recognition of Emotions- Audio, Video, Physiological Modality

DREAM (Deep Learning-based Recognition of Emotions from Multiple Affective Modalities) is built using decision-level fusion of three different affective modalities, namely, audio (speech), video (face expressions) and physiological signals. Transfer learning is used to learn from existing data. The model is similar to earlier proposed model, but is different in terms of modalities and the fusion strategy.

#### 3.2.3.1. Methodology

VGG and ResNet have been used to pre-train the model, followed by fine tuning using CNN for audio signals. Video signals are processed through ResNet twice for facial emotion detection and is fine-tuned on K-EmoCon using convolution network. Physiological signals are processed with help of convolution models 1-D and 2-D. The output of all the modalities is combined using probability-based average decision fusion, it is one of the techniques of late fusion. The publicly available dataset for emotion recognition in conversations, K-EmoCon is used to train and test the approach. The architecture of DREAM-Smart is shown in fig 3.16. Initially nine input files from K-EmoCon dataset were provided to proposed model after pre-processing. Six out of nine files were of Emperica E4, Only EEG file taken from NeuroSky, Polar_H7's data had many missing values, and Heart rate has already been received from E4, so no file from Polar_H7 was processed. Video file of each participant, and

Separate audio file generated after pre-processing for the participant was used as the initial input to the model.

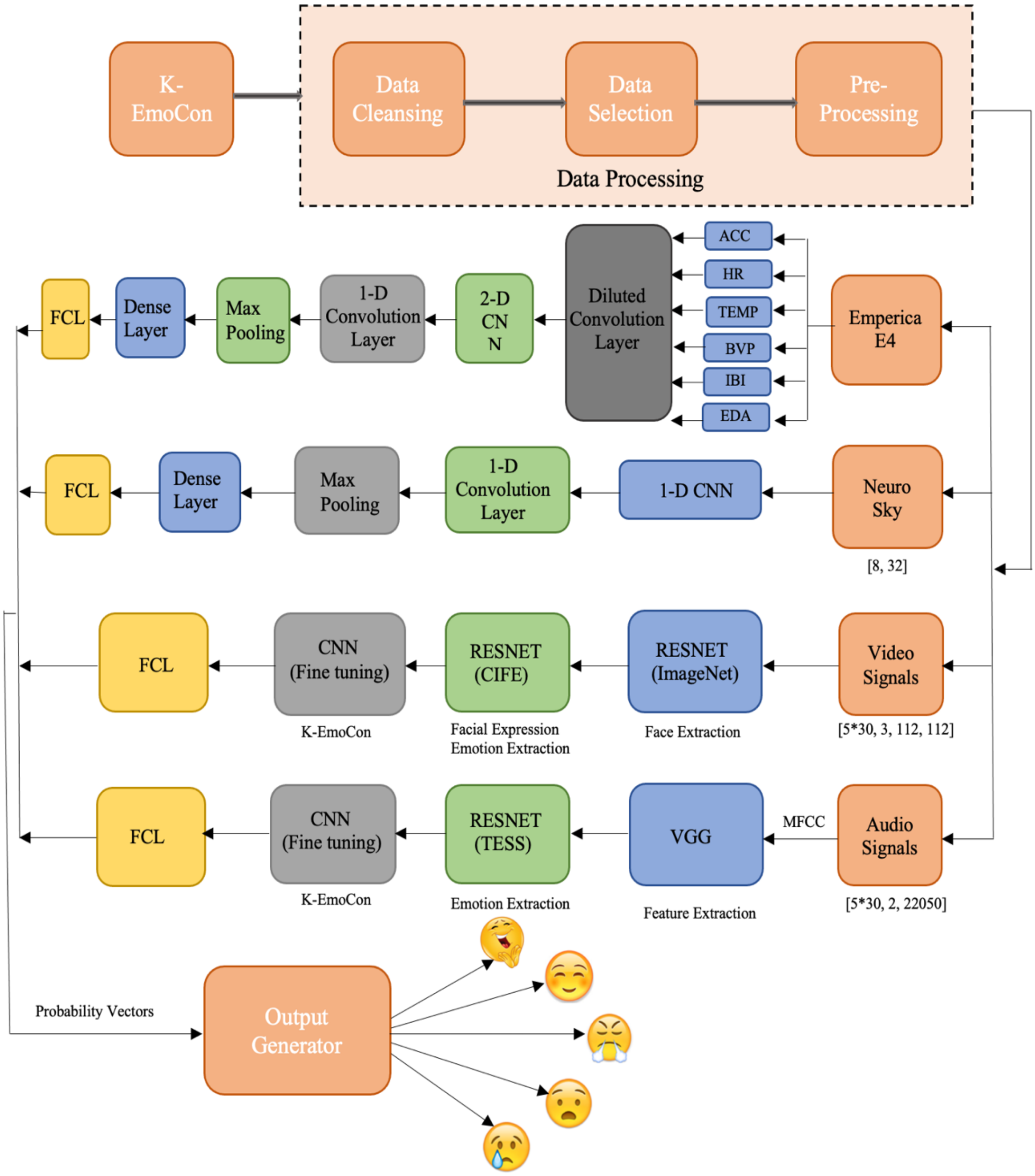



**Fig. 3.16.** Architecture of DREAM

### 3.2.3.2. Dataset

K-EmoCon: A multimodal publicly available dataset for emotion recognition in conversations provided by Park et al. in 2020 [77]. The dataset contains audio, video and bio signals of 32 subjects, who participated in a debate task on a social issue in teams of 2, while wearing Physiological signal measuring devices, Emperica E4, NeuroSky, and Polar H7, along with Video cameras for recording the facial expressions and gestures of the participants. On average, 10 minutes of debate was conducted between each pair, for emotion recognition, audio and video was recorded of this debate, along with Bio-signals from 3 wearable devices. ACC (32Hz), BVP (64 Hz), EDA (4Hz), HR(1Hz), IBI, and Temp (4Hz) was measured through Emperica E4; Attention, Brainwave – EEG (125Hz) and Meditation through NeuroSky; and ECG (1 Hz) from Polar H7.

K-EmoCon is the first publicly available dataset that contains natural conversation between non-actors, while monitoring their physiological changes. Another unique feature in K-EmoCon is that, each instance has been annotated by 7 people, one the person himself, second the opponent/partner in the debate, and lastly by 5 external observers. Although it makes processing the dataset difficult, but it provides different perspective for accurate emotion recognition. Park et al. have used 20 different types of emotions for annotation, including dimensional emotional model (Arousal and Valance), Basic Emotions (Cheerful, Happy, Angry, Nervous, and Sad), along with both common and less common BROMP (Baker Rodrigo Ocumpaugh Monitoring Protocol) Affective categories [77].

***Data Selection****:* K-EmoCon contains multiple files for each modality, not all the files are accessible or are having missing entities, data cleansing was performed. The audio recording of the dataset contains 16 files, each for the debate between the participants, so each file contains audio of 2 participants. Video Signals were recorded through separate cameras for each participant, but 11 of the participants didn't allow the data for public availability, we have only access of 21 participants visual signals. Bio-signal measuring devices too had some troubles. For our proposed model, we have taken audio, video, BVP, EDA, HR, IBI, Temp, and EEG signals for accurate Emotion recognition. As only 18 participant's complete data were available for all these signals, the empirical analysis has been conducted on 18 participants data only.

***Pre-processing****:* For emotion recognition, classification of emotions has been taken into 7 types, 2 affective dimensions and 5 emotion states. Emotion annotation of these 7 categories were ranked on scale 1-5 for dimensional, and 1-4 for emotions by all seven annotators. For our analysis we divided each emotion and dimension into two parts, low arousal if average score of all 7 annotators between 1-3, and high arousal if between 3-5, similarly very cheerful for average score of 2-4, and less cheerful for 1-2.

The experiment was evaluated for both Affective dimensions, and Emotion states separately, with 4 and 10 output classes respectively. To create time synchronous data for proposed work, the 18 participants' all signals were mapped with output (Dimensions, and Emotions), so each signal is annotated at 5 second interval with one of the four affective dimensions for first experiment, and similarly with one of the ten emotion state for second experiment.

All input files of every modality are mapped with the averaged annotated emotion. Speech signals of 22Khz sampling is cut at interval of 5 seconds and mapped to corresponding annotated emotion state. Similarly visual signals sampled with 30 fps with frames resized into 112*112 pixels is mapped to 5 second annotated emotion by mapping 150 (30*5 frames) to single emotion instance. The bio-signals after pre-processing, is mapped to output emotion class, by using statical feature of mean for taking 5 second data as an instance.

The bio-signals and visual signals are recorded for each participant separately, but the audio signals are recorded in sharing for the team of 2 participants. The audio signals of the opponent can trigger certain emotions in the subject, but it cannot reflect the emotions of the subject, so while pre-processing, we manually generate audio file for each participant from original file. First, a duplicate copy of the file was created for the opponent, and then manually the voice of the opponent was identified and removed from the file (sections) and replaced with the voice signals of the participant at start of the debate. Similar process was followed for opponent copy. For bio-signals, in pre-processing, missing values were replaced by means of the next two consecutive instance values.

#### 3.2.3.3. Audio Modality

After generating separate audio files for each participant, the speech signals of sampling rate 22KHz, was provided as an input ([T, C, F]), where T represents time i.e., 5s, C represents Channel i.e., 2, F represents frequency 22050Hz. Mel-Frequency Cepstral Coefficients (MFCC) waveform encoding was used to extract audio features. For mapping audio to visual signals at same time instance, we reframed the audio signal size to that of visual signals, so temporal dimension after MFCC of audio signal became 150 (30*5). These extracted features are then provided as an input to VGG a pre-trained model [100] followed by ResNet [101] for pre-training the model for emotion recognition on TESS (Toronto Emotional Speech set) dataset. VGG creates a simple but unified deepen structure of the network, ResNet uses residual learning, that ease the training process [102]. These pre-trained models are then fine-tuned by using K-EmoCon for mapping the output to 10 emotion states. The output of fine-tuned layer is then passed to a fully connected layer (FCL).

#### 3.2.3.4. Video Modality

After mapping emotional state at each segment of 5 second interval, the size of individual video clip became [5 x 30, 3, 112, 112] represented as ([T x F, C, W, H]), where T is time, F for frames per second, C for color channels, W and H stands for width and height respectively. This input feature vector is feed to pre-trained ResNet on ImageNet for face extraction. The output of this ResNet model is then given to second pre-trained ResNet model on CIFE for emotion mapping of facial expressions. These pre-trained models have different with output classes, so it is then fine-tuned on the dataset itself. These visual signals can be used to identify the body gestures such as hand movements, but as the movement of participants is limited (seated on a chair), we just took facial expressions into consideration for emotion classification. Finally, the output from fine-tuned model is feed into a fully connected layer to predict the emotion state.

#### 3.2.3.5. Physiological Modality

Seven Physiological signals have been measured by 2 wearable sensor devices Emperica E4, and NeuroSky. The band pass filtering was applied on EEG signals (8 bathes, 32 Hz), with a window size of 5 seconds to map the emotion state. The input is then provided to a convolution layer with kernel of size 3 * 1, with 64 kernel and the pooling size 3*3. The output of pooling layer is again provided to 2D convolution layer, followed by max pooling. The output of this pooling layer is provided to dense layer followed by a fully connected layer. For 6 bio signals of Emperica E4 wrist band, as the size of each signal is different, like ACC is 3*126 (for each participant), and HR 1*126, initially a diluted convolution layer is applied to all 6 bio-signals, then after generating the same size feature map, it is provided as an input to two 1D convolution layers similar to EEG signal, followed by Dense layer, and fully connected layer to produce emotion state.

### 3.2.3.6. Average Decision Fusion

The final output generator provided output by taking in emotion state produced by 4 fully connected layers and providing output by using average weighted decision fusion model. In this method, we take output of all 4 sub-models produced by the SoftMax function resulting in probabilities for all 10 emotion states. Suppose the 4 sub-models are represented by $E_V$, $E_A$, $E_{EEG}$, $E_{E4}$, Where $E_V$ stands for output vector of Emotions produced by Visual modality, $E_A$ stands for output vector of Emotion of Audio modality, $E_{EEG}$ for output vector of EEG signal, and lastly $E_{E4}$ gives the output vector produced by 6 bio-signals of Emperica E4 device.

Each output vector will contain probabilities of all 10 classes produced for that instance, say $P_{v1}$, $P_{v2}$, $P_{v3}$, $P_{v4}$, $P_{v5}$, $P_{v6}$, $P_{v7}$, $P_{v8}$, $P_{v9}$, $P_{v10}$, where $P_{v1}$ represents, the probability of the output to be less cheerful and $P_{v6}$ represents probability of belonging to very cheerful for visual modality model. This Probability value is generated by SoftMax activation function displayed in (10), employed after fully connected layers in each sub-model.

$$SoftMax\ (P_{Vi}) = \frac{\exp(P_{Vi})}{\sum_i exp(P_{Vi})} \qquad (10)$$

Then in final output generator, probability of corresponding class will be added with probabilities of same class output from different sub-model as shown in (11).

$$P_i = P_{Vi} + P_{Ai} + P_{EEGi} + P_{E4i} \qquad (11)$$

After calculating probability of each output class, the counter classes' probabilities are compared with each other, the one with more value will be provided as output, resulting in 5 outputs for a single instance.

$$E = max(P_{i,} P_{n-i}), i \in (1, \ldots\ldots, n) \qquad (12)$$

From (12), it can be observed that for "n" output classes, there will be n/2 outputs. For all the fully connected layers in four sub-models, Adam optimizer has been used, and loss is calculated by categorical cross-entropy as shown in (13).

$$Loss = -\sum\left(y'_{i1}log(y_{i1}) + y'_{i2}log(y_{i2}) + \cdots + y'_{in}log(y_{in})\right) \qquad (13)$$

$y_{i1}, y_{i2}, y_{in}$are internal node labels, $y'_{i1}$ , $y'_{i2}$ , $y'_{in}$ are the output layer nodes, produced by SoftMax function.

### 3.2.3.7. Findings

For performance evaluation the training and testing of the model is performed using "leave one out". The size of the dataset is limited, having all the signals recorded only for 10 minutes for each of the 32 participants, generating total of 320 minutes data approximately. But only taken 18 participants data was taken for model execution, resulting in almost half the total data. Therefore, to have large training set, leave one out strategy was employed, implying that first model will be trained on 17 participants complete data, will be tested for performance evaluation of 18$^{th}$ participant. Then the same process will be repeated leaving another participant data for testing. So, the model will be executed 18 times for both types of output categories. The execution of the model has been performed on MacBook Pro with M1 pro processor, 16GB RAM, 10 core GPU, and 14 core CPU. To evaluate the performance of proposed multimodal, F1-score has also been computed along with Accuracy.

The proposed work has three modalities, each executed separately in synchronous manner, combining the output of final layers of each model into Output layer. The performance evaluation of proposed model for both output categories are shown in table 3.12 and table 3.13. Here LA represents Low arousal, LV Low Valance, HA for

High Arousal and HV for High Valance. For emotional states highest F1-score was obtained for very angry emotion for $P_{15}$, and highest accuracy of 81.7 for $P_{21}$ participant for Less Nervous emotion. For Dimensional Affect Arousal and Valance, the highest accuracy of 82.4 and F1-score of 92.6 was achieved.

**Table 3.12.** Model Performance for Dimensional States

| | **$P_{1}$** | **$P_{4}$** | **$P_{8}$** | **$P_{9}$** | **$P_{10}$** | **$P_{11}$** | **$P_{15}$** | **$P_{18}$** | **$P_{19}$** | **$P_{20}$** | **$P_{21}$** | **$P_{22}$** | **$P_{23}$** | **$P_{24}$** | **$P_{25}$** | **$P_{26}$** | **$P_{29}$** | **$P_{30}$** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **F1-Score** | | | | | | | | | | | | | | | | | | |
| **LC** | 57.9 | 71.2 | 69.8 | 61.7 | 64.3 | 66.2 | 68.7 | 70.1 | 65.2 | 67.2 | 72.3 | 59.4 | 65.4 | 68.2 | 67.9 | 63.6 | 61.8 | 68.1 |
| **VC** | 70.5 | 69.3 | 72.3 | 79.7 | 76.1 | 73.5 | 68.5 | 69.3 | 70.4 | 61.9 | 78.4 | 73.7 | 62.7 | 69.5 | 69.8 | 66.3 | 70.3 | 75.8 |
| **LH** | 73.8 | 67.3 | 69.4 | 78.4 | 61.1 | 54.9 | 66.9 | 64.5 | 77.9 | 72.5 | 83.6 | 44.7 | 70.2 | 65.8 | 73.5 | 71.9 | 74.9 | 79.1 |
| **VH** | 81.9 | 74.9 | 78.5 | 80.3 | 72.8 | 75.1 | 83.4 | 73.9 | 71.8 | 72.9 | 51.4 | 78.2 | 68.9 | 72.3 | 70.3 | 68.4 | 63.7 | 75.3 |
| **LA** | 68.7 | 64.5 | 70.3 | 68.4 | 72.7 | 69.3 | 71.6 | 74.8 | 75.4 | 65.2 | 80.1 | 75.4 | 67.3 | 82.5 | 67.1 | 71.9 | 66.9 | 80.7 |
| **VA** | 75.4 | 58.4 | 68.4 | 67.6 | 78.3 | 72.6 | 87.8 | 51.9 | 70.3 | 81.9 | 71.8 | 71.7 | 76.4 | 72.1 | 74.9 | 72.3 | 71.5 | 57.9 |
| **LN** | 56.2 | 65.8 | 71.2 | 75.4 | 74.1 | 79.4 | 75.5 | 66.3 | 73.9 | 72.6 | 79.5 | 72.8 | 79.8 | 77.9 | 64.8 | 80.1 | 78.4 | 82.9 |
| **VN** | 69.2 | 72.5 | 75.3 | 76.1 | 63.8 | 68.4 | 82.7 | 78.2 | 67.3 | 69.8 | 64.9 | 81.6 | 54.9 | 76.8 | 75.8 | 73.6 | 40.6 | 79.4 |
| **LS** | 79.3 | 68.5 | 72.4 | 73.4 | 80.4 | 81.9 | 69.5 | 68.7 | 64.5 | 55.8 | 81.6 | 68.2 | 82.3 | 79.5 | 69.4 | 78.9 | 63.5 | 69.9 |
| **VS** | 74.6 | 83.6 | 76.7 | 72.5 | 75.8 | 78.4 | 78.2 | 66.1 | 62.6 | 79.3 | 76.7 | 79.6 | 74.8 | 73.1 | 77.2 | 82.4 | 79.9 | 73.6 |
| **Accuracy** | | | | | | | | | | | | | | | | | | |
| **LC** | 68.4 | 63.5 | 75.3 | 71.2 | 69.9 | 68.1 | 64.5 | 74.7 | 69.9 | 64.3 | 70.4 | 66.7 | 71.7 | 73.2 | 71.9 | 66.8 | 78.9 | 68.7 |
| **VC** | 66.1 | 68.7 | 51.9 | 68.4 | 59.8 | 70.2 | 62.6 | 76.1 | 73.5 | 68.5 | 68.3 | 49.8 | 68.7 | 71.8 | 72.3 | 72.6 | 74.1 | 57.6 |
| **LH** | 70.8 | 71.8 | 66.3 | 70.2 | 75.3 | 65.7 | 68.4 | 68.1 | 72.7 | 70.8 | 66.1 | 72.5 | 75.1 | 63.8 | 64.5 | 67.6 | 78.3 | 72.6 |
| **VH** | 69.5 | 75.4 | 72.6 | 69.3 | 68.7 | 71.8 | 67.6 | 78.3 | 72.6 | 58.7 | 63.7 | 59.8 | 69.3 | 61.7 | 62.6 | 74.8 | 66.2 | 71.9 |
| **LA** | 64.5 | 75.1 | 68.5 | 70.3 | 76.1 | 69.3 | 72.3 | 74.7 | 75.1 | 62.5 | 66.9 | 69.3 | 80.2 | 66.2 | 63.6 | 61.8 | 68.9 | 72.3 |
| **VA** | 68.9 | 69.3 | 71.9 | 73.9 | 74.9 | 72.4 | 63.9 | 70.3 | 79.4 | 67.4 | 71.5 | 75.3 | 76.1 | 67.8 | 74.7 | 65.4 | 54.9 | 71.7 |
| **LN** | 67.3 | 72.6 | 64.8 | 67.3 | 64.5 | 70.6 | 70.4 | 74.9 | 46.7 | 69.5 | 81.7 | 72.4 | 73.4 | 51.7 | 70 | 62.7 | 72.3 | 64.5 |
| **VN** | 72.1 | 64.5 | 70.6 | 72.6 | 68.9 | 68.4 | 62.1 | 73.9 | 72.5 | 68.3 | 45.8 | 76.9 | 71.8 | 72.9 | 68.4 | 70.2 | 69.4 | 62.6 |
| **LS** | 73.2 | 62.6 | 68.4 | 68.1 | 67.3 | 66.9 | 54.7 | 67.3 | 59.4 | 65.4 | 68.2 | 76.1 | 73.5 | 68.5 | 44.3 | 68.9 | 63.6 | 61.8 |
| **VS** | 71.6 | 48.7 | 69.2 | 63.5 | 76.4 | 62.3 | 66.3 | 70.2 | 68.7 | 71.2 | 73.8 | 63.6 | 61.8 | 69.3 | 72.6 | 67.3 | 71.8 | 72.9 |

**(LC: Less Cheerful, VC: Very Cheerful, LH: Less Happy, VH: Very Happy, LA: Less Angry, VA: Very Angry, LN: Less Nervous , VN: Very Nervous , LS: Less Sad , VS: Very Sad)**

**Table 3.13.** Performance of Proposed Model

| **F1-Score** | **$P_{1}$** | **$P_{4}$** | **$P_{8}$** | **$P_{9}$** | **$P_{10}$** | **$P_{11}$** | **$P_{15}$** | **$P_{18}$** | **$P_{19}$** | **$P_{20}$** | **$P_{21}$** | **$P_{22}$** | **$P_{23}$** | **$P_{24}$** | **$P_{25}$** | **$P_{26}$** | **$P_{29}$** | **$P_{30}$** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **LV** | **89.6** | 87.6 | 83.6 | 83.1 | 80.1 | 79.8 | 68.7 | 73.9 | 71.8 | 72.9 | 68.4 | 79.2 | 69.4 | 68.2 | 77.2 | 82.4 | 79.9 | 73.6 |
| **HV** | 84.3 | 79.7 | 78.9 | 91.7 | 76.1 | 81.3 | 68.5 | 69.3 | 70.4 | 61.9 | 78.4 | 73.7 | 62.7 | 71.5 | 75.3 | 76.1 | **92.6** | 83.7 |
| **LA** | 79.8 | **91.7** | 81.3 | 81.4 | 84.2 | 83.5 | 66.9 | 64.5 | 77.9 | 80.5 | 83.6 | 81.4 | 70.2 | 65.8 | 73.5 | 71.9 | 74.9 | 84.1 |
| **HA** | 83.2 | 78.9 | 78.5 | 80.3 | **90.7** | 86.7 | 83.4 | 73.9 | 71.8 | 72.9 | 80.6 | 78.2 | 68.9 | 72.3 | 70.3 | 71.4 | 81.5 | 75.3 |
| **Accuracy** | | | | | | | | | | | | | | | | | | |
| **LV** | 75.4 | 58.4 | 68.4 | 72.3 | 74.7 | 75.1 | 82 | 57.8 | 77.2 | **82.4** | 79.9 | 73.6 | 76.4 | 72.1 | 74.9 | 72.3 | 71.5 | 68.3 |
| **HV** | 56.2 | 65.8 | 71.2 | **81.4** | 73.7 | 79.4 | 75.5 | 66.3 | 73.9 | 72.6 | 79.5 | 72.8 | 79.8 | 77.9 | 64.8 | 80.1 | 78.4 | 78.1 |
| **LA** | 69.2 | 72.5 | 75.3 | 77.8 | 79.7 | **80.2** | 82.7 | 78.2 | 67.3 | 77.2 | 79.4 | 79.9 | 73.6 | 76.8 | 77.2 | 78.1 | 76.9 | 70.6 |
| **HA** | **79.8** | 69.6 | 76.3 | 70.7 | 76.4 | 72.1 | 74.9 | 72.3 | 76.3 | 77.3 | 74.9 | 72.4 | 74.7 | 65.4 | 72.5 | 75.3 | 76.1 | 76.4 |

The ablation study on the proposed model was performed as well, to analyze the impact of each modality. The Physiological modality performs best individually, followed by video modality, and the accuracy achieved by only speech signals for emotion detection was the least of all three. The best accuracy achieved by each modality in the ablation study is shown in figure 3.17. The variation in results in each individual modality is impacted by the amount of the data available as well, and the bio-signals were available for more participants in comparison to video or audio signals.

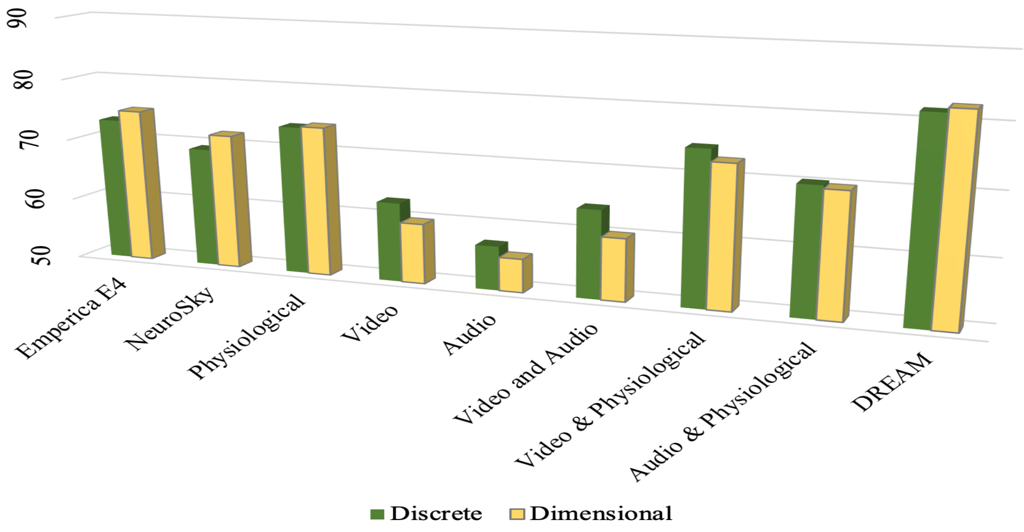



**Fig. 3.17.** Ablation Study

## 3.3. Chapter Summary

This chapter discuss various models proposed for handling multi-modalities of biomarkers for identifying the affective psychological state as well as affective emotional state. For affective psychological state, two models proposed focused on physiological biomarkers, and psycholinguistic biomarkers separately. For affective emotional state, the physiological as well as visual biomarkers were used with different combinations and fusion strategies.

***

# CHAPTER 4

# ENHANCED MULTI-MODAL AFFECTIVE STATE RECOGNITION

Advances in technologies have the potential to influence and shape the society. Upsurge in IoMT devices has made the remote healthcare a reality. The bio-signals of patients can be monitored from remote location by medical professionals, making it possible for everyone to have access to healthcare. Remote monitoring helped the medical community as well, only critical patient needs to be kept in hospitals, making room for emergencies and other patients. With advancement in IoT sensors, the monitoring of patient's health remotely has created a large resource of electronic health records (EHR). The EHR can be accessed by doctors remotely, patient can consult various doctors, without worrying for their multiple body tests, as patient would be able to share their existing EHR with the new doctor.

Though the benefits offered are unparalleled and promise useful decision support information, but all this is challenged by a lot of noise created owing to the large volume and variety of information sent at almost light speed. One of the most significant threats that the IoMT poses is of data security & privacy. As these devices capture and transmit data in real-time with no standard data protocols and data ownership regulations, it makes it highly susceptible to hacks and frauds. Moreover, non-uniformity of the connected device's protocols make integration of multiple devices (data aggregation) difficult thereby reducing the scope of scalability of IoT in healthcare. At the same time, data overload and accuracy further create complications. The IoMT devices record a ton of data and utilize it to gain vital insights. However, the amount of data is so tremendous that deriving insights from it is becoming extremely difficult for doctors which, ultimately affects the quality of decision-making. Undeniably, the sheer amount of IoMT data traffic with the rise in connected devices and its immediacy makes data management & decision analysis a pressing issue. That is, despite this proliferation of data, only a small fraction is being used for decision making because the data cannot be stored or transmitted efficiently. As a result, the full value of analyzing the data for timely decision making has not been realized. When healthcare practitioners must make life-or-death decisions, the quality of information at their disposal is critical. Having more specific data — and being able to access it in real time — leads to more informed decisions. IoMT makes this possible through an infrastructure of connected medical devices, software applications, and health systems powered by 5G wireless technology and edge computing, which enables connected devices to process data closer to where it is created.

## 4.1. Genetically Optimized Fuzzy C-Means Data Clustering

IoMT sensors such as wearables, ingestible sensors and trackers have the potential to provide a proactive approach to healthcare. But grouping, traversing and selectively tapping the IoMT data traffic and its immediacy makes data management & decision analysis a pressing issue. Evidently, the selection process for real-world, time-constrained health problems involves looking at multivariate time-series data generated simultaneously from various wearables resulting in data overload and accuracy issues. Computational intelligence of edge analytics can extend predictive capability by quickly turning digital biomarker data into actions for remote monitoring and trigger alarm during emergency incidents without relying on backend servers. But the pervasive generation of data streams from IoMT levies significant issues in data visualization and exploratory data analysis. In this work, we utilize the genetically optimized Fuzzy C-means data clustering technique for affective state recognition on the edge. Clustering segregates the biomarker data in chunks and generates a summarized data for each subject which is then genetically optimized to avoid stagnation in local optima. A multi-level convolution neural network is finally used to classify the affective states into the baseline, stress and amusement categories. The model is evaluated on the publicly available WESAD dataset.

### 4.1.1. Methodology

The data reduction techniques can certainly reduce the size and complexity of real-time data streams by converting it into more coherent and manageable proportions, highlighting the relevant features of the data more clearly and eventually facilitating more accurate and efficient edge analytics. Based on this, the proposed genetically optimized Fuzzy C-means data clustering model proffers a fast analytical method for detecting the affective state of the user using data reduction at the edge. The underlying impetus is to create quality clusters with reduced time complexity to get the most significant information. The model has been implemented on the publicly available WESAD dataset, containing data of 15 subjects (patients) measured through IoT-based chest wearable and wrist wearable device for a period of 2 hours with total of 12 biomarkers.

The model consists of the following four architectural components:

- Fuzzy C-means clustering.
- Genetic algorithm for optimized dataset.
- Data reduction by summarizing the clusters.
- A deep hierarchical model trained using optimized dataset to detect the affective state. The architecture of the proposed model is shown in figure 4.1.

**4.1.1.1. Fuzzy C-means Clustering:** The first component involves defining and initializing clusters for that we used the observation made by Siirtola in 2019 [103], that the affect perceived depends upon the time window taken and the results concluded that affect can be identified with better accuracy with 120 second time window as compared to 15 seconds, 30 seconds, 60 seconds and 90 seconds time window. Analyzing every single second data may not be effective, but at the same time to be quick to detect any physiological change in the body, the data was reduced by margin of 3 seconds. The base hypothesis taken for identifying the best suitable number of clusters for data reduction is chosen similarly by taking different time window and observing the accuracy in each scenario. As WESAD contains the data for a window of 2-hours, i.e., of 7200 seconds, to initialize the clusters, 1440 clusters are used as the model performed optimally with the hypothesis that a physiological change due to any external stimuli will last for at least 5 seconds, as shown in result section. As a person may exist in neutral state before and after an event it is imperative to analyze both these for determining affective state of an individual. Therefore, to distinguish between pre- and post-state of a subject, it is ensured that no two-time varied data belongs to the same cluster with a time criterion added to the initial data table. Time is provided with the numerical value ranging from 1 to 7200, based on the second it belongs to. This extra criterion while clustering ensures that no pre- and post-data is combined based only on the biomarker measures, as different affective states show different physiological changes in an individual, for some having heartbeat rate 85 could be a warning sign showing a hike, whereas for other it may be average heartbeat rate because of their daily activities. Affect visualization can be done by observing the change in pre- and post-condition, therefore time is an essential feature to understand the affective state.

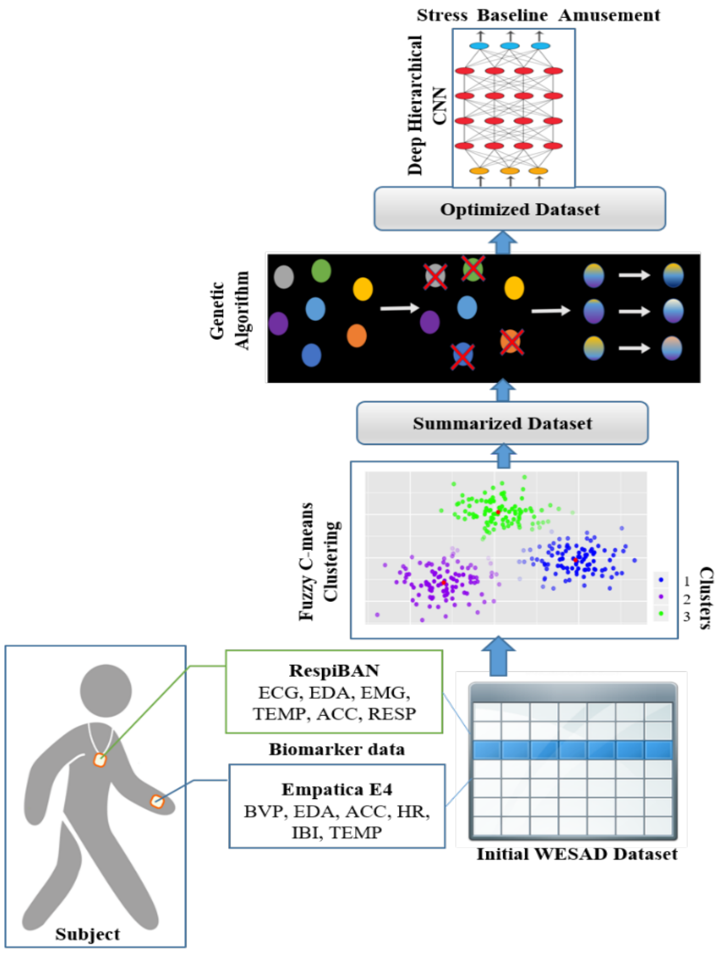



**Fig.4.1.** Architecture of Genetically Optimized Fuzzy C-Means Data Clustering model

Fuzzy C-Means performs clustering by iteratively searching for a set of fuzzy clusters and the associated cluster centres that represent the structure of the data as best as possible. Given a number of clusters c, it partitions the data $X = \{x_1,x_2,\ldots,x_n\}$ into c fuzzy clusters by minimising the within group sum of squared error objective function. The algorithm stops when either the error is below a certain tolerance value or its improvement over the previous iteration is below a certain threshold. Fuzzy C-means uses fuzzy membership, to cluster the data, it allows an instance to belong to different clusters with different membership value [30]. As the minor change in a single feature may be because of any reason, like pinching by a friend, it last for few seconds, but it may not change the affective state of a user, but certainly brings a physiological change. It is imperative to provide number of clusters for executing the Fuzzy c mean, to identify this optimized number of clusters, we have performed empirical analysis on varying time window hypothesis to identify the most accurate time window clustering by generating clusters based on 3 seconds, 4 seconds, 5 seconds, 6 seconds and 7 seconds. Comparison of different number of clusters has been discussed in the result section. The most accurate results have been shown when the initial clusters were defined using 5 second similarity hypothesis, resulting in 1440 clusters for two hours of period. The 1440 clusters initialized will initially having second tuple of the interval, i.e., first cluster will initially contain 2nd tuple, and 2nd cluster will contain 7th second instance, and so on. Once the number of clusters has been defined, and clusters has been initialized, rest of the tuples are now assigned to each cluster based on the objective function (f) as defined in (14).

$$f = \sum_{j=1}^{k} \sum_{x \epsilon C_j} u_{ij}^{m} \left(x_i - u_j\right)^2 \tag{14}$$

$$u_j = \frac{\sum_{x \epsilon C_j} u_{ij}^{m} x}{\sum_{x \epsilon C_j} u_{ij}^{m}} \tag{15}$$

where,

$u_{ij}$ is the membership value, i.e. degree to which a particular observation $x_i$ belongs to a cluster $c_j$.

$u_j$ represents the center of cluster j, which is the mean of all points, weighted by their degree of belongingness to the cluster, as shown in (14).

$\left(x_i - u_j\right)$ is the Euclidean distance between i$^{th}$ data and j$^{th}$ cluster's center.

$m$ is the fuzzifier, i.e., it defined the level of cluster fuzziness, it can have any value greater than 1, we have taken m=2. The membership value and fuzziness can be computed as shown in (16):

$$u_{ij}^{m} = \frac{1}{\sum_{i=1}^{k} \left(\frac{\left|x_i - u_j\right|}{\left|x_i - c_k\right|}\right)^{\frac{2}{m-1}}} \tag{16}$$

i.e., membership $u_{ij}$ is inversely linked to the distance of an entity from the center of the cluster.

The output derived from the first component is a summarized data table, reducing the size by a large margin. For each subject, data can be represented as 1440*17. Initially an individual subject has total of 4498417 instances, a raw discrete decision matrix of subject have 17 attributes, one for time, 8 for wrist-based sensor (5 for BVP, EDA, HR, IBI, TEMP and 3 for ACC, as ACC records the movement of the individual in all 3 directions.) collected through Empatica E4, and 8 from a chest based sensor (5 for ECG, EDA, EMG, TEMP, RESP and 3 for ACC) collected by RespiBAN, where biomarker values of Empatica E4 is variation from a predefined base value, RespiBAN.

**4.1.1.2. Genetic Optimization:** The second component of the proposed model optimizes the clusters, by generating the optimal mapped data to the clusters by minimizing the objective function. As the main issue with a partition clustering Fuzzy C-means is that it requires Apriori specification of the number of clusters and moreover, this clustering is done on the basis of homogenous data points whereas as discussed with the real-time IoMT time series data analysis, considering the pre- and post-data temporal correlations is important, that is, as the physiological change can be observed for more than 5 seconds, the number of clusters can be defined by dividing the time frame accordingly. Therefore, optimizing the clusters are important. A genetic model typically has an initial random population of individuals or solutions, which in this case is provided by the solutions generated by Fuzzy C-means, followed by the fitness evaluation of these individuals, using a fitness function. Once, the fitness function value has been generated for each individual the following procedure is repeated until one of the terminating conditions has been met. First, select the best individual, based on fitness value. Then, generate new population, by using crossover and mutation. Thirdly, evaluate the fitness of new population. And at last, replace the poor performer with new best performers.

This process is repeated until either the number of iterations set has been completed or no new solution can be generated, or the fitness of new population comes to be lower than already existing individuals. The steps involved in the second phase are:

- *Initialization:* Generally, the initial population is generated randomly, allowing the entire range of possible solutions (the search space). But for our normalized dataset, we start by initializing all the initial clusters formed by Fuzzy C-Means as the initial population generation.
- *Selection:* In our problem the Fitness function calculates the position of centroids obtained through Fuzzy c-means. This is done repeatedly till the centroids keep moving. This produces an optimised separation of alternatives into clusters. Once better individuals are determined, they replace the worst individuals in the group and the process is repeated. The breeding of the new generation is done through two Genetic operators: crossover or mutation.
- *Crossover:* Crossover is a genetic operator used to change or "evolve" an individual from one generation to the next. It is analogous to reproduction and biological crossover, upon which genetic algorithms are based. For our dataset, any 2 individuals are chosen for a single-point crossover to produce new children.
- *Mutation:* Mutation alters one or more gene values in a chromosome from its initial state. For our normalized dataset, a Uniform mutation is applied which adds a unit uniform random value to the user defined upper and lower bounds for that gene. The resultant mutants are then operated on by the fitness function.
- *Termination:* Termination can be done on the basis of the following conditions:
    - An optimised solution is obtained.
    - The highest-ranking solution's fitness is about to or has plateaued such that successive iterations no longer produce better results.

**4.1.1.3. Summarization:** Once the clusters have been identified, each cluster is represented by just one instance having median value of each attribute. This reduces the size of data directly to the total number of instances chosen at the initial level. Although the size has been reduced drastically, but accuracy of the model depends highly on the number of clusters.

**4.1.1.4. Deep hierarchal modelling**: After generating the optimized summarized data or information table, it is given as an input to the fourth final component, i.e., deep hierarchical model to train and detect the affective state of an individual. CNN is a sequence of convolutional layers, interspersed with activation functions. It is a deep neural architecture which has the power of self-tuning & learning skills by generalizing from the training data, it enhances feature extraction. The proposed Deep Hierarchical CNN model comprises of three levels, at first level ten 1-dimensional CNN (1D-CNN) layers were used to train the model on individual biomarkers, the output of first level was transformed to second level having two 1D-CNN, one each for generating an output for each wearable device (RespiBAN, & Empatica E4) and finally using model-level fusion the results of two 1D-CNN were combined to classify the affective state of an individual.

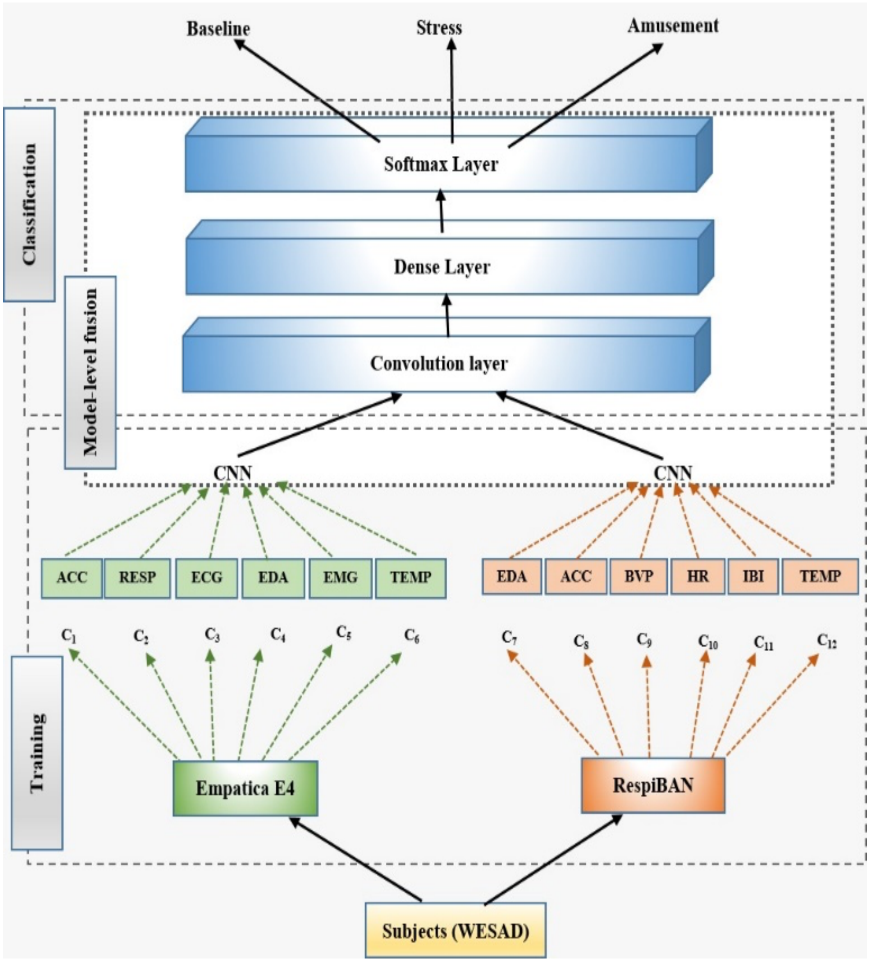



**Fig. 4.2.** Architecture of Deep Hierarchical Model

After the data has been fetched in the scalable format, data is still astronomical and exists as individual units, so we apply the 1D-CNN over each feature signal by constructing a separate CNN ($C_n$). Once all the $C_n$s gets trained, the input is passed in as subject manner, i.e., all the features of a subject at one sliding window or at a particular time is passed through the second level 1D-CNN, and the output of the two CNN layers is then passed into the primary model that classifies the subject at a particular instance into one of the three classes defined during the training phase using SoftMax layer. As each subject produces different bio-signals while being in stress, amusement or baseline, the convolution network while predicting can map a single instance into different classes having different probabilities, therefore application of logistic regression is very important to determine the final output class of the subject at a given instance. Due to 3 classes, SoftMax function is applied on top layer to map the output with highest probability class. The proposed hierarchical network is shown in figure 4.2.

#### 4.1.2. Dataset

WESAD is a publicly available data set for variable stress and affects detection collected by Schmidt et al. in a lab study on 15 subjects. Different biomarkers that were used hey to monitor the stress level of a person are BVP, ECG, EDA, EMG, TEMP, RESP and ACC. The data collected by RespiBAN was at 700Hz, whereas the data collected by the wrist device was at low resolution. Each subject has 12 features, and the results were self-reported by the user. The dataset contains a total of 63000000 instances. Each subject's data contains signals from two different sensor devices, RespiBAN recorded with 700Hz signal, the $C_n$s created for each of these feature is 1-D CNN with input layer 700*1 except for ACC which is provided with input layer of 700*3, as the ACC signals contain 3 dimension data. The $C_n$s for features recorded with Empatica E4, have different sized input layers, for ACC 32* 3, BVP has 64*2, TEMP and EDA have input size of 4*1 input layer.

#### 4.1.3. Algorithm

**Algorithm 1: Proposed Model**

**Input:** WESAD data, k=number of clusters,

**Output:** $Ac$ – Accuracy obtained

1: **Begin**: Choose the initial Cluster centroid

2: **Initialize**: Fuzzy Membership Matrix, U= $[u_{ij}]$

3: **Repeat while (**U(k+1)-U(K)< threshold)

4: Calculate new Euclidean distance

5: Calculate new Cluster Centroid

6: Update the Fuzzy Membership Matrix

7: **End-while**

8: **Initialize:** Fitness Population= U

9: **Define** Tournament for fit individuals

10: **Repeat While** (New Child !$\epsilon$ Population)

11: **Select** Fit individual←Random from Fitness Population

12: **Select** Child1 & Child 2←Random from Fitness Population less chromosome

13: Fit1←Fitness (Child1)

14: Fit2←Fitness (Child2)

15: Optimal Child←(Fit1>Fit2)? Child1: Child2

16: **Select** Parent ←Tournament (Fitness Population)

17: **Define** Crossover

18: **Define** Mutation

19: **End-while**

20: Calculate new Cluster Centroid

21: **Define** new Information Matrix

22: Update Information Matrix with only cluster centroids

23: **Train**← Information Matrix

24: **Return** (Ac)

#### 4.1.4. Findings

The proposed model was initially applied on the original WESAD dataset with 12 bio-signals. To evaluate the model, WESAD was pre-processed by applying min-max normalization for standardization so each subject have test results in the same range. 65% of data is used to train the model, i.e., ten subjects are used to train the model, and rest 35% i.e. 5 Subjects data is used to test the data, and 20% of the data (2 subjects) is used to validate the model. 53% of the total instances belong to the baseline class, 30% belongs to stress class, and 17% belong to the amusement class. The average model accuracy of 87.7% is achieved with the best subject-level accuracy of 96.98% achieved for Subject 2. It is observed that the accuracy of the model varies from 72% to 96%, whereas the F-1 score ranges from 0.612 to 0.998. The average accuracy achieved by the proposed model is better than the state-of-the-art results.

As the impact of each stimuli on a subject can be observed for few seconds, the data reduction is done by clustering the data which belongs to a certain time window and subsequently optimizing the clusters. To find the best time window, in which the affective state can be detected accurately while having fast execution time, the proposed architecture was evaluated multiple times on different variations. The execution time taken for each subject ranges from 1.37 seconds to 1.64 seconds, with an average execution time of 1.51 seconds. The variation in execution time and accuracy in each case can be visualized in figure 4.3 and 4.4 respectively. It was observed that the average model accuracy slightly reduced with increase in reduction ratio of the data, but the execution time was reduced considerably with the summarized data. The optimal trade-off between accuracy and execution time was observed with 1440 clusters, having time frame window of 5 seconds. As we increase the time frame window to reduce the number of clusters, the execution becomes fast, but the accuracy declines as well. Therefore, we propose clustering the data at an interval of 5 seconds which can reduce the latency time and can provide an accurate detection of the affective psychological state of an individual.

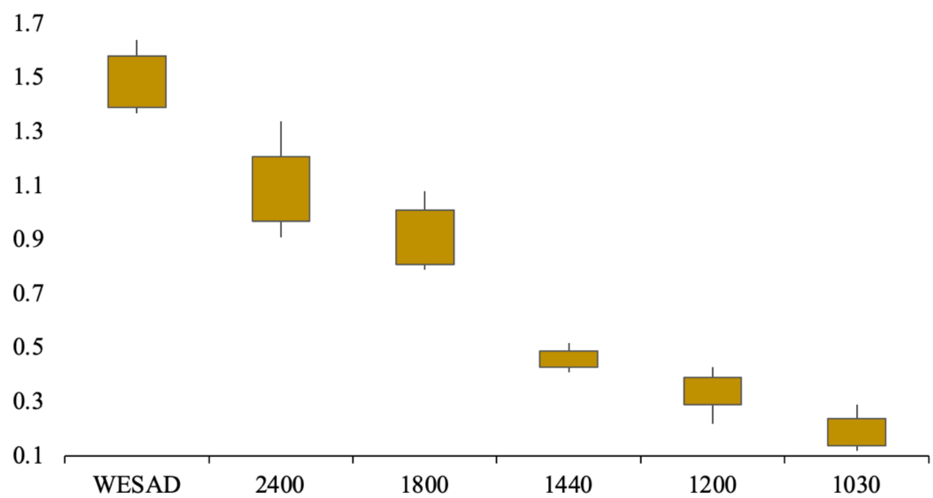



**Fig. 4.3.** Execution time of different clusters

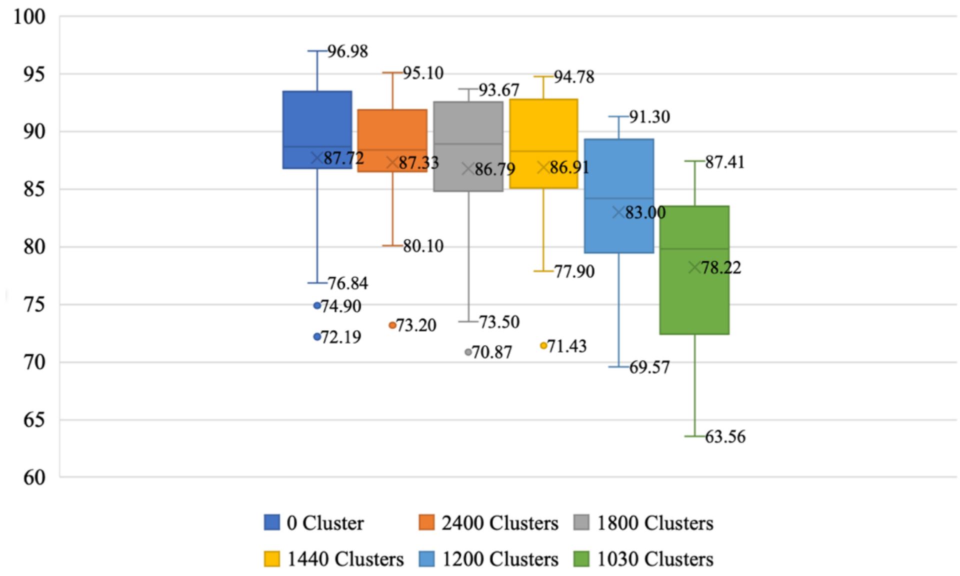



**Fig. 4.4.** Accuracy of different clusters

The accuracy and execution time comparison of both the original WESAD and the summarized WESAD is shown in table 15 and in figs. 4.3 and 4.4. The table shows the performance of the proposed model in terms of average accuracy, F1-score, and execution time of each subject after 5-cross validation for original WESAD dataset, and the summarized WESAD with 1440 clusters with time frame window of 5 seconds for clustering. In case of summarized data, best accuracy of 94.78% was achieved for Subject 11, with an average accuracy of 86.8%. And the execution time taken for each summarized subject data ranges from 0.41 seconds to 0.52 seconds, with an average execution time of 0.46 second.

**Table 4.1.** Performance of Proposed Model

| **Subjects** | **WESAD** | | | **Summarized WESAD** | | |
|---|---|---|---|---|---|---|
| | **Accuracy** | **F-1 Score** | **Execution Time** | **Accuracy** | **F-1 Score** | **Execution Time** |
| **S1** | 93.39 | 0.946 | 1.64 | 92.81 | 0.925 | 0.52 |
| **S2** | 96.98 | 0.983 | 1.53 | 94.03 | 0.974 | 0.49 |
| **S3** | 88.70 | 0.612 | 1.61 | 87.95 | 0.721 | 0.48 |
| **S4** | 95.07 | 0.929 | 1.47 | 92.81 | 0.934 | 0.46 |
| **S5** | 74.9 | 0.968 | 1.58 | 71.43 | 0.815 | 0.42 |
| **S6** | 87.92 | 0.855 | 1.39 | 88.56 | 0.837 | 0.51 |
| **S7** | 86.79 | 0.860 | 1.56 | 84.2 | 0.892 | 0.47 |
| **S8** | 72.19 | 0.684 | 1.37 | 71.5 | 0.656 | 0.43 |
| **S9** | 87.24 | 0.798 | 1.44 | 88.3 | 0.813 | 0.41 |
| **S10** | 92.73 | 0.935 | 1.52 | 91.8 | 0.908 | 0.46 |
| **S11** | 93.46 | 0.834 | 1.48 | 94.78 | 0.851 | 0.51 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **S12** | 87.24 | 0.858 | 1.39 | 86.78 | 0.893 | 0.48 |
| **S13** | 88.72 | 0.931 | 1.62 | 87.34 | 0.899 | 0.44 |
| **S14** | 76.84 | 0.649 | 1.59 | 77.9 | 0.715 | 0.49 |
| **S15** | 93.56 | 0.917 | 1.42 | 92.5 | 0.878 | 0.43 |
| **Average** | 87.7 | 0.852 | 1.51 | 86.8 | 0.848 | 0.46 |

It is observed that the accuracy of the model varies from 71.4% to 94.78%, whereas in case of non-summarized the accuracy varied from 72% to 96% as shown in figure 4.5. Figure 4.6 highlights the time variation the model has shown while execution for summarized vs. non-summarized data.

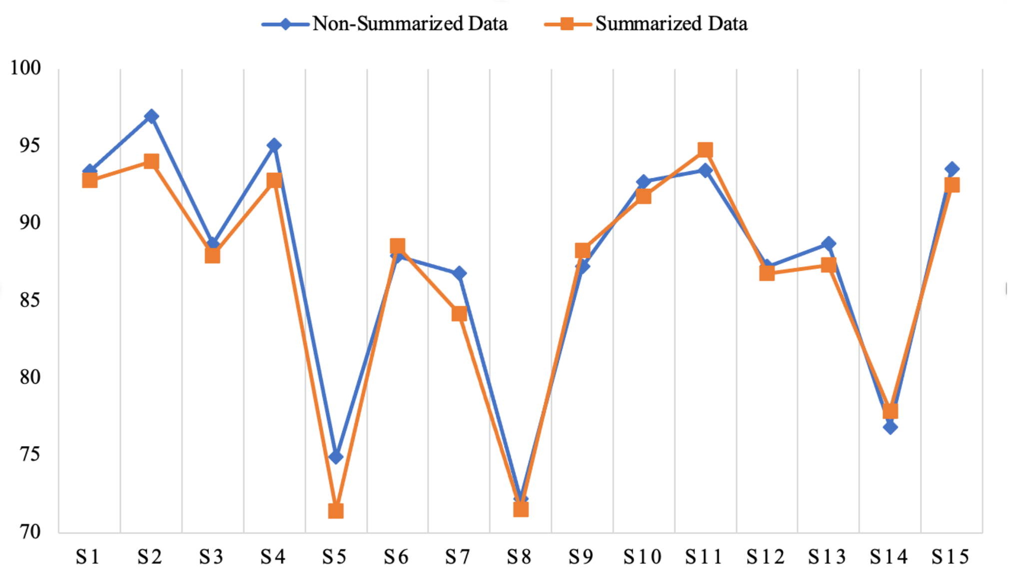



**Fig. 4.5.** Accuracy of Summarized vs. Non-Summarized Data

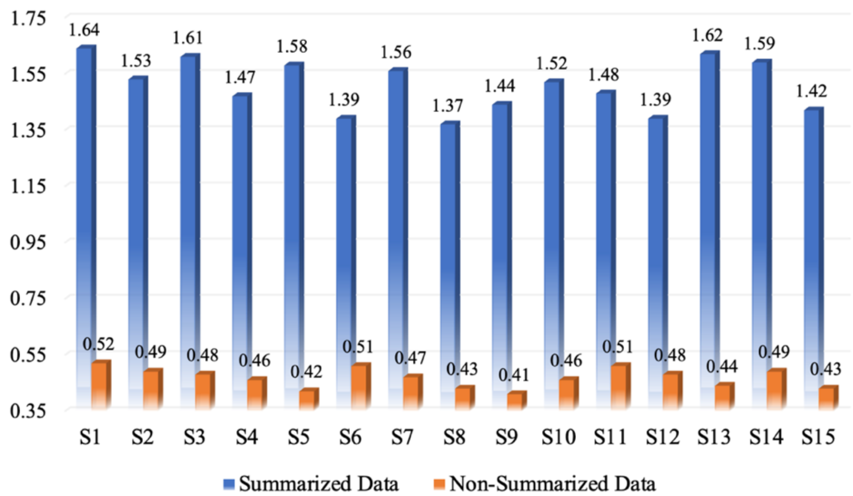



**Fig. 4.6.** Execution Time of Summarized vs Non-Summarized Data

Quite clearly, the decline in accuracy over the summarized data is insignificant as the average model accuracy reduces by 1% as compared to a significant reduction of 27.7% in the execution time. The reduction in accuracy could have been due to reduction in the size of training data. It is plausible to train the model on complete data and then evaluating it on reduced data for better accuracy. Also, as validated from the results obtained, the accuracy of the model depends upon the subject's data, as different individuals have different level of stress under same scenario and the expression of stress is also different for each individual. To understand the affective state of a person it is required to have the health profile of the subjects like medical history, traumas faced along with continuous monitoring using IoMT-based wearable devices. The proposed model currently only takes the subject's bio-signals of 2 hours under controlled lab environment, but this trained model can work as a base model for real-time analysis over the intelligent edge with continuous modifying weights for each individual at the cloud server.

## 4.2. FTL-Emo: Federated Transfer Learning for Emotion Recognition

To ensure privacy protection in emotional identification, a federated-learning based model FTL-Emo was proposed. FTL-Emo: a privacy preserving transfer learning approach for recognition of emotions using EEG. Transfer learning is used to learn from existing data DEAP. In FTL-Emo, CNN has been employed at both client and server side.

### 4.2.1. Methodology

Initially every single user has been considered as a unique client with their own processing power at their edge, where they train their own model, these trained modes' weight is then shared with the centralized server. Twenty-nine input files from K-EmoCon (only EEG files) dataset were provided to proposed model separately at different devices. To have a more effective and accurate model the transfer learning has been used as well. The centralized server was initially trained on DEAP dataset, and these weights were shared with the client models, where they updated it with their own training, this process was continued till the convergence of the model. The experiment was evaluated for Affective dimensions only. The architecture of federated model is shown in figure 4.7. The step-by-step procedure followed is:

- Construct an initial server model employing CNN with publicly available dataset (DEAP).
- Distribute the weights of server model to each client.
- Train client' model with their own data.
- Send Client models' weight to server model.
- Perform weight aggregation at server.
- Distribute the updated weight at server level to each user.
- Repeat this procedure with new coming data, till threshold is reached.

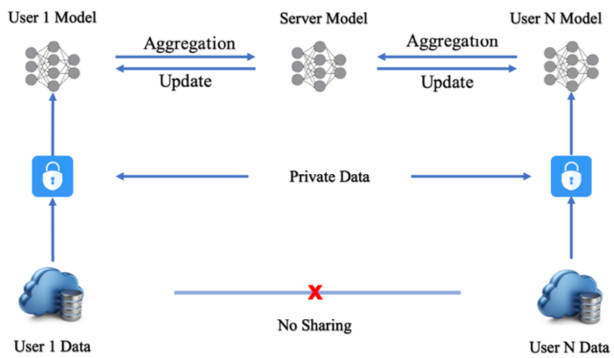



**Fig. 4.7.** Architecture of FTL-Emo

### 4.2.2. Federated Transfer Learning

Federated learning (FL) concept was given by Google in 2017, to reduce the computational costs, by utilizing the computation of the mobile devices, acting them as a node in edge computing [5]. In federated learning, training is performed at individual client level, and then the weights of model from each client are shared with the server, where server collects the weight of each client, compute them, and calculate a new weight, as shown in equation 17.

$$f_s(w) = \frac{1}{K} \sum_{n=1}^{K} (f_n(w)) \tag{17}$$

Where, $f_s(w)$ represents the weight at server/centralized model, and $f_n(w)$ represents the weight of client/user models. This new weight calculated by the server is then communicated back to each client, which again train their own models individually, repeating this process till optimized weights are obtained. To understand the proposed approach, consider that given N different user $U_1$, $U_2$, ...$U_N$ and their sensor collected data (EEG) by $E_1$, $E_2$,..$E_N$. Centralized/server model $M_S$ is first trained with existing dataset (DEAP). The weights of the model, $M_S$ is frozen without performing testing on a deep neural network approach by following validation approach. These frozen weights of server, $W_g$ is then shared with all the clients, where each client then uses this weight as the initialized weight for their own training. The process of initial global weight training is provided in algorithm 2.

---

**Algorithm 2:** Pre-Training Server Model, $M_S$

Input: DEAP Dataset

Output: $W_{pre\text{-}trained}$: Global Model Weight after pre-tuning

1. Train on Convolution Model, Ms
2. Validate
3. Freeze weight of layers
4. Assign Frozen weight to $W_{pre\text{-}trained}$
5. Return : $W_{pre\text{-}trained}$

---

After training the model each client freezes their weight after each epoch, and share this weight with the server model, $M_s$ which calculates the average weight of the model using equation 17, and assign the updated weight to each client, this process is repeated with each epoch, till either the model received global optimization, or the threshold of epoch is reached. The process at local client level is shown in algorithm 3.

---

**Algorithm 3**: Initial Processing on Local Nodes

Input: K-EmoCon, $W_{pre\text{-}trained}$

Output: Wg : Global Server Weight

(1) Share $W_{pre\text{-}trained}$ with each user

(2) For each user i:

  (a) $W_{Ui} \longleftarrow W_{pre\text{-}trained}$

  (b) Train User Model with K-EmoCon

  (c) Update Weight $W_{Ui}$

(d) Share $W_{Ui}$ with Server Model, $M_S$

(e) Aggregate Global weight at Server Model

(f) $f_s(w) = \frac{1}{K} \sum_{n=1}^{K} (f_n(w))$

(3) Update Global Weight, Wg

(4) Share Wg with Each User

(5) $W_{pre\text{-}trained}$ <— Wg

(6) Repeat the Process

---

#### 4.2.3. Dataset

*K-EmoCon*: A multimodal publicly available dataset for emotion recognition in conversations provided by Park et al. in 2020 [77]. The dataset contains audio, video, and bio signals of 32 subjects, who participated in a 10-minute debate task on a social issue in teams of 2, while wearing Physiological signal measuring devices, Emperica E4, NeuroSky, and Polar H7, along with Video cameras for recording the facial expressions and gestures of the participants. Brainwave – EEG (125Hz) and Meditation signals were recorded through NeuroSky.

For our study we have used the data collected by only NeuroSky device, i.e, EEG signal. Although in K-EmoCon each instance has been annotated by 7 people, but we have taken the emotion annotated by the user itself. Park et al. have used 20 different types of emotions for annotation, including dimensional emotional model.

#### 4.2.4. Pre-processing

To create time synchronous data for proposed work, the 29 participants' EEG signals were mapped with output (Arousal, Valance). EEG signals has been collected at a sampling rate of 125 kHz and these has been down sampled to 220 Hz utilizing a Savitzky-Golay filter for smoothing. To extract physiological features, NeuroKit2 Toolbox was used to extract the time domain features of the raw EEG signal with a window size of 4 s (i. e., 4000 steps) and a hop size of 0.5 s (i.e., 500 steps). Then we pad the head and tail of the raw data with neighboring data and combine the above two feature vectors as the physiological feature. In pre-processing, missing values were replaced by means of the next two consecutive instance values.

### 4.2.5.Findings

FTL-EMO has twenty-nine nodes and one centralized server, each executed separately in synchronous manner, combining the weight. The performance evaluation of proposed model for both output categories Arousal and Valance is shown in Table 4.2. For Arousal highest F1-score obtained was 89.03 and the average F1-score obtained on all 29 devices was 86.8. And for Valance the highest F1-score obtained was 94.1, and the average obtained is 88.4. FTL-Emo has obtained high accuracy as well, while maintain the privacy of the data.

**Table 4.2.** Performance of FTL-Emo

| Emotion | Average F-1 Score | Best F-1 Score | Average Accuracy | Best Accuracy |
|---|---|---|---|---|
| **Arousal** | 86.8 | 89.03 | 88.5 | 92.7 |
| **Valance** | 88.4 | 94.1 | 87.3 | 91.5 |
| **Overall** | 87.9 | 93.8 | 88.1 | 91.8 |

The comparison of the model is not possible with exact similar approach, since this is the first work that incorporate privacy preserving approach on K-EmoCon dataset. Although, we have compared the FTL-Emo with simple deep learning-based model to show the performance comparison, and the model has also been compared with other existing models on K-EmoCon, although the data signals taken by them are different, as shown in table 4.3. The proposed work while maintaining the privacy has obtained highly accurate result, the use of transfer learning approach have improved the performance a lot, as can be seen from the results.

**Table 4.3.** Comparison of FTL-Emo

| Model | Average F-1 Score | Average Accuracy |
|---|---|---|
| **Sig-Rep** | 58.9 | 71.3 |
| **LSTM** | 72.5 | 67.4 |
| **RNN** | 76.8 | 71.5 |
| **CNN (DREAM)** | 73.41 | 69.78 |
| **FTL-Emo** | 87.9 | 88.1 |

The ROC curve of two different subjects formed during evaluation of models is shown in figure 4.8. To have accurate emotion recognition, while ensuring privacy of the user in maintaining EHR and sharing it with other doctors or hospitals, FTL-Emo have provided great results. This FTL-Emo approach can enact as baseline for future studies for real time privacy preserving emotion recognition. It can be used for various real time activities like automatic chatbots, HCI in smart industries, schools.

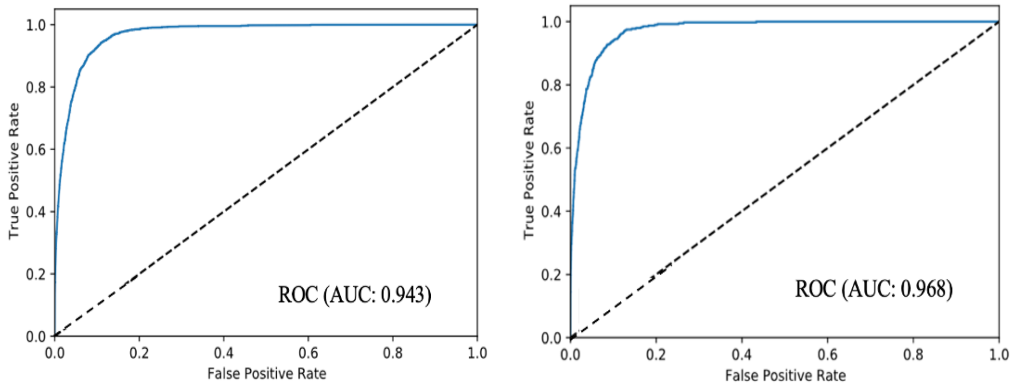



**Fig. 4.8.** ROC curve of U7 & U23

## 4.3. Chapter Summary

This research primarily aimed to identify and major limitation in the existing work, and resolve those limitations by developing novel models. We have identified two major limitations of the existing work, first, handling of large sized data generated during real-time application. Secondly, to ensure privacy preservation while detecting affective state. Two models were developed to resolve both these issues: Genetically optimized fuzzy c-means clustering for data reduction, and FTL-Emo for privacy preservation using federated transfer learning approach.

***

# CHAPTER 5

# CAUSAL THEORY OF AFFECTIVE EXPERIENCE

This work primarily aimed to understand the relation between cause and effect, by analyzing the impact of a stimuli on human affective state. To understand the relation between the two, a affective causal relation theory was proposed. To support the proposed theory, the analysis was made on a case study of student's data gathered through questionaries during the pandemic.

## 5.1. Causal Theory

Affect signals us that things are going all right (e.g., because we are in a good mood or are experiencing joy or serenity) or that things are not going so well (we are in a bad mood, anxious, upset, or angry). It is the visible reaction a person displays toward events. Typically, events cause emotional reactions, which can influence attitudes and behaviours. Affective events are events that impact people in positive or negative ways. This characterizes the affective event theory (AET), where an emotional episode consists of emotional experiences from an event that affect the mood and emotions cycles.

It is imperative to comprehend the cause-effect relationship between affective state, its causes (triggers) and effects as every individual acts differently when encountered with the same situation, and the affective state is not dependent on a single attribute. That is, it works differently for different individuals. AET has been well-studied for organizational (work/job) environment where the impact of affective events and the reactions are analyzed for employees' psychological well-being and work productivity. We extend this theory to generalize a *causal theory of affective experience* (CTAE) which comprehends the relationship between affective state, its causes and effects. As specific events are antecedents of affective reactions and behaviours, it is during the emotion process an individual automatically registers an eliciting stimulus and experiences a feeling state and physiological changes [103].

CTAE defines the relation between the cause (dynamic event/stressor/trigger) which alters routine and brings change, an observable affective reaction characterized using traits and states and a behavioural effect as a response. It helps to typify the affect reasoning where events trigger affective reactions based on exact or perceived cause and lead to an affect-driven behaviour. Consequently, the affect-driven behaviour describes the psychological well-being of an individual.

Table 20 depicts few generic examples that help us understand the CTAE, where behavioural response and affective experience for each causal event is given.

**Table 5.1.** CTAE Examples

| **Event** | **Affective Experience** | **Behavioural Response** |
|---|---|---|
| **Partner Infidelity** | we are angry (negative affectivity) | we may attack |
| **Career Advancement** | we are happy (positive affectivity) | we may socialize with friends & family |
| **Sighting A Snake** | we are fearful (negative affectivity) | we may run away and isolate |
| **Heartbreak** | we are sad (negative affectivity) | we may stop eating |
| **Surprise Birthday Party** | we are amused (positive affectivity) | we may dance |

Fig.5.1. depicts the CTAE in triggered events/situations.

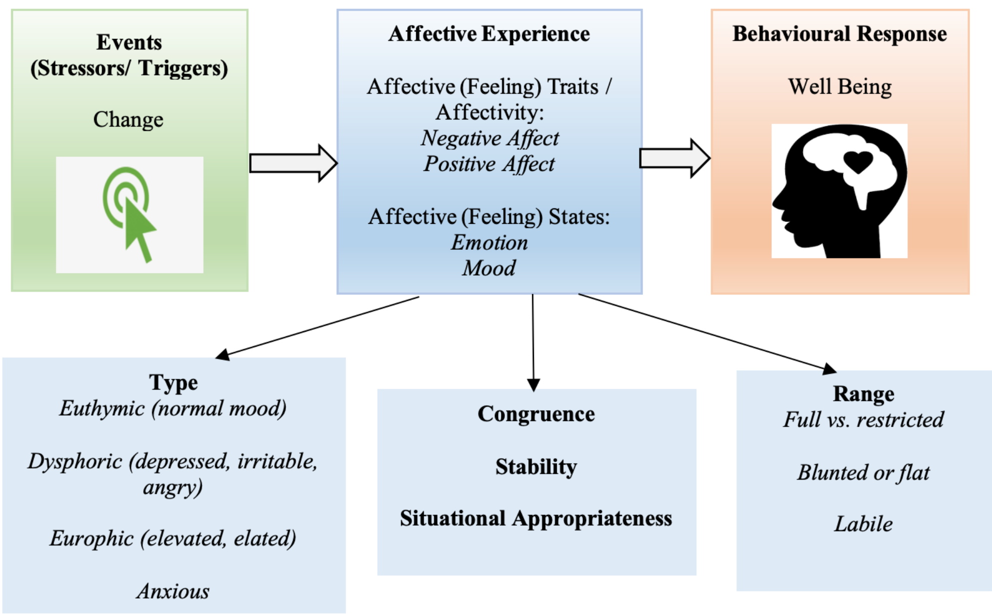



**Fig.5.1.** The Causal theory of affective experience in triggered situations

To study casual theory of emotional processing which comprehends relationship between affective state, its causes and effects with the help of a case study on Indian students during the COVID-19 pandemic.

## 5.2. Case Study

The current pandemic has fostered the academic community to adopt and adapt to an absolute online teaching paradigm at a global level. From pre-schools to post-graduate professional courses, e-learning is now being used to sustain academics and prevent collapsing of educational systems. The primary issue with online method of teaching (online pedagogy) is that the masses have been forced to use it. It was unplanned and rapid move with no training, insufficient resources (devices, bandwidth) and little preparation. This transition from classroom-based to online learning can be quite stressful for children as it is a change from their normal structure, and they are not accustomed to the new way of learning. Undeniably, the students are experiencing psychological and behavioral issues with this new sudden arrangement which is "emotionally and mentally draining".

The psychological effects can lead to a range of reactions from fear to anxiety, from boredom to depression, from frustration to anger, from stigmatization to repulsion. The detection of psychological effects may take months or years as early signs or symptoms of mental disorder may be overlooked or indiscernible until it's too late.

### 5.2.1. Methodology

This study is conducted using a *'phenomenological approach'* which involves examining human experience using data collected (through interviews, questionnaires and focus groups) from people who have experienced the phenomenon and through observations as people are experiencing a phenomenon. Formally, phenomenological study is to trace out precisely the lived experiences of people & generate theories or models of phenomena being studied. It provides a rich description of the 'lived experience'. This qualitative method focuses on describing participants' experience, i.e., depicting to the underlying or core experience, describing participants' perspective of their experience, interpreting participants' experience and interpreting participants' understanding of their experience. Quite clearly, phenomenological studies help in raising an understanding of the relationship between states of individual consciousness & social life and attempts to uncover how human awareness is implicated in the production of a social action, social situation & social world. The forced migration from in-person to online learning during this current pandemic has deepened the digital divide and added to the stress levels of the students. Fig. 5.2 depicts the common COVID-19 triggered mental health issues in students.

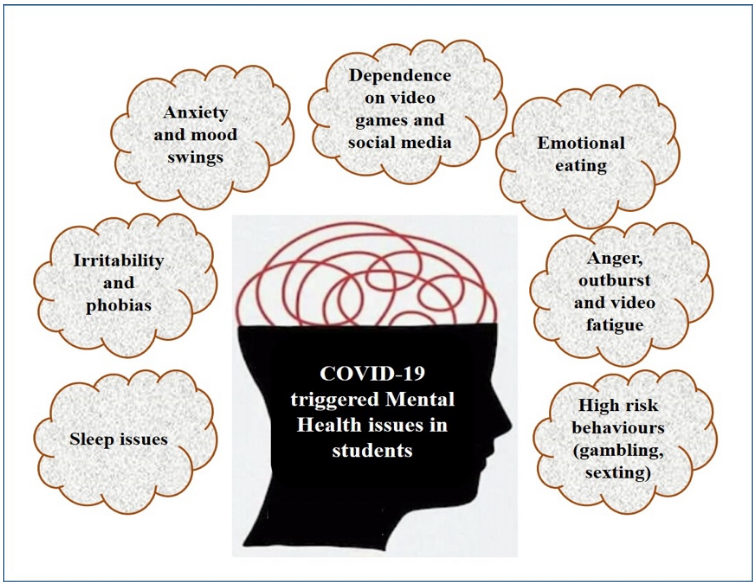



**Fig. 5.2.** COVID-19 triggered common mental health issues in students

### 5.2.2. Participants and C-19 MHQ

A total of 733 students completed the C-19 MHQ, but 1 student with history of clinical depression prior to the pandemic was not taken into account. Therefore, total responses of 732 students from both undergraduate and postgraduate courses were considered for analyzing the causal effect of pandemic and online learning on affective state of an individual. Though stress and pressure are normal among university students, especially during the examinations or placement sessions, but the pandemic and nationwide lockdown has induced increased levels of stress. The questionnaire assessed the basic personality characteristics of the subjects like age, sex, resources available, interests, health condition, and the course they are pursuing. Participants were also asked about a variety of questions related to remote learning, online exams, and other experiences they are having due to lockdown such as the different kind of support they are receiving from friends, family and the university. Subjects indicated the different issues they are facing, such as restlessness, excessive worrying, irrational phobias, altered sleep pattern, video fatigue, changes in eating habits (emotional eating), difficulty in concentration, and feeling of guilt amongst others that triggered stress. Finally, participants were asked to rate their level of stress too on the scale of 1 to 5.

The age of students varied from 17 to 26, with mean age of 20.01 years and the gender distribution included 554 male participants and 178 female participants. The under-graduate study year participation was as follows: only 20 participants were from first year of the undergraduate studies whereas the largest share of participants were from second year, i.e., 402 students (200 alone the student of Bachelor of Technology (B.Tech.) in Computer Science and Engineering or Information Technology). 214 students participated in the study were from third year and 96 students were from the final year. Most of the students involved in the study (427) were from the Department of Computer Science or Information Technology, followed by 183 students pursuing engineering in Computer Science with specialization in a sub-domain (Big Data Analytics, Machine Learning etc.), and 77 belonged to other core engineering branches such as Mechanical Engineering, Civil Engineering, Electronics Engineering, and Electrical Engineering. Post-graduate students who took part in the study were 38 whereas only 7 research scholars pursuing doctorate participated in the survey. These basic characteristics of participants are shown in table 5.2.

**Table 5.2**. Characteristics of the Participants

| CHARACTERISTICS | FREQUENCY | PERCENTAGE |
|---|---|---|
| Gender | | |
| Male | 554 | 76% |
| Female | 188 | 24% |
| Age | | |
| 17-20 | 342 | 46.5% |
| 20-22 | 368 | 50.3% |
| >22 | 22 | 03.1% |
| Course Enrolled in | | |
| B.Tech CSE/IT | 427 | 58.3% |
| B.Tech CSE with Specialization | 183 | 25% |
| B.Tech other than CSE | 77 | 10.5% |
| M.Tech | 38 | 5.2% |
| PhD | 7 | 0.9% |

| | | |
|---|---|---|
| Year of the Study | | |
| $1^{st}$ | 20 | 2.7% |
| $2^{nd}$ | 402 | 54.9% |
| $3^{rd}$ | 214 | 29.2% |
| $4^{th}$ | 96 | 13.1% |

In India, engineering undergraduate students get placement opportunities by organizations/ companies visiting their campus for recruitment. Job losses and economic malaise as an impact of COVID-19 has instigated fear of the future and upheaval in career prospects in students leading to increased stress situations. Job prospects is one of the major factors contributing to the stress and depression in students, as most of the students are from second and third year, who don't have an active job offer with them. Out of the total 732 participants, 560 students had no job offer and 100 had received job offer from one or more than one organization whereas 72 students opted for not applicable option as research scholars and first year students does not sit in any of the recruitment activity. 492 students were worried about their offered job/internship opportunities as COVID 19 might affect their existing offers or they might not receive as many offers as they would have in the pre-pandemic situation. Only 68 students were confident about their job and internship offers, for rest of the 172 students were uncertain about the future in terms of job and internship opportunities. Further, the participants who had job or internship offer were still under pressure as 120 had delays in joining date whereas 184 students had the same joining date for their jobs and internship (428 students doesn't fall into this category so they responded with not applicable option). Table 22 presents this analysis of job prospects.

**Table 5.3.** Analysis of job offer to students

| CHARACTERISTICS | FREQUENCY | PERCENTAGE |
|---|---|---|
| Job Offered received | | |
| Yes | 100 | 14% |
| No | 560 | 77% |
| NA | 72 | 10% |
| Stress that Pandemic will affect the Job/Internship | | |
| Opportunities | 492 | 67% |
| Yes | 68 | 9% |
| No | 172 | 23% |
| Maybe | | |

| | | |
|---|---|---|
| Joining Delayed? | | |
| Yes | 120 | 16% |
| No | 184 | 25% |
| NA | 428 | 58% |

The third main focus of C-19 MHQ was about the adoption and adaption of remote learning, that is, whether the students were able to attend online lectures, for which 430 students responded with yes, whereas 238 students attended the lectures sometimes, depending upon the network availability and 34 students were unable to attend any of the lectures. Different issues that students faced while attending online lectures were internet connectivity issue (518 students), unavailability of laptop/ smart phones (160), inability to focus during online lecture (350), background disturbance either from faculty side or on your own side (240), not enough practice questions in online questions (208), 64 students faced other issues than the mentioned ones. Only 64 students were comfortable with the online classes and did not face any issue. Students were asked about whether their doubts were cleared during the online classes and 190 students responded with a positive Yes, 464 students had their doubts cleared sometimes, whereas 78 students answered with not at all option. Overall, 84 students preferred online classes over traditional classroom teaching, whereas 526 students still prefer traditional classroom teaching, 122 students are inconclusive about the decision. Table 23 summarizes the response of participants regarding remote learning.

**Table 5.4.** Analysis of the Remote Classes

| CHARACTERISTICS | FREQUENCY | PERCENTAGE |
|---|---|---|
| Able to attend online classes? | | |
| Yes | 430 | 59% |
| No | 34 | 37% |
| Sometimes | 268 | 5% |
| Different issues being faced during online classes? | | |
| Internet Connectivity issue | 518 | 70.7% |
| Unavailability of Laptop or smart phone | 160 | 21.8% |
| Difficult of focus | 350 | 47.8% |
| Background disturbance | 240 | 32.78% |
| Not enough practice questions | 208 | 28.4% |
| Other issues | 64 | 8.7% |
| No issues | 60 | 8.1% |
| Doubts cleared during Online classes? | | |
| Yes always | 190 | 26% |
| Sometimes | 464 | 63% |
| Not at all | 68 | 11% |

| | | |
|---|---|---|
| <u>Prefer Online class over class room teaching</u><br>Yes<br>No<br>Maybe | <br>84<br>526<br>122 | <br>11%<br>72%<br>17% |
| Rating for the initiative of the university for starting Online classes | 3.46 | - |

The final section of the C-19 MHQ analyzed the effect of online examinations during pandemic. To understand the students' pre-condition, their comfort level with online exams is a must and so students were asked about their exposure to online assessments, that is whether they have given an exam in online mode earlier too. 664 students responded with yes, whereas there were 68 students who did not have any experience in giving an online exam. Students were enquired about whether they believed that the evaluation process in the online mode was same as the traditional exams and 159 students were in favour of this, whereas 357 students responded with a No, implying they would not be assessed akin to the traditional manner. 216 students were inconclusive, so they answered with Maybe. Table 5.5 outlines the response of participants towards online exam.

**Table 5.5.** Analysis of the Online Exam

| CHARACTERISTICS | FREQUENCY | PERCENTAGE |
|---|---|---|
| <u>Prior Experience in giving Online Exam</u><br>Yes<br>No | <br>664<br>68 | <br>91%<br>9% |
| <u>Evaluation process will be same in both online and traditional exam?</u><br>Yes<br>No<br>Maybe | <br><br>160<br>356<br>216 | <br><br>22%<br>49%<br>30% |
| <u>Preference of Online Exam over Traditional exam</u><br>Online Exam<br>Offline Exam<br>Doesn't Make a difference | <br>288<br>284<br>160 | <br>39%<br>39%<br>22% |
| <u>Considers Grades will not be same as if exam would be in offline mode</u><br>Yes<br>No<br>Maybe | <br><br>420<br>120<br>192 | <br><br>57%<br>16%<br>26% |
| Stress induced due to online exams | 3.58 | 71.6% |

Based on the answers of the self-assessment questionnaire the students stress state was categorized into six different classes, namely, amused, neutral, low stress, high stress, depression and anxiety. It was observed that various factors that caused stress included online exams, job and internships, isolation, competition among friends, physical health issues and unavailability of the resources among others. It was perceived from responses of the participants that the major indication of stress is uncertainty of future, and the economic crisis, as 67% of the participants have responded with having access to MOOC platform as a stress reducing strategy.

### 5.2.3. Methodology

To predict the mental stress level of university students in the pandemic situation, two different machine learning algorithms namely random forest (RF) and artificial neural network (ANN) were trained and tested on the 732 responses collected through the participant data of C-19 MHQ. To implement the ML models, data was pre-processed first by converting the responses of C-19 MHQ to numeric values, by replacing Yes with 1, No with 0, and Maybe, NA, and sometimes to 2 in all the features. Similarly, B.Tech in CSE/IT is replaced with 1, Specialization with 2, B.Tech in other core branches by 3, and M.Tech and PhD with 4. In the similar fashion, all the features present in the data is converted into the numeric values, the final mental health issues defined by the users were further categorized into six different classes, amused, neutral, low stress, high stress, depression and anxiety. The target class assignment was based on the information provided by the participants over the different issues they were facing, and finally the same was verified with the help of a practicing clinical psychiatrist. After, the validation of the target class, the final data contained 97 students in amused category, 126 with neutral mental state, 211 had low stress issue, 192 students were under high stress, and 74 students were under depression, whereas 32 students had severe anxiety. After data pre-processing, feature selection was performed by generating correlation matrix. It analyzed the features that have no effect or minimal effect on the output and removed them and the simultaneously gave more initial weightage to features showing high correlation. Fig. 5.3 shows the generic architecture of ML-based predictive model for affective mental states.

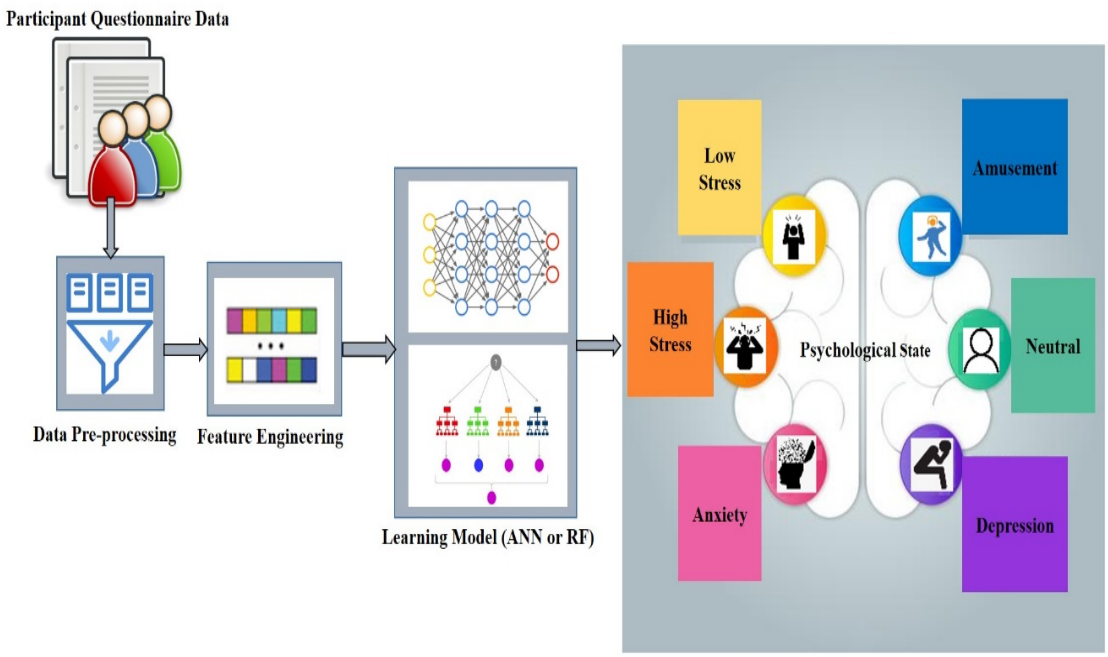



**Fig. 5.3**. Affective Mental State Prediction Model

To train and predict the model, 4-cross validation technique was used over two different models, firstly the data was applied on an ensemble method, i.e. random forest (RF), in which 7 mini decision trees were used to predict the mental state of the user. RF allows collective decisions of different decision trees that is in RF we make a prediction about the class, not simply based on one decision trees, but by an (almost) unanimous prediction, made by 'K' decision trees. Prediction in RF is truly ensemble as for each decision tree we predict the class of instance and then return the class which was predicted the most often (Fig. 5.4).

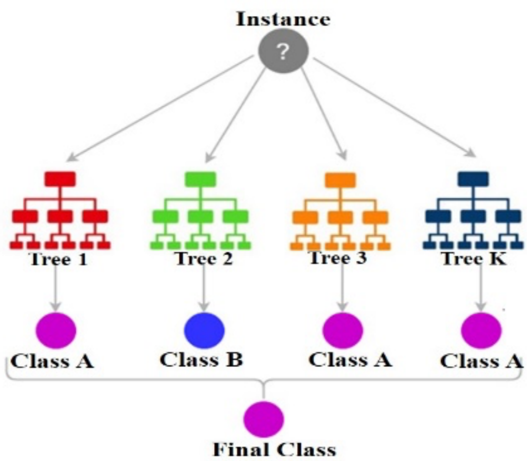



**Fig. 5.4.** Random Forest with ‘K’ decision trees

The second model that was used to predict the mental state of the user was an artificial neural network (ANN). To provide the feature values as an input to the neural network, data is normalized in the range of 0-1, as the decision tree works well on the actual values, the normalized data is only used for the neural network. Model was trained and rather than initializing the each layer with equal weights, job offer, availability of the resources, health issues, and the support of families were initialized with more weight (0.1 for each), and rest of the features were initialized with 0.02 weight. 150 epochs were used to train the neural network with back propagation with learning rate of 0.01 optimized using stochastic gradient descent. Back propagation in an artificial neural network (ANN) is a method of training a network with hidden neurons (i.e. network with multiple hidden layers) . In this method, using training data where input and output is known, the difference or error between desired output and actual output is computed and propagated back into the hidden layers of the neural network to adjust the node weights so as to bring the difference between desired and actual output down.

### 5.2.4. Findings

From case study of 732 students, we can see the multiple factors that are spiking the stress levels in different students, some of the factors include the unavailability of the resources to study online.

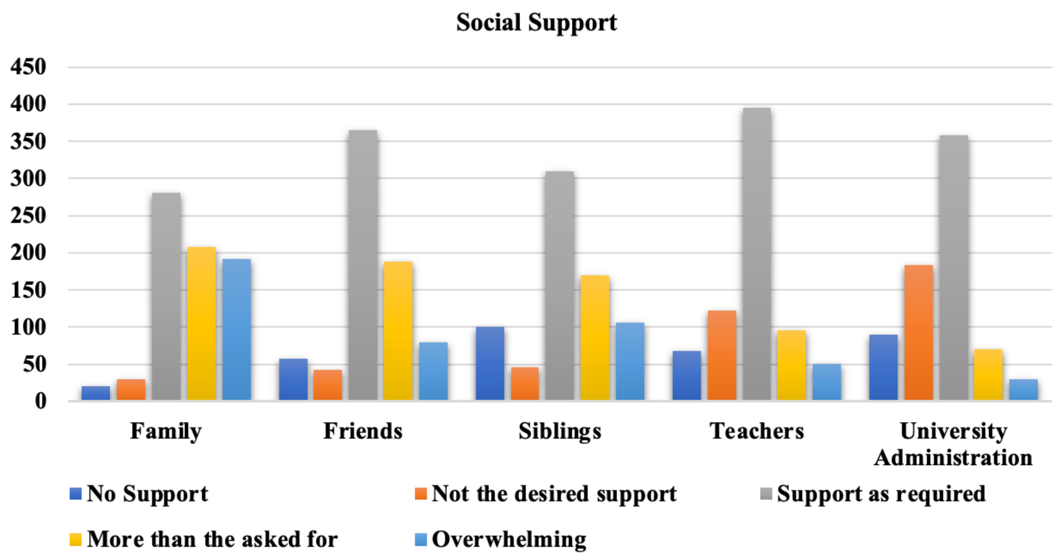



**Fig. 5.5.** Support Received during Pandemic from a social group

The stress to get the job offer, to repay the loan, feeling isolated, not having the support of friends and family, missing the college and other day to day activities, the insecurity about own learning growth in comparison to friends and classmates, and along with that the stress of giving the online examination. Some of these attributes are shown in the following fig. 5.5 & fig. 5.6.

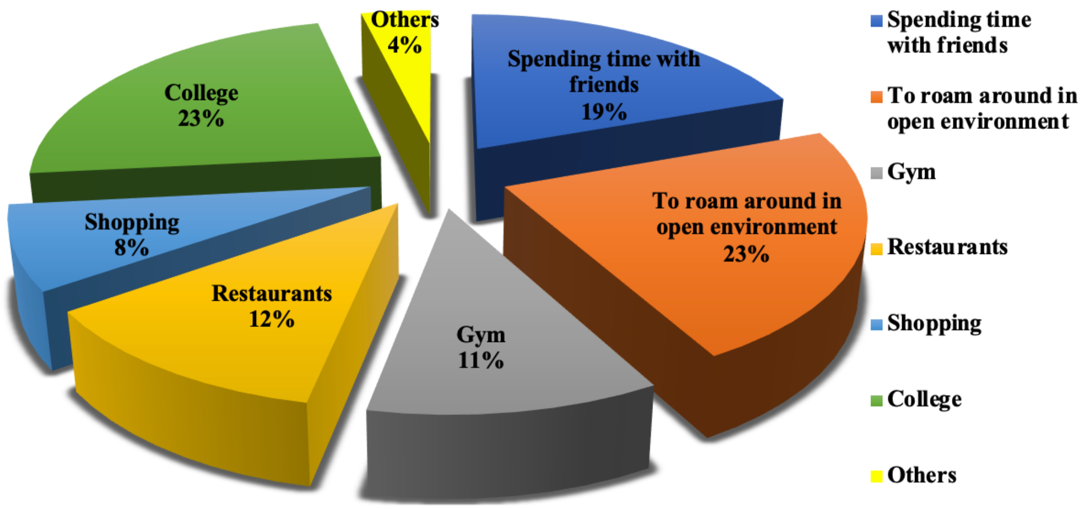



**Fig. 5.6.** Activities missed most during the COVID-19

As some of the companies have already declared the financial year 2020-2021 as the pandemic-induced recession, students have a lot of pressure on them to learn new technologies to be market-ready on completion of their graduation. When asked what technologies they will like to learn, most of the students wanted to learn at least two new technologies, as shown in figure 5.7.

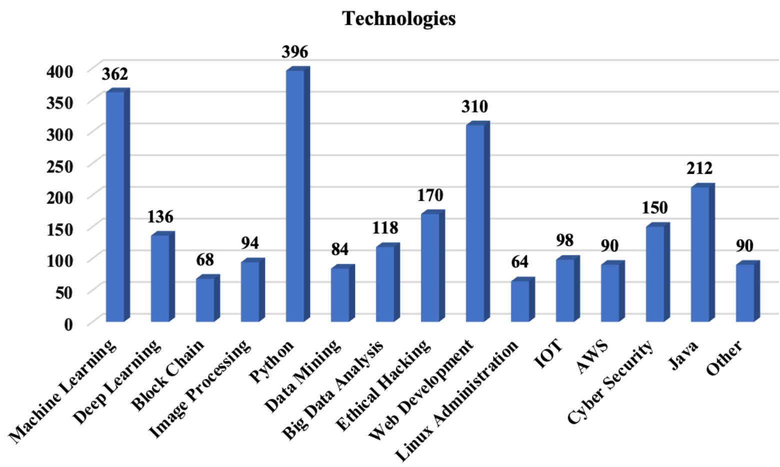



**Fig. 5.7.** Technologies student want to learn during nationwide lockdown

The nationwide lockdown has provided students with the unique opportunity to stay with their families, as most students belong to different states or districts, this was the first and unique opportunity for participants to spend time with their families, but with the ongoing financial crisis, their focus was more on to be industry-ready, rather than to enjoy this time. The effect of availability of Job/internship offer on students was shown in table 5.6.

**Table 5.6.** % distribution of participants' affective psychological state

| Affective Mental State | Participants | Percentage |
|---|---|---|
| Neutral | 126 | 17.21% |
| Low Stress | 211 | 28.82% |
| High Stress | 192 | 26.22% |
| Depression | 74 | 10.11% |
| Anxiety | 32 | 4.37% |
| Amusement | 97 | 13.25% |

The issues that students were facing are illustrated with the help of the pie chart, as shown in fig. 5.8.

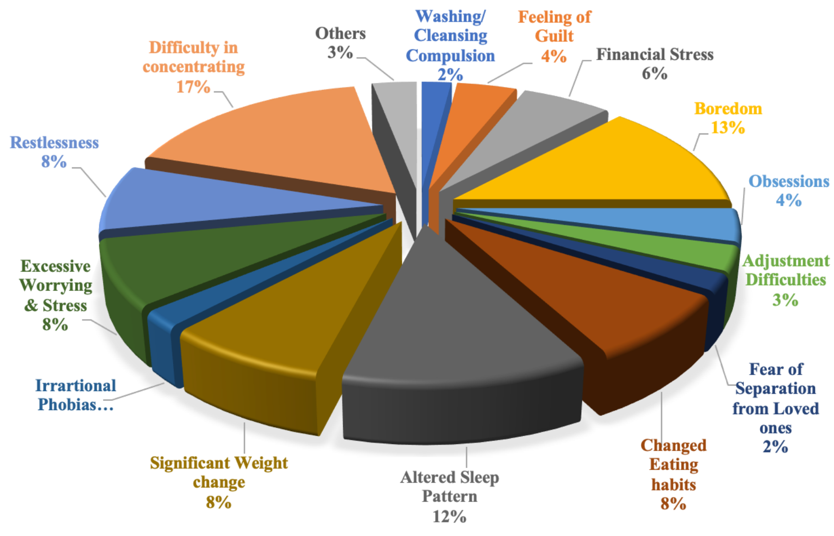



**Fig. 5.8.** Issues faced by the participants

For, the predictive model, RF and ANN were trained using features for identifying the affective psychological state of the participants. The models were trained using 75% data and tested with 25% of the data, 4-cross validation was performed. An accuracy of 90.4% was obtained for RF and 89.15% for ANN. The results are shown in table 5.7.

**Table 5.7.** Performance results of the predictive models

| Model | Precision | Recall | Accuracy | F-Measure |
|---|---|---|---|---|
| Neural Network | 88.0 | 87.4 | 89.15 | 87.48 |
| Random Forest | 88.7 | 89.9 | 90.4 | 89.29 |

The confusion matrix generated by both the models is shown below in fig. 5.9 and fig. 5.10 for ANN and RF respectively. It indicates that RF has an overall better performance in comparison to ANN.

| N=732 | Amused | Neutral | Low Stress | High Stress | Depression | Anxiety |
|---|---|---|---|---|---|---|
| Amused | 78 | 6 | 5 | 4 | 2 | 2 |
| Neutral | 9 | 111 | 3 | 2 | 1 | 0 |
| Low-Stress | 1 | 2 | 187 | 19 | 2 | 0 |
| High-Stress | 0 | 0 | 5 | 183 | 4 | 0 |
| Depression | 0 | 0 | 1 | 4 | 66 | 3 |
| Anxiety | 0 | 0 | 1 | 1 | 3 | 27 |

**Fig. 5.9.** Confusion matrix of artificial neural network

| N=732 | Amused | Neutral | Low Stress | High Stress | Depression | Anxiety |
|---|---|---|---|---|---|---|
| Amused | 77 | 11 | 4 | 3 | 1 | 0 |
| Neutral | 7 | 115 | 3 | 1 | 0 | 0 |
| Low-Stress | 3 | 0 | 188 | 15 | 2 | 1 |
| High-Stress | 0 | 2 | 3 | 181 | 5 | 1 |
| Depression | 0 | 0 | 1 | 2 | 69 | 2 |
| Anxiety | 0 | 0 | 0 | 1 | 2 | 29 |

**Fig. 5.10**. Confusion matrix of random forest

Evidently, the long-term mental health effects of pandemic and consequent remote learning affect the well-being of students. Conclusively, this study emphasizes that a rapid assessment of outbreak-associated psychological disorders among various sections of society is needed as the pandemic may lead to severe public mental health implications. As stress, depression, anxiety for a prolonged period of time can leave a permanent affective change on the psychological state of an individual it is imperative for public health and policy makers along with the university administration to take extra care of the students, by providing them with all the kind of the support that can keep them motivated during this crucial period. It is vital to identify students at-risk for emotional difficulties and develop a plan of action to connect for effective support. The educational centres must adopt a multi-tier system which:

- creates mental health first aid services in centres and community with counsellors and psychologists.
- assists in identifying at-risk students through AI-based predictive learning.
- advocates for universal screening of the school population during and following online learning phases; and
- increases awareness of the importance of mental health screening with their teacher colleagues as well as school administrators and parents with the help of webinars, talks, close group discussions and virtual one-on-ones.

## 5.3. Chapter Summary

This chapter presents a theory for understanding the causal relationship between the events and their consequences on the affective state of a person. To validate this theory, this chapter contains a case study on stress induced in Indian students during COVID-19.

***

# CHAPTER 6

# CONCLUSION & FUTURE WORK

In the digitalized and interconnected world, with advancements in sensor technologies, the traceability of every single activity of any individual is possible. IoT sensors can track the activities happening inside the human body as well. Stress is common in day to day lives, but overwhelming stress can lead to many psychological and emotional disorders. These psychological and emotional disorders can be difficult to detect, until it is too late. The early identification and management of any disorder can save the lives. To identify any psychological or emotional disorder, it is important to understand the cognitive thinking of a person, as every person reacts to every situation differently, the impact of each situation on a person's mind depend upon their past experiences, their cognitive process, and their personality. To analyze the impact of any stimuli on a person, we proposed to used multi-modalities for affective state recognition. Different biomarkers measure different aspects of human body, physiological biomarkers measure the changes happening inside the body like heart rate, blood pressure, temperature etc., whereas changes in hormones are measured by the electrochemical biomarkers, and the visible changes that are observable by human eye, are measured through visual biomarkers.

This research incorporated different models developed and designed to identify the impact of different biomarkers in identifying the psychological state, and emotional state. Fusion of different biomarkers have also been explored for emotion detection models. For affective psychological state, two models has been proposed, one for physiological biomarker, by developing a hierarchal deep learning based model on WESAD dataset. The second model for psychological state, is developed for psycholinguistic biomarkers, by detecting anxious depression from social media posts. The study was conducted on 100 users of Twitter that are prone to stress. For affective emotional state recognition, three models have been designed using transfer learning approach with combination of different modalities; first with audio, video, and linguistic modalities. Second with Video and physiological modalities, and lastly the third model, taken into account video, audio and physiological modalities. Designed models has shown that, incorporation of different modalities enhance the performance of the emotion recognizing model. The resolution of issues faced with real-time deployment of these models were also proposed in this research work.

IoMT sensors are readily available in the form of fitness watches and other small wearable devices nowadays. These devices generate a large amount of volume every day. IOMT based devices have made it possible to monitor the wellness of a person remotely. To handle the volumes of data produced every second the IoMT devices, a genetically optimized fuzzy c-means clustering technique is proposed that is used to summarize the data. This reduction in data using summarization helps in handling data overload issue and time latency in transferring the data from edge to the servers. The proposed technique has a unique approach of clustering the data of a concise time frame and then converting each cluster to a single entity by performing the summarization of the cluster (identifying the centroid point). The proposed method has reduced the size of the data while maintaining the quality. Although the results obtained have shown the decline of the accuracy by 1%, which can be lethal in healthcare, this work can act as the baseline model for handling data overload and time latency for IOMT based data. An improvement of the model can make remote wellness tracking a reality for every person.

Another major issues faced by smart psychological healthcare is the privacy of the user. As every subject's data is shared to a centralized server, the threat of cyber-attacks is real. To resolve this issue, a federated learning based approach is proposed, that employs transfer learning in a decentralized environment. Initially every single user has been considered as a unique client with their own processing power at their edge, where they train their own Model, and these trained modes' weight is then shared with the centralized server.

To define the relation between an event and the affective psychological state, a causal theory of affective experience is proposed. CTAE defines the relation between the cause (dynamic event/stressor/trigger) which alters routine and brings change, an observable affective reaction characterized using traits and states and a behavioural effect as a response. This theory is validated through a phenomenological study conducted through questionnaire on Indian students during COVID-19. Study was conducted on 732 students, empirical analysis of the questionnaire, have shown that due to COVID-19 restrictions, the stress levels in students has emerged a lot.

## 6.1. Future Scope

One of the challenges of automatic affective state detection is the availability of data, the models developed are on the very small datasets, that contains the biomarkers of young adults, while performing activities in a controlled lab environment. To develop effective automatic affective recognition models, user specific data is required. Researchers should work on developing a large user specific datasets, recorded while performing day to day activities. The past medical history of the users should also be considered.

Additionally, this research has proposed a method to resolve the issue of data overload. The proposed model has reduced the size of the data, but it also impacted the accuracy of the model as well, the proposed genetically optimized fuzzy c-mean clustering could serve as the base model, for future researchers. To have real-time automatic psychological and emotional detection, novel data reduction techniques need to be developed.

***

# REFERENCES


[1] Bienertova-Vasku, Julie, Peter Lenart, and Martin Scheringer. "Eustress and Distress: Neither Good Nor Bad, but Rather the Same?." *BioEssays* (2020): 1900238.

[2] S. Kumar, 'Biochemistry of the Brain', in Examining Biological Foundations of Human Behavior, 2020.

[3] World Health Organization. Depression and other common mental disorders: global health estimates. World Health Organization; 2017.

[4] Esposito, Anna, Antonietta M. Esposito, and Carl Vogel. "Needs and challenges in human computer interaction for processing social emotional information." *Pattern Recognition Letters* 66 (2015): 41-51.

[5] Dautov, Rustem, Salvatore Distefano, and Rajkumaar Buyya. "Hierarchical data fusion for Smart Healthcare." *Journal of Big Data* 6, no. 1 (2019): 19.

[6] Kumar, A., Sharma, A., & Arora, A. (2019, March). Anxious depression prediction in real-time social data. In International Conference on Advances in Engineering Science Management & Technology (ICAESMT)-2019, Uttaranchal University, Dehradun, India.

[7] Tong, Lei, Qianni Zhang, Abdul Sadka, Ling Li, and Huiyu Zhou. "Inverse boosting pruning trees for depression detection on Twitter." *arXiv preprint arXiv:1906.00398* (2019).

[8] Garcia-Ceja, Enrique, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J. Oedegaard, and Jim Tørresen. "Mental health monitoring with multimodal sensing and machine learning: A survey." *Pervasive and Mobile Computing* 51 (2018): 1-26.

[9] Chandrasekaran, M., M. Muralidhar, C. Murali Krishna, and U. S. Dixit. "Application of soft computing techniques in machining performance prediction and optimization: a literature review." *The International Journal of Advanced Manufacturing Technology* 46, no. 5-8 (2010): 445-464.

[10] Crestani, Fabio, and Gabriella Pasi, eds. *Soft Computing in Information Retrieval: techniques and applications*. Vol. 50. Physica, 2013.

[11] Priya, Anu, Shruti Garg, and Neha Prerna Tigga. "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms." *Procedia Computer Science* 167 (2020): 1258-1267.

[12] Halim, Zahid, and Mahma Rehan. "On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning." *Information Fusion* 53 (2020): 66-79.

[13] Ihmig, Frank R., et al. "On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals." *Plos one* 15.6 (2020): e0231517.

[14] Penchina, Brian, et al. "Evaluating Deep Learning EEG-Based Anxiety Classification in Adolescents with Autism for Breathing Entrainment BCI." (2020).

[15] Elgendi, Mohamed, and Carlo Menon. "Machine Learning Ranks ECG as an Optimal Wearable Biosignal for Assessing Driving Stress." *IEEE Access* 8 (2020): 34362-34374.

[16] Kumar, Prince, Shruti Garg, and Ashwani Garg. "Assessment of Anxiety, Depression and Stress using Machine Learning Models." *Procedia Computer Science* 171 (2020): 1989-1998.

[17] Flesia, Luca, et al. "Predicting Perceived Stress Related to the Covid-19 Outbreak through Stable Psychological Traits and Machine Learning Models." *Journal of clinical medicine* 9.10 (2020): 3350.

[18] Richter, Thalia, Barak Fishbain, Andrey Markus, Gal Richter-Levin, and Hadas Okon-Singer. "Using machine learning-based analysis for behavioral differentiation between anxiety and depression." *Scientific reports* 10, no. 1 (2020): 1-12.

[19] Can, Yekta, Said, et al. "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study." *Sensors* 19.8 (2019): 1849.

[20] Giannakakis, Giorgos, Eleftherios Trivizakis, Manolis Tsiknakis, and Kostas Marias. "A novel multi-kernel 1D convolutional neural network for stress recognition from ECG." In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 1-4. IEEE, 2019.

[21] Islam, Md Rafiqul, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, and Anwaar Ulhaq. "Depression detection from social network data using machine learning techniques." *Health information science and systems*6, no. 1 (2018): 8.

[22] Giannakakis, G., Matthew Pediaditis, Dimitris Manousos, Eleni Kazantzaki, Franco Chiarugi, Panagiotis G. Simos, Kostas Marias, and Manolis Tsiknakis. "Stress and anxiety detection using facial cues from videos." *Biomedical Signal Processing and Control* 31 (2017): 89-101.

[23] Subhani, Ahmad Rauf, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, and Aamir Saeed Malik. "Machine learning framework for the detection of mental stress at multiple levels." *IEEE Access* 5 (2017): 13545-13556.

[24] Sau, A., Bhakta, I. (2017)"Predicting anxiety and depression in elderly patients using machine learning technology."*Healthcare Technology Letters* **4 (6)**: 238-43.

[25] Gjoreski, Martin, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. "Continuous stress detection using a wrist device: in laboratory and real life." In *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*, pp. 1185-1193. 2016.

[26] Picard, Rosalind W., Elias Vyzas, and Jennifer Healey. "Toward machine emotional intelligence: Analysis of affective physiological state." *IEEE transactions on pattern analysis and machine intelligence* 23, no. 10 (2001): 1175-1191.

[27] J Healey, Jennifer A., and Rosalind W. Picard. "Detecting stress during real-world driving tasks using physiological sensors." *IEEE Transactions on intelligent transportation systems* 6, no. 2 (2005): 156-166.

[28] Koelstra, Sander, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. "Deap: A database for emotion analysis; using physiological signals." *IEEE transactions on affective computing* 3, no. 1 (2011): 18-31.

[29] Maxhuni, Alban, Pablo Hernandez-Leal, L. Enrique Sucar, Venet Osmani, Eduardo F. Morales, and Oscar Mayora. "Stress modelling and prediction in presence of scarce data." *Journal of biomedical informatics* 63 (2016): 344-356.

[30] Moors, Theories of emotion causation: A review, Cognition & Emotion 23 (2009) 625–662.

[31] G. Lange, W. James (Eds.), The emotions, Vol. 1., Williams & Wilkins Co, 1922.

[32] W. B. Cannon, The james–lange theory of emotions: A critical examination and an alternative theory, The American Journal of Psychology 39 (1927) 106–124.

[33] S. Schachter, J. Singer, Cognitive, social, and physiological determinants of emotional state., Psychological Review 69 (1962) 379–399.

[34] Ortony, G. L. Clore, A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, 1988.

[35] G. H. Bower, Mood and memory, The American psychologist 36 (1981) 129–148.

[36] L. F. Barrett, The conceptual act theory: A pr´ecis, Emotion Review 6 (2014) 292–297.

[37] Russell, L. F. Barrett, Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant., Journal of Personality and Social Psychology 76 (1999) 805–819.

[38] R. Plutchik, The Emotions: Facts, Theories, and a New Model, Random House, 1962.

[39] L. F. Barrett, Discrete emotions or dimensions? the role of valence focus and arousal focus, Cognition and Emotion 12 (1998) 579–599.

[40] S. Hamann, Mapping discrete and dimensional emotions onto the brain: controversies and consensus, Trends in Cognitive Sciences 16 (2012) 458–466.

[41] H. Hoffmann, A. Scheck, T. Schuster, S. Walter, K. Limbrecht, H. C. Traue, H. Kessler, Mapping discrete emotions into the dimensional space: An empirical approach, in: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2012, pp. 3316–3320.

[42] P. Ekman, Basic emotions, in: Handbook of cognition and emotion, John Wiley & Sons, Ltd, 2005, pp. 45–60.

[43] W. V. F. Paul Ekman, P. Ellsworth, Emotion in the Human Face, Pergamon, 1972.

[44] R. E. Jack, O. G. Garrod, P. G. Schyns, Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time, Current Biology 24 (2014) 187–192.

[45] B. Mauss, M. D. Robinson, Measures of emotion: A review, Cognition and Emotion 23 (2009) 209–237.

[46] Orthony, A. C. GL & Collins, A.(1988). The cognitive structure of emotions.

[47] R. David C, T. Jennifer M, A comparison of dimensional models of emotion: evidence from emotions, prototypical events, autobiographical memories, and words, Memory (Hove, England) 17 (2009).

[48] J. A. Russell, A circumplex model of affect., Journal of Personality and Social Psychology 39 (1980) 1161– 1178. URL:

[49] D. Watson, A. Tellegen, Toward a consensual structure of mood., Psychological Bulletin 98 (1985) 219–235.

[50] R. Plutchik, The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, American Scientist 89 (2001) 344–350.

[51] J. A. Russell, A. Mehrabian, Evidence for a three-factor theory of emotions, Journal of Research in Personality 11 (1977) 273–294.

[52] Y. Fang, R. Rong, J. Huang, Hierarchical fusion of visual and physiological signals for emotion recog- nition, Multidimensional Systems and Signal Processing 32 (2021) 1103–1121.

[53] X. Zhang, J. Liu, J. Shen, S. Li, K. Hou, B. Hu, J. Gao, T. Zhang, B. Hu, Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine, IEEE Transactions on Cybernetics 51 (2021) 4386–4399.

[54] F. Panahi, S. Rashidi, A. Sheikhani, Application of fractional fourier transform in feature extraction from electrocardiogram and galvanic skin response for emotion recognition, Biomedical Signal Processing and Control 69 (2021) 1746–8094.

[55] Z. Farhoudi, S. Setayeshi, Fusion of deep learning features with mixture of brain emotional learning for audio-visual emotion recognition, Speech Communication 127 (2021) 92–103.

[56] D. Nguyen, D. T. Nguyen, R. Zeng, T. T. Nguyen, S. Tran, T. K. Nguyen, S. Sridharan, C. Fookes, Deep auto-encoders with sequential learning for multimodal dimensional emotion recognition, IEEE Transactions on Multimedia (2021) 1–10.

[57] J. Shi, C. Liu, C. T. Ishi, H. Ishiguro, Skeleton-based emotion recognition based on two-stream self-attention enhanced spatial-temporal graph convolutional network, Sensors (Switzerland) 21 (2020) 205.

[58] P. Tzirakis, J. Chen, S. Zafeiriou, B. Schuller, End-to-end multimodal affect recognition in real-world environments, Information Fusion 68 (2021) 46–53.

[59] D. Ayata, Y. Yaslan, M. E. Kamasak, Emotion recognition from multimodal physiological signals for emotion aware healthcare systems, Journal of Medical and Biological Engineering 40 (2020) 149–157.

[60] D. Y. Choi, D.-H. Kim, B. C. Song, Multimodal attention network for continuous-time emotion recognition using video and EEG signals, IEEE Access 8 (2020) 203814–203826.

[61] Y. Cimtay, E. Ekmekcioglu, S. Caglar-Ozhan, Cross-subject multimodal emotion recognition based on hybrid fusion, IEEE Access 8 (2020) 168865–168878.

[62] M. F. H. Siddiqui, A. Y. Javaid, A multimodal facial emotion recognition framework through the fusion of speech with visible and infrared images, Multimodal Technologies and Interaction 4 (2020) 46.

[63] E. Ghaleb, M. Popa, S. Asteriadis, Metric learning based multimodal audio-visual emotion recognition, IEEE Multimedia (2019) 1–1.

[64] Y. Huang, J. Yang, S. Liu, J. Pan, Combining facial expressions and electroencephalography to enhance emotion recognition, Future Internet 11 (2019) 1–17.

[65] C. Tan, G. Ceballos, N. Kasabov, N. P. Subramaniyam, FusionSense: Emotion classification using feature fusion of multimodal data and deep learning in a brain-inspired spiking neural network, Sensors (Switzerland) 20 (2020) 5328.

[66] Z. Wang, X. Zhou, W. Wang, C. Liang, Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video, International Journal of Machine Learning and Cybernetics 11 (2020) 923–934.

[67] Wang, Zhongmin, Xiaoxiao Zhou, Wenlang Wang, and Chen Liang. "Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video." *International Journal of Machine Learning and Cybernetics* 11, no. 4 (2020): 923-934.

[68] Gal, Viviane, Soumya Banerjee, and Dana V. Rad. "Identifying emotion pattern from physiological sensors through unsupervised EMDeep model." *Journal of Intelligent & Fuzzy Systems* Preprint (2020): 1-19.

[69] Chen, Genlang, Yi Zhu, Zhiqing Hong, and Zhen Yang. "EmotionalGAN: Generating ECG to Enhance Emotion State Classification." In *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, pp. 309-313. 2019.

[70] L. Cai, Y. Hu, J. Dong, S. Zhou, Audio-textual emotion recognition based on improved neural networks, Mathematical Problems in Engineering 2019 (2019) 1–9.

[71] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, G. Fortino, Human emotion recognition using deep belief network architecture, Information Fusion 51 (2019) 10–18.

[72] Mostafa, Amr, Mahmoud I. Khalil, and Hazem Abbas. "Emotion recognition by facial features using recurrent neural networks." In *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pp. 417-422. IEEE, 2018.

[73] Wiem, Mimoun Ben Henia, and Zied Lachiri. "Emotion classification in arousal valence model using MAHNOB-HCI database." *International Journal of Advanced Computer Science and Applications* 8, no. 3 (2017).

[74] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, I. Patras, AMIGOS: A dataset for affect, personality and mood research on individuals and groups, IEEE Transactions on Affective Computing 12 (2021) 479–493.

[75] M. Doyran, A. Schimmel, P. Baki, K. Ergin, B. Tu ̈rkmen, A. A. Salah, S. C. J. Bakkes, H. Kaya, R. Poppe, A. A. Salah, MUMBAI: multi-person, multimodal board game affect and interaction analysis dataset, Journal on Multimodal User Interfaces 15 (2021) 373–391.

[76] J. Kossaifi, R. Walecki, Y. Panagakis, J. Shen, M. Schmitt, F. Ringeval, J. Han, V. Pandit, A. Toisoul, B. Schuller, K. Star, E. Hajiyev, M. Pantic, SEWA DB: A rich database for audio-visual emotion and sentiment research in the wild, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2021) 1022–1040.

[77] Park, C.Y., Cha, N., Kang, S., Kim, A., Khandoker, A.H., Hadjileontiadis, L., Oh, A., Jeong, Y. and Lee, U."K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations." *Scientific Data* 7.1 (2020): 1-16.

[78] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, R. Mihalcea, MELD: A multimodal multi-party dataset for emotion recognition in conversations, in: ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, Association for Computational Linguistics, 2019, p. 527–536.

[79] J. Pinto, Exploring Physiological Multimodality for Emotional Assessment, Technical Report, Instituto Su- perior T´ecnico (IST), Lisboa, Portugal, 2019.

[80] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, N. Sebe, ASCERTAIN: Emotion and personality recognition using commercial sensors, IEEE Transactions on Affective Computing 9 (2018) 147– 160.

[81] S. Katsigiannis, N. Ramzan, DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices, IEEE Journal of Biomedical and Health Informatics 22 (2018) 98–107.

[82] S. R. Livingstone, F. A. Russo, The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north american english, In PLoS ONE 13 (2018) e0196391.

[83] L. Zhang, S. Walter, X. Ma, P. Werner, A. Al-Hamadi, H. C. Traue, S. Gruss, "BioVid emo DB": A multimo dal database for emotion analyses validated by sub jective ratings, in: 2016 IEEE Symp o- sium Series on Computational Intelligence, IEEE, 2016, pp. 1–6.

[84] C. Busso, S. Parthasarathy, A. Burmania, M. AbdelWahab, N. Sadoughi, E. M. Provost, MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception, IEEE Transactions on Affective Computing 8 (2017) 67–80.

[85] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, N. Sebe, DECAF: MEG-based multimodal database for decoding affective physiological responses, IEEE Transactions on Affective Computing 6 (2015) 209–222.

[86] H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova, R. Verma, CREMA-d: Crowd-sourced emotional multimodal actors dataset, IEEE transactions on affective computing 5 (2014) 377–390.

[87] S. Park, H. S. Shim, M. Chatterjee, K. Sagae, L.-P. Morency, Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach, in: In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14), Association for Computing Machinery, ACM, 2014, pp. 50–57.

[88] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, DEAP: A database for emotion analysis using physiological signals, IEEE Transactions on Affective Computing 3 (2012) 18–31.

[89] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, A multimodal database for affect recognition and implicit tagging, IEEE Transactions on Affective Computing 3 (2012) 42–55.

[90] Barra, S. F. PhysioUnicaDB: A dataset of EEG and ECG simultaneously acquired. *Pattern Recognition Letters*, (2019). 119-122.

[91] Schmidt, P. R. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. *International Conference on Multi- modal Interaction (ICMI '18)* (p. 9). 2018. Boulder, CO,USA: ACM.

[92] Lin, J. P. an Explainable Deep Fusion Network for Affect Recognition Using Physiological Signals. *28th ACM International Conference on Information and Knowl- edge Management (CIKM '19),* (p. 4). 2019. Beijing, China: ACM.

[93] Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, *42*(4), 335-359.

[94] Li, W., Abtahi, F., & Zhu, Z. (2015, November). A deep feature based multi-kernel learning approach for video emotion recognition. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 483-490).

[95] Li, W., Tsangouri, C., Abtahi, F., & Zhu, Z. (2018). A recursive framework for expression recognition: from web images to deep models to game dataset. *Machine Vision and Applications*, *29*(3), 489-502.

[96] Hossain, M. S., & Muhammad, G. (2019). Emotion recognition using deep learning approach from audio–visual emotional big data. *Information Fusion*, *49*, 69-78.

[97] Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, 1-41.

[98] Hazarika, D., Poria, S., Zimmermann, R., & Mihalcea, R. (2021). Conversational transfer learning for emotion recognition. *Information Fusion*, *65*, 1-12.

[99] Wang, Zhongmin, Xiaoxiao Zhou, Wenlang Wang, and Chen Liang. "Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video." *International Journal of Machine Learning and Cybernetics* 11, no. 4 (2020): 923-934.

[100] Chowdary, M. Kalpana, Tu N. Nguyen, and D. Jude Hemanth. "Deep learning-based facial emotion recognition for human–computer interaction applications." *Neural Computing and Applications*, 1-18, 2021.

[101] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[102] Kriman, Samuel, Stanislav Beliaev, Boris Ginsburg, Jocelyn Huang, Oleksii Kuchaiev, Vitaly Lavrukhin, Ryan Leary, Jason Li, and Yang Zhang. "Quartznet: Deep automatic speech recognition with 1d time-channel separable convolutions." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6124-6128. IEEE, 2020.

[103] Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., & Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, *46*(5), 1256-1266.

# LIST OF PUBLICATIONS

**Journal:**

1. A. Kumar, K. Sharma, A. Sharma, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers". *Pattern Recognition Letters*, Elsevier, ISSN: 0167-8655, Vol. 145(2021), pp: 81-87, DOI: 10.1016/j.patrec.2021.01.030 **[SCI, Impact Factor: 4.757].**

2. A. Kumar, K. Sharma, A. Sharma, "Genetically optimized Fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics". *Applied Soft Computing*, Elsevier ISSN: 1568-4946, Vol. 109 (2021), pp: 107525, DOI: 10.1016/j.asoc.2021.107525 **[SCIE, Impact Factor: 8.263].**

3. A. Sharma, K. Sharma, A. Kumar, "Real-Time Emotional Health Detection using Fine-Tuned Transfer Networks with Multimodal Fusion" *Neural Computing and Applications,* Springer, ISSN: 1433-3058 (2022), DOI: https://doi.org/10.1007/s00521-022-06913-2 **[SCIE, Impact Factor: 5.102].**

4. A. Kumar, K. Sharma, A. Sharma, "MEmoR: A Multimodal Emotion Recognition using Affective Biomarkers for Smart Prediction of Emotional Health for People Analytics in Smart Industries". *Image and Vision Computing*, Elsevier ISSN: 0262-8856, Vol. 123, 104483 (2022), DOI: https://doi.org/10.1016/j.imavis.2022.104483 **[SCI, Impact Factor: 3.860].**

**Conference:**

1. A. Kumar, A. Sharma, A. Arora, "Anxious Depression Prediction in Real-time Social Data" in proceedings of *International Conference on Advances in*

*Engineering Science Management & Technology 2019*, Uttaranchal University, Dehradun, arXiv preprint arXiv:1903.10222. **[SSRN].**

2. A. Kumar, K. Sharma, A. Sharma, "Empirical Analysis of Psychological Well-Being of Students during the Pandemic with Rebooted Remote Learning Mode", in proceedings of *International Conference on Data Analytics and Management (ICDAM-2022)*, Springer, 2022, Poland, Europe. **[WoS, Scopus].** DOI: 10.1007/978-981-19-7615-5.
3. "FTL-Emo: Federated Transfer Learning for Privacy Preserved Biomarker based automatic Emotion Recognition", Accepted at 4th International Conference on Data Analytics & Management (ICDAM-2023), London Metropolitan University London, 23rd – 24th June 2023, **[WoS, Scopus]**.

**Under Review**:

1. "DREAM: Deep Learning-based Recognition of Emotions from Multiple Affective Modalities using consumer-grade body sensors and video cameras". Communicated in IEEE Transactions on Consumer Electronics**, [SCI, Impact Factor: 4.414].**

**To be Communicated:**

1. "*Multimodal Affective Human State Recognition: A survey*". *ACM Computing Surveys,* ACM. **[SCI, Impact Factor: 10.282].**

<u>*Curriculum Vitae*</u>


# *ADITI SHARMA*

**B.Tech., M. Tech., Ph.D.**

*Mobile: +91-9877613335*

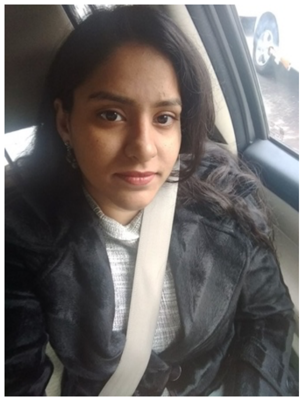
*Email: [aditisharma9420@gmail.com](mailto:aditisharma9420@gmail.com)*

*[https://orcid.org/0000-0003-2840-3330](https://orcid.org/0000-0003-2840-3330)*

*WOS: AAC-6040-2021*

*Scopus:* Aditi Sharma

*Google Scholar:* Aditi Sharma

---

## HIGHLIGHTS

- All India Rank- **571 in GATE** 2015, Score - 717.
- **UGC NET 2018** qualified with **99.3 Percentile**.
- Secured **2nd** position in Master's at DTU.
- Ranked **170** in all India merit list of NPIU by TEQIP.
- **Presented 4** Research Papers at International Conferences.
- **5** SCI/SCIE Journal Papers with Cumulative **IF: 23.98**.

Received ***"Commendable Research Award for Excellence in Research"*** from Delhi Technological University in 2022 & 2023.

## EDUCATIONAL QUALIFICATIONS

- **Ph.D. (Doctor of Philosophy), Computer Science Engineering, 2023**, Delhi Technological University, Delhi, India.

  *THESIS: Modelling and Application of Biomarkers for Affective State Mining Using Soft Computing Techniques.*
- **M.Tech. (Master of Technology), Software Engineering, 2017,** Degree with Distinction: CGPA-8.68, Delhi Technological University, Delhi, India.

  *THESIS: A Fuzzy logic based Text Summarization.*
- **B.Tech. (Bachelor of Technology), Computer Science & Engineering, 2015,** Degree with Distinction: 85%, Punjabi University, Patiala, Punjab, India.

  *PROJECT: QULOG- A question answering and blog website.*
- **GATE (Graduate Aptitude Test in Engineering),** Indian Institute of Technology, **2015, 2018, 2021, 2022**.

  *BEST SCORE (2015): All India Rank: 571, Percentile: 99.55, Score: 717.*
- **UGC-NET (National Eligibility Test- Assistant Professor),** National Testing Agency, **2018**.

  *PERCENTILE: 99.3.*

## AWARDS

- Commendable Research Award for excellence in Research from Delhi Technological University in 2022 & 2023.

## ACADEMIC /TEACHING EXPERIENCE **4 Years +**

- **Assistant Professor (Grade-1),** Department of Computer Science, Thapar Institute of Engineering and Technology, Patiala, India.
  **Jan' 2023 – Till Date.**
- **Assistant Professor (Grade-2),** Department of Computer Science/Information Technology, Jaypee University of Information Technology, Waknaghat, India.
  **Aug' 2022 – Jan' 2023.**
- **Assistant Professor (Guest Faculty),** Department of Information Technology, Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India.
  **Jan' 2021 – Jun 2021.**
- **Assistant Professor,** School of Computing, DIT University**,** Dehradun, India.
  **Sep' 2017 - Jan' 2021**.

## SUBJECTS TAUGHT

- FALL, 2017 — Discrete Mathematics
- WINTER, 2018 — Software Engineering, Web Technology
- FALL, 2018 — Discrete Mathematics, Cryptography and Network Security
- FALL, 2019 — Discrete Mathematics, Cryptography and Network Security
- WINTER, 2020 — Theory of Computation, Compiler Design
- FALL, 2020 — Discrete Mathematics
- WINTER, 2021 — Data Base Management System
- FALL, 2022 — Artificial Intelligence Techniques, Programming for Problem Solving Lab
- WINTER, 2023 — Object Oriented Programming

## TEACHING INTEREST

- Discrete Mathematics
- Software Engineering
- Machine Learning
- Theory of Computation
- Compiler Design
- Data Base Management System

## UNDERGRADUATE PROJECTS MENTORERD

- Detection of Suicidal Ideation using Machine Learning
- Result Prediction Using Machine Learning
- Sentiment Analysis of Tweets using Neuro-Fuzzy Approach
- Depression detection using deep learning
- Identifying opinion leaders on tweeter using soft computing
- HR Analytics using Machine Learning

## METHODOLOGY OF TEACHING

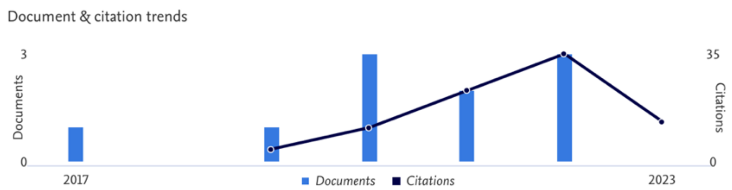
- **Lectures:** Important material from the text and outside sources are covered in class using white-board teaching & PowerPoint slides. During the lecture portion, all concepts are explained at length with help of real time applications, and then they are reinforced by using in-class lab activities.
- **Assignments:** Periodical homework assignments and tutorials are assigned to further reinforce the knowledge discovery process and to encourage the use of the supplemental material found in the text and discussing the solutions of all assignments and tutorials in detail, encouraging different methodologies to solve each problem incorporating Flip-Classroom methodology.
- **Class Project:** A group project consisting of 2-3 members for further substantiating the theoretical concepts studied in the subject course to embolden team building activities.

## RESEARCH

| Journal Paper | | Conference Papers | | | Book Chapter | |
|---|---|---|---|---|---|---|
| SCIE | Scopus | WOS | Scopus | SSRN | Scopus | Total |
| 5 (Cumulative IF: 23.988) | 1 | 2 (1 IEEE, 1 Springer) | 2 (1 ACM, 1 Springer) | 2 (Elsevier) | 1 (Springer) | 13 |

### Research Indicators*:*

<u>Google Scholar:</u>
Citations: 218
H-Index: 8
i10-Index: 7
<u>Scopus:</u>
Documents: 10
Citations: 86
H-Index: 6
<u>WoS:</u>
Documents: 6
Citations: 39
H-Index: 3




### Reviewer*:*

- ACM Transactions on Asian and Low-Resource Language Information Processing
- Information Fusion
- Wireless Communications and Mobile Computing

### TPC Member*:*

- Seventh International Conference on Parallel, Distributed and Grid Computing, Organized by Jaypee University of Information Technology, Waknaghat, India.

## AREAS OF RESEARCH INTEREST

- Affective Computing

• Machine Learning
• Sentiment Analysis
• Text Summarization
• Fuzzy Logic

# PUBLICATIONS

### *JOURNALS (SCI/SCIE)*

1. A. Kumar, **A. Sharma***, "Systematic Literature Review of Fuzzy Logic Based Text Summarization". *Iranian Journal of Fuzzy Systems*, ISSN: 2676-4334, Vol. 16 (2019), Issue. 5, pp: 45-59. DOI:10.22111/ijfs.2019.4906 **[SCIE, Impact factor: 2.006].**

2. A. Kumar*, K. Sharma, **A. Sharma**, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers". *Pattern Recognition Letters*, Elsevier, ISSN: 0167-8655, Vol. 145(2021), pp: 81-87, DOI: 10.1016/j.patrec.2021.01.030 **[SCI, Impact Factor: 4.757].**

3. A. Kumar, K. Sharma, **A. Sharma***, "Genetically optimized Fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics". *Applied Soft Computing*, Elsevier ISSN: 1568-4946, Vol. 109 (2021), pp: 107525, DOI: 10.1016/j.asoc.2021.107525 **[SCIE, Impact Factor: 8.263].**

4. **A. Sharma**, K. Sharma, A. Kumar*, "Real-Time Emotional Health Detection using Fine-Tuned Transfer Networks with Multimodal Fusion" *Neural Computing and Applications,* Springer, ISSN: 1433-3058, DOI: https://doi.org/10.1007/s00521-022-06913-2 **[SCIE, Impact Factor: 5.102].**

5. A. Kumar, K. Sharma, **A. Sharma***, "MEmoR: A Multimodal Emotion Recognition using Affective Biomarkers for Smart Prediction of Emotional Health for People Analytics in Smart Industries". *Image and Vision Computing*, Elsevier ISSN: 0262-8856, Vol. 123, 104483 (2022), DOI: https://doi.org/10.1016/j.imavis.2022.104483 **[SCI, Impact Factor: 3.860].**

### *JOURNALS (Scopus)*

**1.** **A. Sharma**, R. Ranjan, "Software Effort Estimation using Neuro Fuzzy Inference System: Past and Present". *International Journal on Recent Innovation Trends in Computing and Communication (IJRITCC),* ISSN: 2321-8169, Vol.5, Issue.8, pp: 78-83.

### *CONFERENCES (WOS Indexed)*

1. A. Kumar, **A. Sharma***, S. Sharma, S. Kashyap, "Performance Analysis of Keyword Extraction Algorithms Assessing Extractive Text Summarization" In *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, pp. 408-414, IEEE, 2017, Jaipur, India.
2. A. Kumar, K. Sharma, **A. Sharma***, "Empirical Analysis of Psychological Well-Being of Students during the Pandemic with Rebooted Remote Learning Mode", presented in *International Conference on Data Analytics and Management (ICDAM-2022)*, Springer, 2022, Poland, Europe.
3. A. Kumar, A. Sharma*, *"FTL-Emo: Federated Transfer Learning for Privacy Preserved Biomarker based automatic Emotion Recognition", Accepted at 4th International*

*Conference on Data Analytics & Management (ICDAM-2023), London Metropolitan University London, 23rd – 24th June 2023, [WoS, Scopus].*

***CONFERENCES (Scopus Indexed)***

1. R. Ranjan, **A. Sharma***, "Voice Control IOT Devices Framework for Smart Homes", in Proceedings of *International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, pp:57-67, DOI: 10.1007/978-981-15-3369-3, Springer, 2020, Chandigarh, India.

2. A. Kumar, **A. Sharma***, A. Nayyar, "Fuzzy Logic Based Hybrid Model for Automatic Text Summarization", in proceedings of *5th International Conference on Intelligent Information Technology (ICIIT 2020)*, pp: 7-15, ACM, 2020, Vietnam, ACM.

***CONFERENCES***

1. A. Kumar, **A. Sharma***, A. Arora, "Anxious Depression Prediction in Real-time Social Data" in proceedings of *International Conference on Advances in Engineering Science Management & Technology 2019*, Uttaranchal University, Dehradun, arXiv preprint arXiv:1903.10222.

2. R. Ranjan, **A. Sharma***, "Evaluation of Frequent Itemset Mining Platforms Using Apriori and FP-Growth Algorithm", $4^{th}$ International Conference on Computers & Management, Delhi, published in *International Journal of Information Systems & Management Science*, Vol. 2, No. 2, 2019, SSRN.

***BOOK CHAPTER (Scopus indexed)***

1. **A. Sharma***, D. Jain, "Development of Industry 4.0", In *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development*, pp: 23-38, Springer, Cham, 2020, ISSN: 2522-8714, DOI: 10.1007/978-3-030-14544-6.

## ACADEMIC DUTIES PERFORMED (AT DIT UNIVERSITY)

- Editor of quarterly newsletter of School of computing for AY 2020-2021.
- Academic advisor for first and second year students for AY 2020-2021.
- Coordinator of Summer training for AY 2019-2020
- Course Coordinator of Discrete Mathematics for AY 2019-2020.
- Coordinator of Course-file and Lab-File Management for AY 2019-2020.
- Coordinator of Result Analysis for AY 2019-2020.
- Coordinator of Attendance Compilation generation of debar list for AY 2019-2020.
- Coordinator of Aptitude Building for AY 2017-2018 and AY 2018-2019.
- Coordinator of Employee Enhancement Training for AY 2018-2019.
- Coordinator of "BTech in Computer Science Engineering in Big Data Analysis" in collaboration with IBM for AY 2018-2019.
- Admission Counsellor for Computer Science Students for June 2018.
- Course Coordinator of Software Engineering for Summer Term (2018).

- In-charge of Project Lab from 2018.
- Coordinator of Academic Probation students from 2018.

## FDP/WORKSHOP/SEMINAR/ CONFERENCE ATTENDED

- Attended FDP on "*Human Centre Computing*", offered by ATAL Academy from $15^{th}$ Feb-$19^{th}$ Feb 2021.
- Attended FDP on "*Blockchain Technology*", offered by EICT, at IIT Roorkee from $01^{st}$ - $08^{th}$ May 2020.
- Attended FDP on "*Machine Learning & its Applications*", offered by EICT, from IIT Roorkee on $13^{th}$ - $20^{th}$ April 2020.
- Presented paper at Springer Conference "*International Conference on Computing, Communications, and Cyber-Security*" (IC4S 2019) organized by C-DAC Mohali.
- Attended FDP on "*Deep Learning and Applications*", offered by EICT, at IIT Roorkee from $27^{th}$ - $31^{th}$ May 2019.
- Attended FDP on "Integration of Best International Pedagogical Practices to Indian Higher Education System" on 2 Jan 2019, at DIT University.
- Presented Research paper at Conference "*$4^{th}$ International Conference on Computers & Management (ICCM-2018)*", organized by G B Pant Govt. Engineering College, Delhi.
- Attended FDP on "*Machine Learning and Data Analytics with Python*", offered by EICT, IIT Roorkee at DIT University on $27^{th}$ - $31^{th}$ Dec 2018.
- Attended FDP on "*Basic programming using Python*", offered by FOSSEE group, IIT Bombay at DIT University on $14^{th}$ - $17^{th}$ May 2018.
- Presented paper at IEEE Conference "*International Conference on Computer, Communications and Electronics 2017*" (COMPTELIX 2017) organized by Manipal University, Jaipur.
- Attended workshop on "*Android*" at PEC, Chandigarh during $16^{th}$ and 17 March 2013.
- Attended workshop on "*Ethical Hacking"* at Punjabi university, Patiala on $8^{th}$ and 9 Feb 2012.
- Presented Seminar on "*Big Data Analysis using Soft Computing*" and "*Automatic Text Summarization*" in DTU.

## TRAINING

- Six Month Industrial Training at HCL CDC, Patiala
  Course: J2EE, January-May 2015

- Six week Summer Training at SPIC Centre of Excellence, Chandigarh
  Course: Application development in CORE JAVA, July-August 2013

- Six-week Summer Training at SLR InfoTech, Chandigarh
  Course: Red Hat Administration, June-August 2013

## PROFESSIONAL MEMBERSHIP

- Member IEEE - 92409345.
- Member ACM – 0278285.
- Life Member of CSI (Computer Society of India) - 2010000313.

- Life Member of IAENG (International Association of Engineers) - 260192.

## PERSIONAL DETAILS

- DOB: 21 April 1994.
- Marital Status: Un-Married.
- Permeant Residence: #749, HIG, Urban Estate, Phase-1, Patiala, Punjab, India.

## REFERENCES

- **Dr. Akshi Kumar** (PhD. Supervisor)
  Senior Lecturer, Dept. of Computing & Mathematics,
  **Manchester Metropolitan University,** Manchester, UK.
  *Mobile:* +447442577120
  *E-mail:* Akshi.kumar@mmu.ac.uk

- **Prof. Kapil Sharma (**PhD. Supervisor)
  Professor & Head of Dept., Dept. of Information Technology,
  **Delhi Technological University**, Delhi, India.
  *Mobile:* 9717470444
  *E-mail:* kapil@ieee.org

- **Dr. Vishal Bharti**
  Additional Director, Dept. of Computer Science & Engineering,
  **Chandigarh University**, Mohali, Punjab, India.
  *Mobile:* 9814837334
  *E-mail:* dhavleesh@gmail.com