EMOTION RECOGNITION USING FEDERATED PARADIGM BASED ON PHYSIOLOGICAL SIGNALS

A Thesis Submitted
In Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

Submitted by

NEHA GAHLAN

Roll No. 2K21/PHD/CS/02

Under the Supervision of

DIVYASHIKHA SETHIA

Associate Professor
Department of Software Engineering

To the

Department of Software Engineering

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India.

February, 2025

Acknowledgement

I want to express my sincere gratitude to my supervisor Dr. Divyashikha Sethia, Associate Professor, Department of Software Engineering for her motivation, counselling, constant support and encouragement, which have helped me in formulating an urge for this research thesis. Her exemplary hard work and insightful guidance have always helped me in overcoming difficulties.

My sincere regards to Prof. Prateek Sharma, Vice-Chancellor, Delhi Technological University for providing a platform for pursuing my Ph.D. work. I express my thankfulness to Prof. Ruchika Malhotra, Head, Department of Software Engineering, for her kind support and for providing the necessary facilities to undertake this research.

I would also like to express my heartfelt regards to my father (Mr. Dharmvir Singh), mother (Mrs. Sudesh Gahlan), brother (Deepak Gahlan), sister (Kiran Gahlan), and my dearest niece Saavi, whose support, inspiration, and kind help had been strong pillars of support throughout this venture.

Also, a huge thanks to my lab members who have continuously supported and helped me with their valuable suggestions as and when required.

Neha Gahlan
2K21/PHD/CS/02

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Neha Gahlan hereby certify that the work which is being presented in this thesis entitled **"Emotion Recognition using Federated Paradigm based on Physiological Signal"** in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out under the supervision of Dr. Divyashikha Sethia, Associate Professor, Dept. of Software Engineering, DTU.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Neha Gahlan
2K21/PHD/CS/02

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

This to certify that **Neha Gahlan** (Roll no. 2K21/PHD/CS/02) has carried out her research work presented in this thesis entitled **"Emotion Recognition using Federated Paradigm based on Physiological Signal"** for the award of Doctor of Philosophy from Department of Software Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Supervisor
Dr. Divyashikha Sethia
Associate Professor
Department of Software Engineering
Delhi Technological University

Date:

ABSTRACT

Emotions are complex psychological states that involve physiological arousal, cognitive interpretation, and behavioural expression, influencing how individuals experience and respond to events. Various modalities, including facial expressions, subjective psychological tests, and physiological signals, can recognize human emotional states, out of which physiological signals have several advantages over the other modalities, including greater sensitivity to internal feelings and the ability to provide continuous, real-time data for accurate emotional monitoring.

In this context, automated *Emotion Recognition Systems* (ERS) are gaining popularity in predicting human emotions and enhancing health and decision-making. These systems utilize machine learning (ML) and deep learning (DL) algorithms to process wearable biosensor data and classify emotions with high precision and reliability. The automated ERS using traditional ML and DL algorithms directly accesses the user's raw physiological data to train the model and further classify emotions. It results in a significant loss of privacy protection for the user's sensitive physiological information.

This thesis aims to enhance the automated ERS by improving data privacy concerns and integrity using a novel Federated Learning (FL) paradigm. Unlike traditional machine learning techniques, the FL paradigm creates a decentralized environment (client and server ends), allowing users to transmit only the model weights generated locally on their device rather than the complete raw physiological data to a central server. The server aggregates these weights to create a global model aggregator that updates after each iteration. Apart from privacy, this thesis addresses the research gaps for (1) Lack of multi-modality in FL-based automated ERS with physiological data input; (2) Restricted emotion dimensions in FL-based automated ERS, exploring a smaller range of emotions, (3) Existing FL-based automated ERS fails to address the data heterogeneities occurring in a federated environment.

Firstly, the thesis presents a comprehensive literature review of automated ERS using physiological signals. It includes emotion models, physiological signals, the relation between emotions and physiological signals, technical background including data processing for physiological signals, ML and DL models, and the related works of FL for ERS.

Secondly, the thesis proposes a privacy-preserved emotion recognition architecture for multi-modal physiological data combining EEG, ECG, GSR and RESP signal data. For this, the thesis proposes an **F**L-based **M**ulti-modal **E**motion **R**ecognition **S**ystem (**F-MERS**) for classifying emotions using Valence, Arousal, and Dominance emotion dimensions. The thesis validates the proposed F-MERS with three different emotion datasets, proving it robust achieving an average testing accuracy of 83.02% with AMIGOS, 86.51% with DEAP, and 75.19% with DREAMER. It assesses its classification performance, scalability with different client distributions, convergence speed, and communication computation (in terms of averaging and training times) and discusses the experimental results. The F-MERS did not address data heterogeneity present in the multi-modal physiological data and the federated environment.

Thirdly, the thesis overcomes the challenge of data heterogeneity, lacking in the existing works of FL, by proposing an enhanced **A**ttention-based **F**ederated **L**earning for **E**motion recognition using **M**ulti-modal **P**hysiological data (**AFLEMP**) architecture. AFLEMP removes the Variation Data Heterogeneity (VDH) occurring while combining multiple physiological data together by implementing attention mechanisms at the client end. It proposes a novel Scaled-Weighted Federated Averaging (SWFA) algorithm for the server end to reduce the Imbalanced Data Heterogeneity (IDH) occurring due to imbalanced data distribution at the client end within a federated environment. The thesis validates the proposed AFLEMP with two different emotion datasets, achieving the testing accuracy of 90.11% with AMIGOS and 85.12% with DREAMER, proving it to be robust. For assessing the AFLEMP, the thesis evaluates and contrasts it with other FL algorithms for ERS in terms of classification performance, convergence speed and communication computation (in terms of averaging and training times) and discusses its experimental results.

Fourthly, the thesis presents that the proposed AFLEMP provides multi-dimensionality in terms of emotion dimensions, which is lacking in the existing works of FL for ERS.

For this, the thesis proposes AFLEMP to classify a wide spectrum of emotions using a 3D-VAD model of emotions (including Valence-Arousal-Dominance together). The thesis presents the experimental results of the proposed AFLEMP for its classification performance for Valence-Arousal-Dominance together and individually.

The research presented in this thesis contributes to the field of emotion recognition based on physiological signals by exploring FL. The FL techniques can assist for providing privacy of emotion recognition systems using multi-modal physiological signals. The proposed F-MERS and AFLEMP are robust, efficient in communication, and scalable.

List of Publications

The research presented in this thesis is primarily based on the following peer-reviewed articles:

Papers Accepted/Published in International Journals

Journal 1: N. Gahlan and D. Sethia. "Federated Learning in Emotion Recognition Systems based on Physiological Signals for Privacy Preservation: A Review" **Multimedia Tools and Applications**: 1-69. June 2024. (**SCIE, Impact factor-3.0, Publisher: Springer**). Doi: https://doi.org/10.1007/s11042-024-19467-3.

Journal 2: N. Gahlan and D. Sethia. "AFLEMP: Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data." **Biomedical Signal Processing and Control** 94 (2024): 106353. April 2024. (**SCIE, Impact Factor: 4.9, Publisher: Elsevier**). Doi: https://doi.org/10.1016/j.bspc.2024.106353.

Journal 3: N. Gahlan and D. Sethia. "Federated Learning Inspired Privacy Sensitive Emotion Recognition based on Multi-Modal Physiological Sensors." **Cluster Computing** (2023): 1-23. September 2023. (**SCIE, Impact Factor: 3.6, Publisher: Springer**). Doi: https://doi.org/10.1007/s10586-023-04133-4.

Papers Accepted/Published in International Conferences

Conference 1: N. Gahlan, D. Sethia, and S. Ray. "Emotion Analysis using ASMR and Physiological signals (EMAP) with Federated Learning." In *Proc. IEEE EMBS International Conference on Body Sensor Networks, 2024.* Doi: 10.1109/BSN63547.2024.10780682

Conference 2: N. Gahlan and D. Sethia. "Emotion Recognition from Facial Expressions using Deep Recurrent Attention Network." In *Proc. IEEE 16th International Conference on COMmunication Systems & NETworkS (COMSNETS), 2024.* Doi: 10.1109/COMSNETS59351.2024.10427068.

Conference 3: **N. Gahlan**, et al. "Multifaceted Discrete Emotion Recognition from EEG Physiological Signals via Machine Learning Techniques." In *Proc. IEEE 5th International Conference for Emerging Technology (INCET), 2024.* Doi: 10.1109/INCET61516.2024.10592936.

Conference 4: **N. Gahlan** and D. Sethia. "Three-Dimensional Emotion State Classification based on EEG via Empirical Mode Decomposition." In *Proc. IEEE International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), 2023.* Doi: 10.1109/ICAIA57370.2023.10169633.

Conference 5: **N. Gahlan** et al. "FedEmo: A privacy-preserving framework for emotion recognition using EEG physiological data." In *Proc. IEEE 15th International Conference on COMmunication Systems & NETworkS (COMSNETS). Jan 2023.* Doi: 10.1109/COMSNETS56262.2023.10041308.

Conference 6: **N. Gahlan** and D. Sethia. "Fed-MLP: Federated Framework for Recognizing Human Emotions via Physiological Signal EEG." *International Conference on Communication, Security and Artificial Intelligence. IEEE, 2022.* **Presented.**

Papers Under Review in Journals/Conferences

Journal 4: **N. Gahlan** and D. Sethia. "Emotional impacts of auditory Autonomous Sensory Meridian Response using Multi-modal Physiological Signals: A Quantitative Analysis" **Journal of Machine Learning and Cybernetics**. Submitted on Oct 2024. **Under Review. SCIE, Impact factor-3.1, Publisher: Springer**).

Contents

ABSTRACT v

List of Publications viii

List of Tables xvii

List of Figures xx

Chapter 1: Introduction 1

1.1 Motivation and Background 4

1.1.1 Why Emotion Recognition? 4

1.1.2 Background for Automated Emotion Recognition using Physiological Signals 4

1.1.3 What are the Privacy Concerns in Automated in ERS? 6

1.1.4 Why Federated Learning (FL) Paradigm? 6

1.2 Research Gaps 7

1.2.1 Research Gap 1: Lack of Privacy 7

1.2.2 Research Gap 2: Limited FL for ERS 8

1.2.3 Research Gap 3: Less Emotion Dimensions 8

1.2.4 Research Gap 4: Restricted Modality 8

1.2.5 Research Gap 5: No handling of Data Heterogeneity 8

1.3 Problem Statement 9

1.4 Research Objectives and Contribution of the Thesis 9

1.5 Significance of the Thesis 11

1.6 Thesis Overview 11

Chapter 2: Background for Emotions and Physiological Signals **14**

2.1 Emotion Models 14

2.1.1 Discrete Emotion Models 15

2.1.2 Continuous Emotion Models 16

2.1.2.1 2-Dimensional Valence-Arousal (2D-VA) Emotion Model 16

2.1.2.2 3-Dimensional Valence-Arousal-Dominance (3D-VAD) Emotion Model 17

2.2 Emotion Annotation and Ground Truth 17

2.2.1 Creation of Emotion Classification Labels 18

2.3 Physiological Signals and Emotions 21

2.3.1 Electroencephalographic (EEG) Signals 21

2.3.1.1 Significance of EEG for Emotions 23

2.3.1.2 Features from EEG signals 24

2.3.1.3 Related Works of EEG for Emotion Recognition . . 25

2.3.2 Electrocardiographic (ECG) Signals 26

2.3.2.1 Significance of ECG for Emotions 27

2.3.2.2 Heart Rate (HR) from ECG Signal 27

2.3.2.3 Significance of HR for Emotions 28

2.3.2.4 Features from ECG Signal 28

2.3.2.5 Related Works of ECG for Emotion Recognition . . 30

2.3.3 Galvanic Skin Response (GSR) 31

2.3.3.1 Significance of GSR for Emotions 33

2.3.3.2 Features from GSR Signal 33

2.3.3.3 Related Works of GSR for Emotion Recognition . . 35

2.3.4 Respiration (RESP) 36

2.3.4.1 Significance of RESP for Emotions 37

2.3.4.2 Features from RESP Signal 37

2.3.4.3 Related Works of RESP for Emotion Recognition . . 38

2.4 Multi-Modal Physiological Signals 38

2.5 Data Processing of Physiological Signals 40

2.5.1 Data Pre-Processing 40

2.5.1.1 Filtering . . . 40
2.5.1.2 Artefact Removal and Denoising . . . 41
2.5.2 Feature Extraction . . . 43
2.6 Summary . . . 43

Chapter 3: Open Emotion Datasets **45**
3.1 Wearable Stress and Affect Detection (WESAD) [195] . . . 46
3.2 SJTU Emotion EEG Dataset (SEED) [299] . . . 46
3.3 A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices (DREAMER) [135] . . . 47
3.4 Continuously Annotated Signals of Emotion (CASE) [134] . . . 48
3.5 A Database for Emotion Analysis using Physiological signals (DEAP) [233] 48
3.6 BioVid Emo DB (EMDB) [148] . . . 49
3.7 A dataset for Multi-modal research of affect, personality traits and mood on Individuals and GrOupS (AMIGOS) [125] . . . 50
3.8 A multi-modal databaASe for impliCit pERsonaliTy and Affect recognitIoN (ASCERTAIN) [221] . . . 51
3.9 Chosen Emotion Datasets for Proposed FL-based Automated ERS . . . 51
3.9.1 Data processing for Emotion Datasets (AMIGOS, DEAP, DREAMER) . . . 52
3.9.2 Data Clipping . . . 53
3.9.3 Data Labelling . . . 53
3.10 Summary . . . 53

Chapter 4: Machine Learning for Automated ERS using Physiological Signals **55**
4.1 Machine Learning Models for Emotion Recognition using Physiological Signals . . . 55
4.2 Deep Learning Models for Emotion Recognition using Physiological Signals . . . 57
4.3 Privacy Concerns in traditional ML and DL for ERS using Physiological signals . . . 62
4.4 Summary . . . 63

Chapter 5: Literature Review of Federated Learning Paradigm for ERS **64**
5.1 Introduction to Federated Learning Paradigm 64
5.1.1 Types of Architecture of FL 65
5.1.1.1 Client-Server Architecture 65
5.1.1.2 Peer-to-Peer Architecture 66
5.1.2 Approaches of FL 66
5.1.2.1 Horizontal Federated Learning (HFL) 67
5.1.2.2 Vertical Federated Learning (VFL) 68
5.1.2.3 Federated Transfer Learning (FTL) 68
5.1.3 Aggregation Algorithm of FL 69
5.1.4 Popular Tools for Federated Learning Setup 72
5.2 Research Methodology 72
5.2.1 Research Questions 72
5.3 Related work of FL in Emotion Recognition 73
5.4 Answers to RQs 75
5.5 Limitations of existing Federated Learning in ERS 76
5.6 Evaluation Measures for Assessing Proposed FL Architectures 77
5.6.1 Classification Performance Measures 78
5.6.1.1 Confusion Matrix 78
5.6.1.2 Classification Accuracy 79
5.6.1.3 Precision 79
5.6.1.4 Recall 80
5.6.1.5 F1-Score 80
5.6.2 Scalability Measures 81
5.6.2.1 Training Time 81
5.6.2.2 Model Accuracy 81
5.6.3 Communication Computation Measures 81
5.6.3.1 Averaging Time 82
5.6.3.2 Convergence Speed 82
5.6.3.3 Binary cross-entropy 82
5.6.3.4 Categorical Cross-Entropy 83
5.7 Summary 84

Chapter 6: F-MERS for Privacy to Emotion Recognition System 85
6.1 Experimental Methodology . . . 86
6.1.1 Data Processing . . . 86
6.1.2 Architecture of F-MERS . . . 89
6.1.3 Experimental Setup . . . 92
6.1.4 Validation Scenarios of the architecture . . . 93
6.1.5 Evaluation Measures for F-MERS . . . 93
6.2 Results and Discussion . . . 94
6.2.1 Subject-Dependent Results . . . 94
6.2.2 Subject-Independent Results . . . 99
6.2.3 Communication and Scalability Measures . . . 104
6.2.4 Discussion . . . 105
6.2.4.1 Limitations . . . 105
6.3 Summary . . . 106

Chapter 7: Attention for Variation Data Heterogeneity in multi-modal ERS 108
7.1 Variation Data Heterogeneity (VDH) . . . 110
7.2 Experimental Methodology . . . 111
7.2.1 Data Processing . . . 111
7.2.2 Centralized ANN with Attention . . . 113
7.2.3 Attention Mechanisms . . . 114
7.2.4 Architecture of AFLEMP . . . 116
7.2.5 Experimental Setup . . . 118
7.2.6 Evaluation Measures for AFLEMP . . . 119
7.2.7 Testing of the architecture . . . 119
7.3 Results and Discussion . . . 120
7.3.1 Accuracy comparison for proposed multi-modal AFLEMP with different attention mechanisms . . . 120
7.3.2 Other performance measures for proposed multi-modal AFLEMP 122
7.3.2.1 Classification measures . . . 122
7.3.2.2 Scalability measures . . . 123
7.3.2.3 Communication Computation Measures . . . 123

7.3.3 Performance comparison of proposed multi-modal FL (AFLEMP) with existing FL works . . . 125
7.4 Summary . . . 125

Chapter 8: Scaled Weighted Federated Averaging (SWFA) for Imbalanced Data Heterogeneity in Multi-modal ERS **127**
8.1 Imbalanced Data Heterogeneity (IDH) . . . 128
8.2 Experimental Methodology . . . 129
8.2.1 Scaled Weighted Federated Averaging (SWFA) . . . 129
8.2.2 AFLEMP with SWFA . . . 132
8.2.2.1 Data Description . . . 132
8.2.2.2 Feature Extraction . . . 134
8.2.2.3 Attention Mechanism - Transformer . . . 135
8.2.2.4 Base Classifier - ANN . . . 135
8.2.2.5 Federated data Partitioning . . . 135
8.2.2.6 Federated Environment via SWFA . . . 136
8.2.3 Experimental Setup . . . 136
8.2.4 Evaluation Measures for AFLEMP with SWFA . . . 137
8.2.5 Testing of the architecture . . . 137
8.3 Results and Discussion . . . 137
8.3.1 Performance measures for proposed multi-modal AFLEMP with SWFA . . . 138
8.3.1.1 Classification Accuracies . . . 138
8.3.1.2 Other Predictive Measures . . . 139
8.3.1.3 Communication Computation Measures . . . 140
8.3.2 Performance comparison of proposed multi-modal AFLEMP with SWFA and other FL aggregation algorithms . . . 141
8.3.3 Performance comparison of proposed multi-modal AFLEMP with SWFA with existing FL works . . . 143
8.4 Summary . . . 143

Chapter 9: Conclusion, Future Work & Social Impact **145**
9.1 Conclusion & Future Work 145
9.1.1 Lessons Learnt 147
9.1.2 Ethical Implications and Responsible Use 150
9.1.2.1 Privacy Concerns 150
9.1.2.2 Potential Biases 151
9.1.2.3 Responsible Uses 151
9.1.2.4 Limitations 152
9.1.3 Future Work 153
9.2 Social Impact & Applications 153

References **158**
Biography 184
Paper Proofs 185

List of Tables

2.1 Mapping of emotions using valence, arousal, and dominance individually (from 3D-VAD). . . . 19
2.2 Mapping of emotions using valence, arousal, and dominance together (from 3D-VAD) 20
2.3 Wearable devices that use physiological signals to track emotions [238]. 22
2.4 Description of EEG signal features in time and frequency domain. . . 25
2.5 Description of ECG signal features in time and frequency domain. . . 31
2.6 Description of GSR signal features in time and frequency domain. . . 35
2.7 Description of Respiration signal features in time domain. 37
2.8 Benefits and limitations of physiological signals. 38

3.1 Details of the 18 film clips shown to each subject in DREAMER [135]. 47
3.2 Details of the 15 film clips shown to each subject in EMDB dataset [148]. 49
3.3 Details of the 16 short videos shown to each subject in AMIGOS dataset [125]. . . . 50
3.4 Summary of openly accessible datasets for ERS based on physiological signals. . . . 52
3.5 Descriptions with pre-processing of the emotion dataset used in the thesis. 54
3.6 Emotions ratings for open emotion datasets using 3D-VAD 54

4.1 Summary of ML models for automated ERS. 58
4.2 Summary of DL models for automated ERS. 61

5.1 Characteristics comparison between FL aggregation algorithms. . . . 71
5.2 Tools for Federated Learning 72
5.3 Research questions 73
5.4 Recent studies on emotion recognition based on physiological signals using Federated Learning. . . . 74

5.5 Confusion matrix representation for binary classification 78
5.6 List of evaluation measures for proposed Federated Learning based emotion recognition architectures. 83

6.1 Brief description of all the datasets for F-MERS. 87
6.2 Extracted features from ECG, EEG, GSR, RESP signals for F-MERS. 87
6.3 Partitioning subjects into clients for each dataset (1 Client = 1 Subject) for F-MERS. 90
6.4 Model parameters of the proposed F-MERS. 91
6.5 Partitioning subjects into clients: Subject-Independent scenario for F-MERS. 92
6.6 Training and testing accuracies for proposed F-MERS with 500 rounds in **subject-dependent** scenario. 96
6.7 F1-Score results (during testing) for the proposed F-MERS with all datasets in a **subject-dependent** scenario 97
6.8 Confusion matrix for the proposed F-MERS in **subject-dependent** scenario. 98
6.9 Testing accuracy for the proposed F-MERS and the Non-FL across all datasets in a **subject-dependent** scenario with maximum number of clients. 98
6.10 Training and testing accuracies for the proposed F-MERS with 500 rounds in **subject-independent** scenario. 101
6.11 F1-Score results (during testing) for the proposed F-MERS with all datasets in **subject-independent** scenario 102
6.12 Confusion matrix for the proposed F-MERS in **subject-independent** scenario. 102
6.13 Testing accuracy comparison for the proposed F-MERS and the Non-FL across all datasets in a **subject-independent** scenario with maximum number of clients. 103
6.14 Training and averaging time (in seconds) for proposed F-MERS in subject-independent scenario. Clients (C), Rounds (R). 104
6.15 Comparison of F-MERS with existing FL-based ERS. 106

7.1 Brief description of datasets for AFLEMP. 111
7.2 Mapping of emotions using 3D-VAD for AFLEMP. 112
7.3 Extracted features from ECG, EEG, and GSR signals for AFLEMP. . . 113
7.4 Model parameters of AFLEMP architecture at the client end. 114
7.5 Federated data distribution for all clients (C) with different subjects (S) for AFLEMP. 118
7.6 Train and test set of clients for proposed AFLEMP for both the datasets. 119
7.7 Training and testing accuracies for **AFLEMP** for different attention mechanisms. 120
7.8 Evaluation measures for testing the AFLEMP (**with Transformer**) for 200 rounds. 122
7.9 Training and averaging time in seconds for the AFLEMP **(with Transformer).** 124
7.10 Performance comparison with existing FL works for ERS with physiological signals. 125

8.1 Brief description of datasets for AFLEMP with SWFA. 133
8.2 Extracted features from ECG, EEG and GSR signals for AFLEMP with SWFA. 134
8.3 Model parameters of the local model run at the client end in AFLEMP with SWFA. 135
8.4 Federated data distribution for all clients (C) with different subjects (S) for AFLEMP with SWFA. 136
8.5 Train and test set of clients for AFLEMP with SWFA. 137
8.6 Training and testing accuracies for proposed AFLEMP with SWFA. . 138
8.7 Evaluation measures for testing the proposed AFLEMP with SWFA for 200 rounds. 140
8.8 Training and averaging time in seconds for the proposed AFLMEP with SWFA. 141
8.9 Testing accuracies and averaging time for AFLEMP with SWFA and other FL algorithms. 142
8.10 Performance comparison with existing FL works for emotion recognition with physiological signals. 143

List of Figures

1.1 Branching representation for the emotion indicators 2
1.2 Sample indicators for emotion recognition. 3
1.3 Methodology for automated Emotion Recognition System (ERS) [244]. 5
1.4 Workflow of steps for Emotion Recognition System (ERS) 5
1.5 Schematic diagram of the Federated Learning environment. 7

2.1 Plutchik's model of emotions [219] 15
2.2 Models of emotions: (A) 2D-VA [286], (B) 3D-VAD [184] 16
2.3 Self-Assessment Manikins (SAM) with ratings for emotion recognition. 18
2.4 Brain lobe's functions for emotional activities. 21
2.5 Frequency band distribution of raw EEG signal and its impacts on brain 22
2.6 19 Channel EEG headset montage mapping over brain regions. 23
2.7 ECG original signal waveform and its filtered signal [201] 26
2.8 Heart Rate and Heart Rate Variability relation. 27
2.9 Heartbeats, RR interval, Histogram of RR interval [257] for a subject in neutral state. 28
2.10 GSR signal and its components [203]. 32
2.11 GSR signal and SCR components: Onsets, Peaks and Recovery [203] 34
2.12 Raw and filtered Respiration signal [203]. 36
2.13 Illustration for multi-modal physiological signals fusion. 39
2.14 Workflow for physiological data pre-processing steps. 40
2.15 Overlapping windows on EEG signal [104]. 43

3.1 Experimental setup for WESAD dataset collection [195]. 46
3.2 Experimental setup for SEED dataset collection [299]. 47
3.3 Experimental setup for CASE dataset collection [134] 48

3.4 Experimental setup for DEAP dataset collection [233]. 49
3.5 Experimental setup for AMIGOS dataset collection [125]. 50
3.6 Experimental setup for ASCERTAIN dataset collection [221]. 51

4.1 Distribution of ML models for automated motion Recognition System (ERS). 56
4.2 Distribution of DL models for automated Emotion Recognition System (ERS). 59

5.1 Federated Learning for Emotion Recognition System (ERS) with physiological signals. 65
5.2 Architectures of Federated Learning: (A). Client-Server Architecture, (B). Peer-to-Peer Architecture 66
5.3 Horizontal data distribution of physiological data from two clients. . . 67
5.4 (a) Horizontal FL, (b) Vertical FL and (c) Federated Transfer Learning 67
5.5 Vertical data distribution of physiological data from two clients. . . . 68
5.6 Illustration of different Federated Learning aggregation algorithms. . . 70
5.7 Components of Federated Learning 71
5.8 Search domains for reviewing. 73

6.1 Feature Fusion for F-MERS. 88
6.2 Architecture of proposed F-MERS with AMIGOS [125]. 89
6.3 Data divisions of clients for F-MERS (with AMIGOS [125]). 90
6.4 Testing accuracies for proposed F-MERS for AMIGOS [125] in **Subject-dependent** scenario. 94
6.5 Testing accuracies for proposed F-MERS for DEAP [233] in **Subject-dependent** scenario. 95
6.6 Testing accuracies for proposed F-MERS for DREAMER [135] in **Subject-dependent** scenario. 95
6.7 Testing accuracies for proposed F-MERS for AMIGOS [125] in **Subject-independent** scenario. 99
6.8 Testing accuracies for proposed F-MERS for DEAP [233] in **Subject-independent** scenario. 100

6.9 Testing accuracies for proposed F-MERS for DREAMER [125] in **Subject-independent** scenario. 100

7.1 Architecture of the ANN with attention layers for AFLEMP. 116
7.2 Complete workflow for AFLEMP architecture with AMIGOS [125]. . 117
7.3 Testing accuracies for **AFLEMP** for Valence, Arousal, Dominance, and 3D-VAD. 121
7.4 Testing accuracies for the **AFLEMP** for 3D-VAD in different iterations with different attention mechanisms. 122
7.5 Loss curves for the AFLEMP for AMIGOS [125] and DREAMER [135] for 3D-VAD **(with Transformer)**. 124

8.1 Imbalanced Data Heterogeneity (IDH) in a federated environment. . . 128
8.2 Architecture of proposed AFLEMP with proposed aggregation algorithm SWFA. 133
8.3 Testing accuracies for the proposed AFLEMP with SWFA for Valence, Arousal, Dominance, and 3D-VAD for all rounds of iterations. 139
8.4 Loss curves for the proposed AFLEMP with SWFA for 3D-VAD. . . . 142

9.1 Health monitoring using smart wearable sensors 154
9.2 In-Home health monitoring using smart wearable sensors 154
9.3 Emotional health monitoring while E-learning. 155
9.4 Emotional health monitoring in office corporate environments. 156
9.5 Pet Emo robot with interactive sensors 156
9.6 Animoji: Augmented reality masks to express emotions. 157

List of Abbreviations

A	Arousal
BPM	Beats Per Minute
BVP	Blood Volume Pulse
CNN	Convolutional Neural Networks
CVNNI	Coefficient of Variation of NN Intervals
CVSD	Coefficient of Variation of Successive Differences
D	Dominance
DL	Deep Learning
DLF	Decision-Level Fusion
DT	Decision Tree
DWFA	Dynamic Weighted Federated Averaging
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
EMG	Electromyogram
ERS	Emotion Recognition Systems
FATE	Federated AI Technology Enabler
FFT	Fast Fourier Transform
FL	Federated Learning
FLF	Feature-Level Fusion
FTL	Federated Transfer Learning
GSR	Galvanic Skin Response
HFL	Horizontal Federated Learning

List of Abbreviations

HR	Heart Rate
HRV	Heart Rate Variability
ICA	Independent Component Analysis
IDH	Imbalanced Data Heterogeneity
IMF	Intrinsic Mode Functions
KNN	K-Nearest Neighbour
LOOCV	Leave-One-Out Cross-Validation
LOSO	Leave One Subject Out
LSTM	Long Short-Term Memory
MIA	Model Inversion Attacks
ML	Machine Learning
MLP	Multilayer Perceptron
NB	Naive Bayes
PANAS	Positive and Negative Affect Schedule
PCA	Principal Component Analysis
POMS	Profile of Mood States
PPG	Photoplethysmography
PSD	Power Spectral Density
RF	Random Forest
RNN	Recurrent Neural Network
RMSSD	Root Mean Square of Successive Differences
SAM	Self-Assessment Manikin
SC	Skin Conductivity
SCR	Skin Conductance Responses

List of Abbreviations

SCL	Skin Conductance Level
SKT	Skin Temperature
SR	Skin Resistance
STAI	State-Trait Anxiety Inventory
SVM	Support Vector Machine
SWFA	Scaled-Weighted Federated Averaging
TFF	TensorFlow Federated
TINN	Triangular Interpolation of NN Interval
TSST	Trier Social Stress Test
V	Valence
VA	Valence-Arousal
VAD	Valence-Arousal-Dominance
VDH	Variation Data Heterogeneity
VFL	Vertical Federated Learning

Chapter 1

Introduction

Recognizing emotions is a technique for determining a human being's current state of feelings and thoughts, which varies substantially. The use of technology to assist individuals in identifying emotions is a relatively more explored field of study. Emotions are essential to people's lives because they influence their feelings and decision-making. The two most common types of emotions that humans display are Positive and Negative emotions [196]. The *Positive emotions* support an individual's well-being, emotional stability, and productivity in day-to-day activities. Conversely, *negative emotions* can result in stress [243], anxiety [242], health issues, and other problems, and in severe cases, they can even lead to suicide.

Psychologists primarily represent emotions using two approaches: (1) classify emotions on discrete levels, defining each as high, low, and neutral [219], and (2) define emotions into groups based on their polarity (positive or negative - named as Valence) and intensity (arousal levels - named as Arousal) [126, 286]. Valence (V) and Arousal (A) are the two dimensions of the 2-dimensional (2D-VA) model of emotion [126] in which Valence indicates whether the emotion is positive or negative, and Arousal indicates the intensity levels, i.e., low or high. This 2D model of emotion was upgraded into a 3-Dimensional (3D-VAD) model of emotion [33, 127], which included an addition of the Dominance (D) dimension to describe how submissive and dominant emotions are for an individual. Each method contributes to communicating a certain aspect of human emotion while putting forth perspectives on how emotions are portrayed and experienced by people. These aid in assessing an individual's emotional condition at any moment.

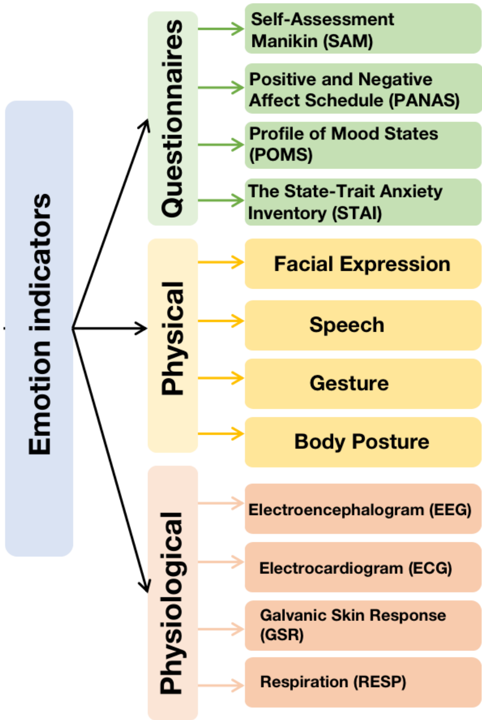


Figure 1.1: Branching representation for the emotion indicators

There are different indicators for identifying emotions as represented in Figure 1.1 and described below:

1. **Psychological Subjective Questionnaires**: It involves self-report measures where individuals describe their feelings, experiences, or reactions to various stimuli via common tests, non-verbal pictorial assessment techniques, and questionnaires (as shown in Figure 1.2). These include Self-Assessment Manikins (SAM) [119], Positive and Negative Affect Schedule (PANAS) [71], Profile of Mood States (POMS) [220], The State-Trait Anxiety Inventory (STAI) [63]. ***Disadvantages:*** These indicators depend on memory, which can be unreliable and influenced by recent events. Individuals might only sometimes be honest or may alter their responses due to social desirability bias or a lack of self-awareness. Additionally, a person's current mood can impact their answers, compromising the data's accuracy [130, 200, 259].

2. **Physical Indicators**: It includes Speech [209], Gestures, Facial Expressions [289] and Postures (as shown in Figure 1.2). ***Disadvantages:*** While these indicators are relatively easy to gather, they are not always reliable. Physical

Figure 1.2: Sample indicators for emotion recognition.

indicators might not capture the full spectrum of emotions, particularly more subtle or complex emotions. Physical behaviours can be influenced by context and social norms, leading to misinterpretation. People can mask their genuine feelings by manipulating body indicators like facial expressions or voices, especially in social situations. For instance, someone might smile even if experiencing negative emotions [226].

3. **Physiological Indicators**: It includes physiological signals for Electroencephalogram (EEG) [294], Electrocardiogram (ECG) [89], Galvanic Skin Response (GSR) [152], Heart Rate (HR) [57], Blood Volume Pulse (BVP) [196] and Respiration (RESP) [79]. ***Disadvantages:*** Recording these signals requires a little complicated installation and maintenance of equipment susceptible to movement artefacts and noises. ***Advantages:*** These indicators can directly measure the body's response (as shown in Figure 1.2) to emotional stimuli, offering insights into the physiological processes underlying emotions. Physiological indicators allow for continuous monitoring of emotional states, providing real-time data. It benefits populations that cannot self-report, such as infants, non-verbal individuals, or people with communication impairments. These physiological signals reflect sensory-motor expressions and help predict more accurate emotional shifts. Using single physiological signals leads to inaccurate emotion classification due to noises and artefacts in them. [77, 118, 169]. However, combining these signals in a multimodal approach leads to more accurate and reliable classifications [50, 92, 109, 135].

This thesis focuses on utilizing the physiological indicators for recognizing emotions. It examines the effectiveness of the physiological signals EEG, ECG, GSR and RESP (single and multi-modal) to accurately identify emotions.

1.1 Motivation and Background

1.1.1 Why Emotion Recognition?

As emotions play a vital role in a human being's life, they influence how to see and understand things related to daily life scenarios. Recognizing emotions via different physical [289],[209] and physiological[294],[89], [152],[57] parameters in the past research concludes that physical parameters like facial expression, speech, and text can be fake, but physiological signals like EEG [294], ECG [89], HR [57], GSR [152] and others can map the correct and accurate state of emotions. Emotion recognition holds the potential to transform healthcare [39, 74, 166, 298] (improving mental health monitoring), entertainment [202] (revolutionizing user experience), and education [183, 260] (enhancing learning environments) industries to understand and respond to human emotions in real-time.

1.1.2 Background for Automated Emotion Recognition using Physiological Signals

Automated Emotion Recognition Systems (ERS) use Machine Learning (ML) and Deep Learning (DL) models to recognize human emotions via identifying patterns from physiological responses, which can improve health and decision-making [74, 183, 298]. Automated ERS using ML and DL models requires physiological signal data as input for classifying different emotions by evaluating behavioural tendencies, physiological reactions, motor expressions, cognitive assessments, and subjective feelings. Figure 1.3 illustrates the process of automated ERS using physiological signals, such as heart rate, skin conductance, and brain activity, through a multi-step process [120].

Initially, physiological signals are collected using wearable sensors (described later in Chapter 2), which undergo preprocessing to remove noise and artefacts. It follows

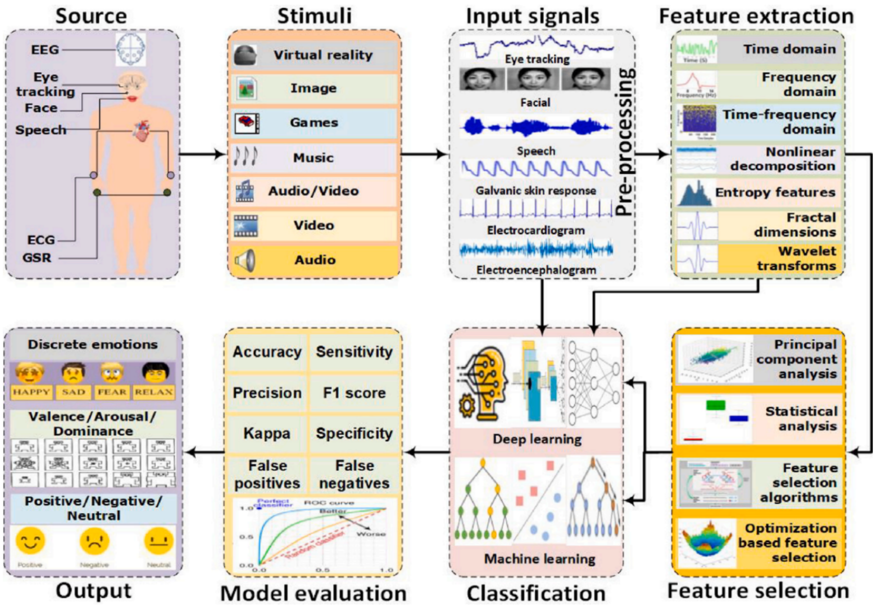


Figure 1.3: Methodology for automated Emotion Recognition System (ERS) [244].

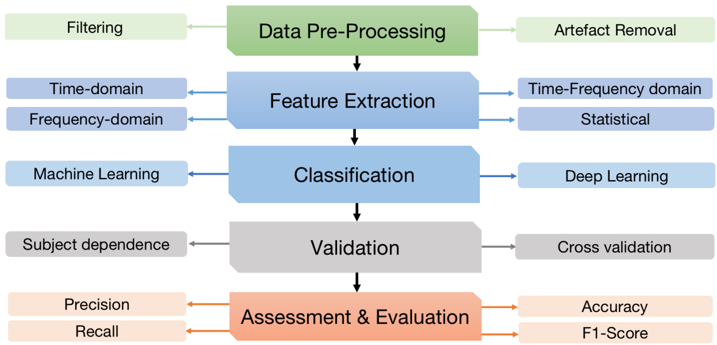


Figure 1.4: Workflow of steps for Emotion Recognition System (ERS)

feature extraction, where relevant features such as statistical metrics, frequency domain features, and time-domain features are derived from the signals (as shown in Figure 1.4). Then, feature selection techniques, such as Principal Component Analysis (PCA) [47], are implemented to help identify the most informative features. For model training, both ML and DL models like Support Vector Machines (SVM) [159, 193] and Random Forests [82, 185, 194], Convolutional Neural Networks (CNN) [101, 143, 173], and Long Short-Term Memory (LSTM) networks [31, 98, 267, 287], are employed to learn the patterns associated with different physiological signal inputs reflecting different emotions from labelled datasets. These models are then validated using cross-validation techniques and performance metrics like accuracy, precision, recall, and F1-score. After validation, the trained models can be deployed in real-time systems for continuous sensor data processing to classify emotions in real time. This comprehensive pipeline leverages the power of ML and DL to enable accurate and robust emotion recognition using physiological signals and can enhance model performance and adaptability. It has wide applications in mental health monitoring, decision making and human-computer interaction with user feedback.

1.1.3 What are the Privacy Concerns in Automated in ERS?

The automated ERS uses traditional ML and DL methods, which require complete raw data for training the model, resulting in a significant loss of privacy protection for a human's sensitive physiological information [36, 45, 100]. These methods require access to vast amounts of raw physiological data for effective model training, making the ERS more vulnerable to unauthorized data access and misuse risks. Directly feeding physiological signals into the model exposes sensitive data to third parties and potential attackers, resulting in a substantial loss of privacy. Safeguarding this sensitive physiological data has become a major concern in human emotion recognition, particularly within certain environments. Conventional ML and DL models for ERS allow multiple users to access each other's emotional states [28, 82, 145, 146, 154, 268]. Given humans' diverse range of daily emotions, individuals may prefer not to share these emotions with others or an audience [56]. For instance, individuals facing health challenges and navigating unstable emotional states might choose not to express themselves openly [163]. Several potential challenges emerge as a result of privacy concerns with automated ERS that utilize sensitive physiological signals, such as data leakage incidents recently happened with the automated ERS [95, 164, 301], In 2019 [25], biometric data and records got leaked. External hacking attempts leading to data breaches result in the unauthorized access of sensitive physiological data utilized by the ERS [123, 270].

1.1.4 Why Federated Learning (FL) Paradigm?

Federated Learning (FL) is one paradigm that cares for data privacy, creating a decentralized environment [39, 99, 124]. It is a specialized architecture in which the global model takes weights of the raw data for secure aggregation and

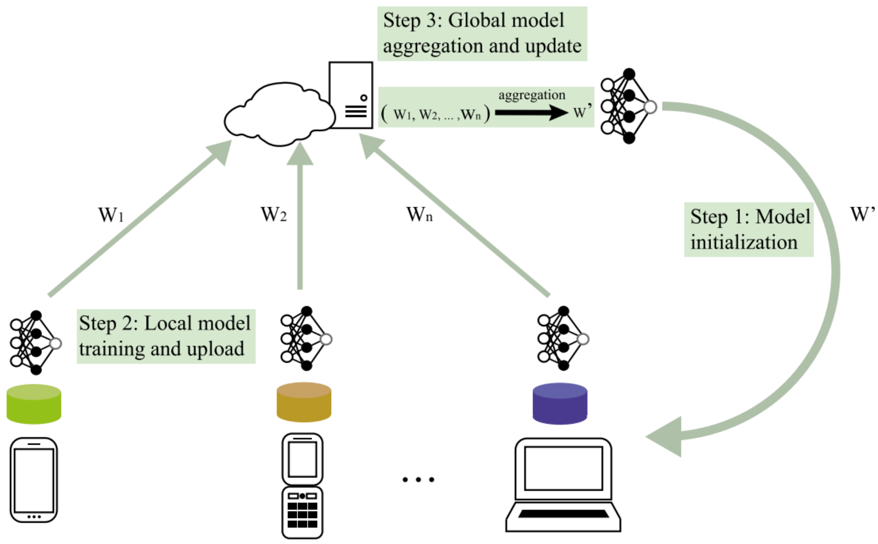


Figure 1.5: Schematic diagram of the Federated Learning environment.

analysis [136, 249]. In 2017, Google AI proposed FL [1]), a new machine learning technology [137, 249]. Figure 1.5 illustrates how several remote devices are allowed for training models without moving or releasing complete datasets using Federated Learning [136]. Unlike traditional learning techniques, FL uses weights of the raw data to train the models instead of users' complete raw data [214, 249]. Due to the relevance of what it brings to the table for machine learning in a hyper-connected world, it is fruitful for emotional recognition. It is a distributed algorithm for training data (stored locally) that does not collect complete raw data from the local servers [69]. Hence, it preserves the privacy of sensitive input data (physiological information) [36, 100].

1.2 Research Gaps

1.2.1 Research Gap 1: Lack of Privacy

Automated Emotion Recognition Systems using traditional machine learning and deep learning methods for classifying emotions fail to preserve the data privacy of users' sensitive physiological information during model training [36, 124]. This vulnerability allows data attackers to access and exploit personal biometrics and physiological information [39, 124, 270].

[1] `https://www.tensorflow.org/federated`

1.2.2 Research Gap 2: Limited FL for ERS

The FL paradigm for emotion recognition based on physiological signals is an unexplored area of research. To the best of our knowledge, very limited research papers have been published using FL for ERS based on physiological signals, and they all have implemented only one aggregation algorithm of FL (i.e. FedAvg) [36, 100, 149].

1.2.3 Research Gap 3: Less Emotion Dimensions

The existing FL works for ERS fail to recognize complex emotions and feelings, as they employ a two-dimensional emotion model (2D-VA) that only encompasses positive-negative emotions with arousal levels [36, 67, 100, 165, 248]. While the 2D-VA emotion model can easily differentiate between positive and negative emotions. However, recognizing identical emotions within the 2D emotion space remains challenging [33]. Therefore, there is a need to implement an updated model (3D-VAD) with more dimensions to capture a wider range of human emotions.

1.2.4 Research Gap 4: Restricted Modality

The existing FL works for ERS classifying emotions uses only one of the single physiological signals EEG, EDA, ECG, or HR [98, 251, 252, 264, 287]. However, using a single physiological signal is unreliable as artefacts and noises in the signals distort the signals' features, which are input for training the ML models to classify emotions. Noises and artefacts in the signal result in unjustified and inaccurate recognition, as they lead the model to misinterpret the underlying emotional state of the subject. [77, 118, 169].

1.2.5 Research Gap 5: No handling of Data Heterogeneity

The existing FL works for ERS do not address the data heterogeneity in the federated environment [36, 67, 100, 165]. The data heterogeneity can significantly impact the performance and generalization of federated models for ERS in real-world scenarios and must be addressed.

1.3 Problem Statement

This thesis addresses the paramount problem of privacy preservation of sensitive multi-modal physiological signals in an automated emotion recognition system using the Federated learning paradigm. Apart from privacy, this thesis reduces two types of data heterogeneities: (1) Variation Data Heterogeneity (VDH) occurring in the multi-modal physiological signal data input while combing multiple physiological signals via feature-level fusion, and (2) Imbalanced Data Heterogeneity (IDH) occurring in a federated environment due to the different data distribution at the client end. Additionally, it works on expanding emotion dimensions using a 3D-VAD model of emotions to explore a wide range of emotions via the three emotion dimensions: Arousal, Valence, and Dominance, which is lacking in the existing works of FL for ERS.

1.4 Research Objectives and Contribution of the Thesis

The main objective of the thesis is to preserve the privacy of the Emotion Recognition System based on physiological signals. The thesis achieves the following four objectives:

- **Objective 1: Literature Review -** To perform a comprehensive literature review on the Emotion Recognition System based on Physiological Signals using a Federated Paradigm.

  Contribution: This thesis conducts an exhaustive, comprehensive literature review to examine the strengths and weaknesses of various machine learning and deep learning methods for classification models for automated emotion recognition systems. This literature review adds the concept of federated learning and its usage for emotion recognition. (Completed by **Journal 1**).

- **Objective 2: Multi-modal Privacy Architecture -** To propose privacy preserved multi-modal architecture for emotion recognition using Federated Learning.

  Contribution: This thesis proposes two multi-modal FL architectures for emotion recognition: (1) *Federated Learning-based Multi-modal Emotion Recognition System* (***F-MERS***), and (2) *Attention-based Federated Learning* *for Emotion recognition using Multi-modal Physiological data* (***AFLEMP***). The thesis validates both the proposed F-MERS and AFLEMP for three different emotion datasets: A dataset for Multi-modal research of affect, personality traits and mood on Individuals and GrOupS (AMIGOS) [125], Database for Emotion Analysis using Physiological signals (DEAP) [233], and A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices (DREAMER) [135]. To prove the generalizability and robustness of the proposed architectures (F-MERS and AFLEMP), the thesis computes their scalability, communication computation and performance measures. (Completed by **Journal 2, 3**).

- **Objective 3: FL Algorithms -** To study and implement aggregation algorithms of FL for emotion recognition systems.

  Contribution: This thesis proposes a novel *Scaled Weighted Federated Averaging* ***(SWFA)*** algorithm for handling IDH in a federated ERS for enhanced and better averaging at the server end. It also explores and implements other existing aggregation algorithms of FL - FedAvg, FedBoost, FedPer, and Dynamic Weighted Federated Averaging (DWFA) for ERS. To prove the efficiency and robustness of the proposed SWFA, the thesis implements it for two emotion datasets (AMIGOS [125] and DREAMER [135]) and compares their averaging time and classification accuracy. (Completed by **Journal 2**).

- **Objective 4: Increased Emotion Dimensions -** To study and implement more dimensions of emotional states with their intensities while classifying emotions in an FL environment for ERS.

  Contribution: This thesis proposes the federated F-MERS architecture with the three-dimensional model of emotion (3D-VAD) for binary classification of emotions using Arousal, Valence, and Dominance individually. It proposes the AFLEMP architecture with the 3D-VAD for the octal classification of emotions for a wider range of emotions (eight different emotions), using Arousal, Valence, and Dominance together. (Completed by **Journal 2**).

1.5 Significance of the Thesis

Automated ERS are pivotal in discerning an individual's emotional states, indicating overall health [122, 135, 233, 287]. These systems leverage physiological signals such as heart rate, blood pressure, and skin temperature collected via smart wearables like smartphones and smart bands, forming an integral part of innovative healthcare systems [122, 189, 274]. The thesis explores different techniques to improve the privacy of emotion recognition systems using a federated learning approach and multi-modal physiological signals. It explains how emotion recognition models are trained across multiple decentralized devices while keeping the data local and offers a robust solution for securely handling personal sensitive data.

The research contributes to the field of FL by exploring its application in emotion recognition based on physiological signals. It encompasses a broad spectrum of emotional states, addressing the limitations of traditional automated emotion recognition systems that rely on conventional machine learning techniques. The findings have significant implications across several domains. Improved emotion recognition can lead to more precise assessments and timely interventions in mental health [39, 298]. The technology can enhance user experience in human-computer interaction by enabling systems to better interpret and respond to emotional cues. Privacy preserves emotional healthcare and facilitates more adaptive and responsive healthcare solutions.

1.6 Thesis Overview

This section outlines the structure of the thesis, providing an overview of the content and purpose of each chapter. **Chapter 2** gives the background for Emotions and Physiological Signals. It details the different emotion models for mapping emotions using discrete and dimensional structures. This chapter details four physiological signals, their significance for emotions, and how they can be recorded using wearable biosensor devices. The limitations and benefits of each physiological signal are described in this chapter, specifying the need and importance of multi-modality. It describes the complete steps for data processing techniques for physiological signals, which include filtering, artefact removal, and feature extraction.

Chapter 3 discusses the openly accessible emotion datasets with the availability of physiological signals. It provides the data descriptions and pre-processing for the three emotion datasets chosen for evaluating the architectures proposed by the thesis.

Chapter 4 presents the automated ERS using different ML and DL models, taking the physiological signals as input for training the models for emotion classification and reports their results. It describes the contribution of different ML and DL models for emotion classification. This chapter also discusses the limitations and privacy concerns in the traditional ML and DL models for emotion recognition.

Chapter 5 presents a detailed introduction to FL, including its workflow, types of architecture, approaches, aggregation algorithms, tools and its importance in handling sensitive data. It presents the literature review of existing FL state-of-the-art for emotion recognition using physiological signals and reports their limitations. Additionally, it outlines the different evaluation measures required to assess the FL architectures proposed in this thesis.

Chapter 6 proposes a novel federated learning architecture: **F**L-based **M**ulti-modal **E**motion **R**ecognition **S**ystem (**F-MERS**) for emotion recognition to solve the problem of lack of privacy preservation for sensitive physiological signals. This chapter presents the methodology for the proposed F-MERS and assesses its classification performance based on accuracy, precision, recall, and F-1 score. It also measures its scalability and communication computation in terms of time and discusses the experimental results.

Chapter 7 proposes a novel architecture: **A**ttention-based **F**ederated **L**earning for **E**motion recognition using **M**ulti-modal **P**hysiological data (**AFLEMP**) to address the Variation Data Heterogeneity (VDH). The issue of VDH occurs due to variations in the multi-modal physiological input data stream. The proposed AFLEMP uses different attention mechanisms to reduce the VDH. This chapter presents the methodology for the proposed AFLEMP, assesses its classification performance, scalability, convergence speed and communication computation, and discusses the experimental results.

Chapter 8 addresses the problem of Imbalanced Data Heterogeneity (IDH) occurring at the client end in a federated environment. It proposes a novel **S**caled **W**eighted **F**ederated **A**veraging (**SWFA**) algorithm as a solution to IDH by implementing AFLEMP with SWFA. This chapter presents the methodology for the proposed AFLEMP with SWFA, evaluates and contrasts it with other FL Algorithms for ERS in terms of classification performance, convergence speed, and averaging time and discusses its experimental results.

Lastly, **Chapter 9** summarizes the conclusions inferred from this research work, highlights the ethical considerations, potential future work, and the social applications in this area.

Chapter 2

Background for Emotions and Physiological Signals

This chapter provides a comprehensive overview of emotions and the physiological signals essential for emotion recognition. It explores the various theoretical models psychologists use to categorize and understand emotions, offering insights into discrete and dimensional models such as arousal and valence. These models serve as the basis for interpreting human emotional experiences and their corresponding physiological responses, including changes in heart rate, skin conductance, and brain activity. The chapter delves into the significance of the physiological signals, explaining how they offer insights into the emotional state of individuals and how they are measured using modern biosensor technologies.

2.1 Emotion Models

Psychologists typically model emotions in two ways. One approach divides emotions into categories such as arousal and valence [33, 127]. Another approach classifies emotions on three levels: high, low, and neutral [65, 188, 219, 296]. Both models offer insights into how humans represent and perceive emotions, highlighting different aspects of human emotion.

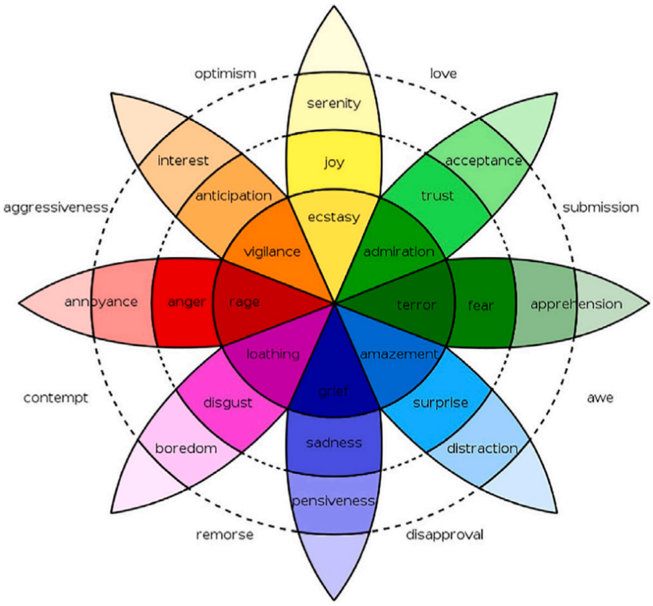


Figure 2.1: Plutchik's model of emotions [219]

2.1.1 Discrete Emotion Models

Human emotions can be described using terms such as pain, anger, fear, happiness, and many others [296]. These emotions are innate and do not require learning, manifesting differently in various environments and revealing unique individual feelings.

Many psychologists have explored human emotions and responses in different contexts, classifying them as discrete emotions. Cicero and Graver [65] suggested that emotions are inherently natural and universally experienced across cultures. The authors grouped emotions into four primary categories: pain, fear, pleasure, and lust. Ekman [188] described emotions as clearly defined, quantifiable, physically connected, originating from past physiological and communicative needs. The author named anger, happiness, fear, sadness, disgust, and surprise the six fundamental emotions. These emotions produced physiological reactions that functioned as alerts, occasionally telling the difference between life and death. Figure 2.1 illustrates Plutchik [219]'s model of emotions in a wheel consisting of eight primary emotions: anger, trust, joy, sadness, fear, surprise, disgust, and anticipation. This model shows varying intensity, with powerful emotions in the centre and weaker feelings on the wheel's edges.

However, this emotion model cannot capture complicated emotions such as liking, disliking, or hatred. In order to work over these constraints, the idea of a continuous multi-dimensional space model is proposed.

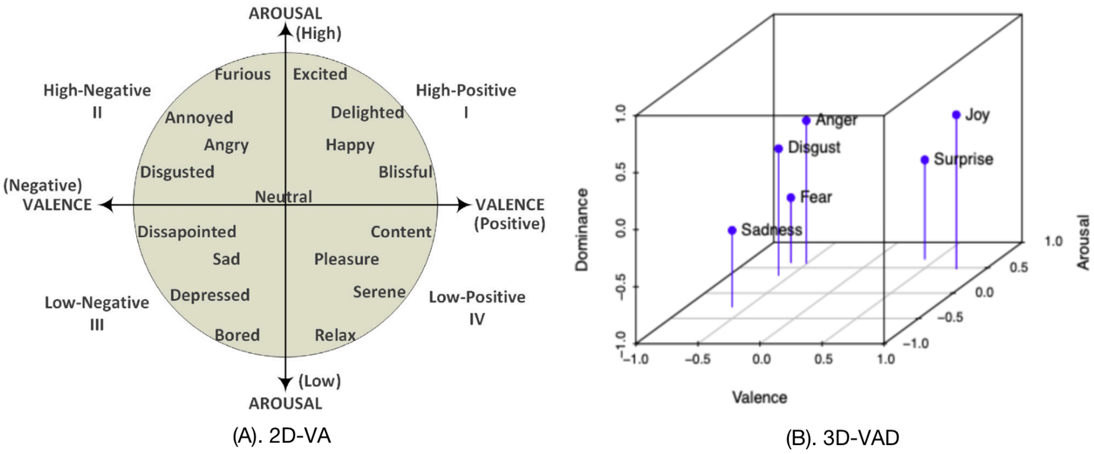


Figure 2.2: Models of emotions: (A) 2D-VA [286], (B) 3D-VAD [184]

2.1.2 Continuous Emotion Models

Continuous emotion models assess emotions along a single axis, simplifying comparison and classification [126]. However, this approach addresses two main difficulties. First, it can express correlations between different emotional states, such as admiration versus trust, grief versus sadness, and quantify specific conditions, such as very sad versus not sad [91]. Second, it accounts for varying intensities in emotions with similar descriptions. For instance, joy can range from a little to a lot of happiness. Consequently, psychologists have developed multi-dimensional emotion space models, including 2D and 3D models.

2.1.2.1 2-Dimensional Valence-Arousal (2D-VA) Emotion Model

Russell introduced the circumplex model of emotions in 1980. This model uses valence and arousal axes to create a two-dimensional space, representing affective states as discrete points [126]. The valence axis indicates the positivity or negativity of the current state. In emotion recognition, valence is typically measured on a scale that ranges from highly negative to highly positive. In contrast, the arousal axis grades the condition based on the arousal level, such as how energized or enervated one feels. Figure 2.2 (A) illustrates Russell's circumplex model schematically. The arousal-valence model is widely used due to its ease of integration in assessing emotions through questionnaires and its simplicity in training machine learning algorithms, resulting in effective outcomes [221, 233].

2.1.2.2 3-Dimensional Valence-Arousal-Dominance (3D-VAD) Emotion Model

It will be difficult to identify similar emotions in a two-dimensional emotion space, despite its ease of discriminating between positive and negative, good and bad emotions. For instance, anger and fear are both in the negative valence and high arousal range. To distinguish such emotions, a new dimension, Dominance is introduced by Mehrabian and Russel [33, 127]. Dominance refers to the degree of control or power one feels in response to the emotion. Anger is associated with high dominance, where a person feels in control or empowered such as taking action or confronting a situation. In contrast, fear is often linked with low dominance, where a person feels powerless or threatened such as feeling overwhelmed or wanting to escape. Figure 2.2 (B) shows Mehrabian and Russel's model of emotions, which expanded the 2-dimensional model to 3-dimensional by adding a Dominance dimension.

2.2 Emotion Annotation and Ground Truth

Emotion annotation involves labelling or tagging the stimuli data with specific emotional states from which emotions are to be induced. This process is essential for training and evaluating automated emotion recognition systems. Annotators can include humans [35, 108] and annotation tools [139, 254] may assist in labelling data based on pre-trained models or algorithms, though human oversight is often needed to ensure accuracy [139, 277]. They review the data and assign emotion labels based on predefined categories (e.g., happy, sad, angry, and others) by listening to audio recordings or watching videos used for emotion elicitation stimuli [35, 119, 233]. There are two types of annotation schemes. One involves discrete emotions, assigned based on specific, predefined emotions (such as joy, fear, surprise), and the other is with dimensional models, annotated based on dimensions such as valence (pleasantness vs. unpleasantness), arousal (activation vs. deactivation), and dominance (control vs. lack of control) [140].

Ground truth refers to the accurate and reliable reference data used as a benchmark for evaluating the performance of emotion recognition systems [29, 78, 140]. It represents the actual emotional states or labels against which model predictions are

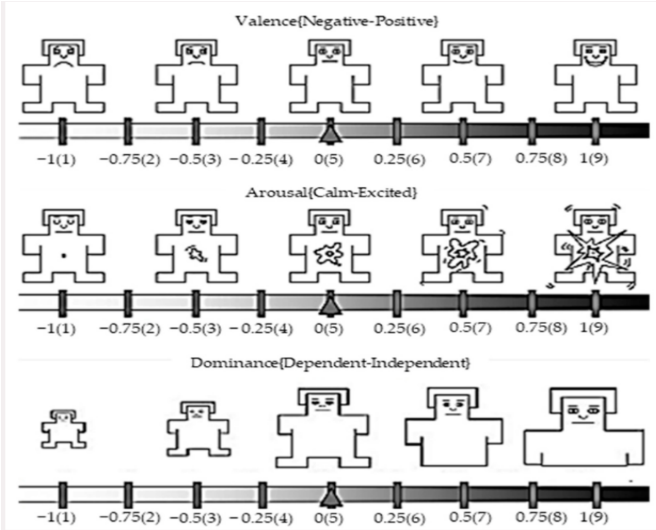


Figure 2.3: Self-Assessment Manikins (SAM) with ratings for emotion recognition.

assessed. Determining the ground truth for emotions is challenging due to the lack of clear definitions. However, the most effective approach for annotating emotions experienced during experiments is through subjective ratings of emotional trials or self-reports. The most widely used subjective rating tool is Self-Assessment Manikin (SAM) [119]. Figure 2.3 shows the SAM ratings on a scale of 1 to 9 for arousal, valence, and dominance using the pictorial representation of manikins. It has been effectively used to measure emotional responses across various scenarios, including reactions to images, sounds, videos, and other types of stimuli. This thesis uses the datasets that recorded subjects' self-ratings using SAM [119].

2.2.1 Creation of Emotion Classification Labels

Using the self-ratings discussed in the previous section provided by the subjects while watching the stimuli, a threshold value is set to categorize the emotional responses into distinct classes (or labels). For each dimension of the 3D-VAD model (Valence, Arousal, Dominance), a threshold is determined based on the distribution of ratings. For example, on a scale of 1-9, a threshold of 4.5 can be used to quantify the low (<4.5) and high (>4.5) classes.

- **Binary Classification:** For binary classification, the emotional labels are divided into two categories (low and high) for each dimension as follows (shown in Table 2.1), and described below:

Table 2.1: Mapping of emotions using valence, arousal, and dominance individually (from 3D-VAD).

Valence	
Low	**High**
Sorrow, Anger, Fear, Disgust	Happiness, Calm, Surprise, Excitement
Arousal	
Low	**High**
Sorrow, Calm, Fear, Happiness	Disgust, Anger, Surprise, Excitement
Dominance	
Low	**High**
Sorrow, Disgust, Happiness, Surprise	Fear, Anger, Calm, Excitement

1. *Valence:* Positive emotions as High and Negative emotions as Low based on whether the rating is above or below the set threshold. For example, rating 3 on a scale of 1-9 will be referred to as a low valence and can imply one of these emotions: Sorrow, Fear, Anger and Disgust.

2. *Arousal:* High arousal level emotions as High activation and Low arousal level emotions as Low activation based on the arousal rating relative to the threshold. For example, rating 3 on a scale of 1-9 will be referred to as low arousal and can imply one of these emotions: Sorrow, Calm, Fear or Happiness.

3. *Dominance:* High control and Low control are classified based on the dominance rating in comparison to the threshold. For example, rating 3 on a scale of 1-9 will be referred to as low dominance and can imply one of these emotions: Sorrow, Happiness, Surprise and Disgust.

- **Octal Classification:** In octal classification, the emotional states are classified into eight categories formed by all possible combinations of the three dimensions (as shown in Table 2.2):

  1. *Low valence, Low arousal, Low dominance* (**LVLALD**): Represents negative emotions with low intensity and little to no control such as Sorrow.
  2. *Low valence, Low arousal, High dominance* (**LVLAHD**): Represents negative emotions with low intensity and strong control such as Fear.
  3. *Low valence, High arousal, Low dominance* (**LVHALD**): Represents negative emotions with high intensity but little to no control such as Disgust.

Table 2.2: Mapping of emotions using valence, arousal, and dominance together (from 3D-VAD)

Valence	**Arousal**	**Dominance**	**Emotions**
Low	Low	Low	Sorrow
Low	Low	High	Fear
Low	High	Low	Disgust
Low	High	High	Anger
High	Low	Low	Happiness
High	Low	High	Calm
High	High	Low	Surprise
High	High	High	Excitement

4. *Low valence, High arousal, High dominance* (**LVHAHD**): Represents negative emotions with high intensity and strong control such as Anger.
5. *High valence, Low arousal, Low dominance* (**HVLALD**): Represents positive emotions with low intensity but little to no control such as Happiness.
6. *High valence, Low arousal, High dominance* (**HVLAHD**): Represents positive emotions with low intensity and strong control such as Calm.
7. *High valence, High arousal, Low dominance* (**HVHALD**): Represents positive emotions with high intensity but little to no control such as Surprise.
8. *High valence, High arousal, High dominance* (**HVHAHD**): Represents positive emotions with high intensity and strong control, such as excitement.

The previous work of FL for ERS with physiological signals used valence and arousal from a 2D model of emotions to recognize emotions [36, 67, 100]. They lacked the presence of dominance in their proposed frameworks for ERS. Incorporating dominance with valence and arousal in emotion recognition is required for enhanced and diverse emotional representation [172] and improved accuracy in emotion recognition [172]. To overcome this research gap, this thesis used valence, arousal, and dominance together, from the 3D-VAD model of emotion for the proposed AFLEMP architecture, for which the results are given in later Chapter 7.

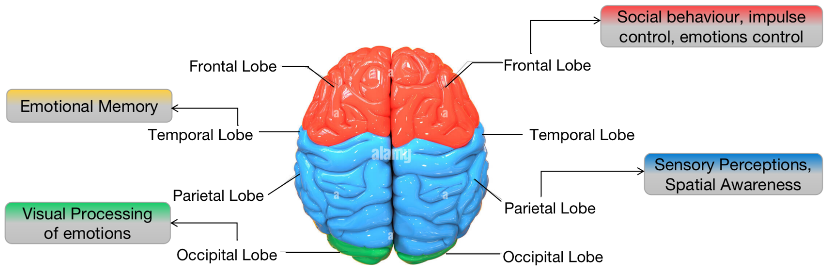


Figure 2.4: Brain lobe's functions for emotional activities.

2.3 Physiological Signals and Emotions

Physiological signals are the biochemical signals produced in response to external events. These are used for emotion recognition as they directly map an individual's current state of emotions. They require considerable signal pre-processing and assessment. The following are the most widely used physiological signals for recognizing emotions.

2.3.1 Electroencephalographic (EEG) Signals

Electroencephalographic (EEG) signals are the electric impulses recorded to assess brain functions [90, 294]. These signals are crucial for understanding brain activities, as the brain controls all human emotions, including physical movement, sensory processing, language and communication, memory, and emotions [237, 294]. EEG has been widely used to investigate brain neuronal functions and human emotion with high time resolution and does not cause radiation exposure [237, 299].

The human brain serves as a central hub where different emotions and feelings are generated, processed, and regulated. Different regions of the brain work together to interpret external stimuli and internal states, translating them into emotional experiences. The human brain is divided into four regions called as lobes: Frontal, Temporal, Parietal, and Occipital [97, 123, 262]. Figure 2.4 shows the different brain lobes with their emotion-relative activities. The frontal lobe, particularly the prefrontal cortex, manages emotions, impulse control, and social behaviour. The temporal lobe processes emotions like fear and pleasure and aids in emotional memory formation. While the parietal lobe focuses on sensory perception and spatial awareness, it also helps interpret emotional expressions and body language. The occipital lobe handles visual processing, indirectly influencing emotional responses to visual stimuli.

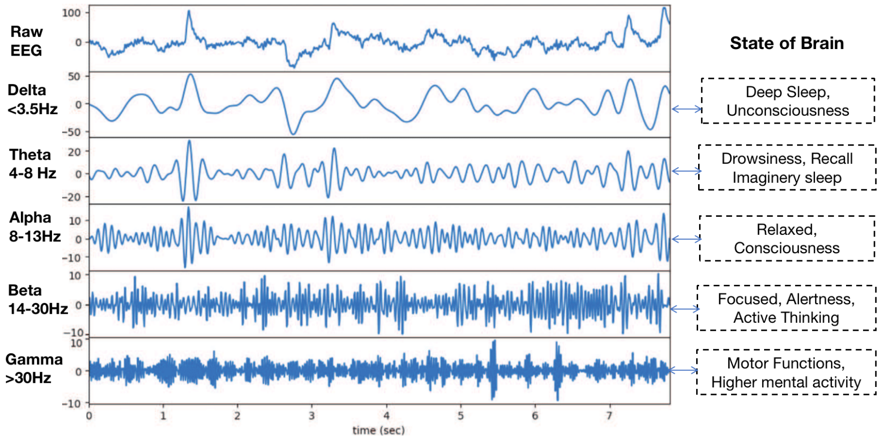


Figure 2.5: Frequency band distribution of raw EEG signal and its impacts on brain

Table 2.3: Wearable devices that use physiological signals to track emotions [238].

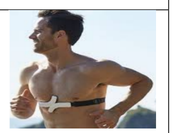
Physiological Signal	Wearable Device	Example
EEG	Emotiv EPOC [5], NeuroSky MindWave [9], Emotiv Insight [116], DSI-24 [18], Neuphony Headband [19], CGX Quick-20r V2 [17]	
ECG	VitalPatch [16], Polar H10 [11], Garmin HRM-DUAL [8]	
GSR	EMPATICA E4 [6], RING [13], SHIMMER3 [15]	
RESP	EMPATICA E4 [6], VitalPatch [16], Samsung Gear Live Watch [20], Garmin HRM-DUAL [8]	

Figure 2.5 presents the raw EEG signal classification into five types of frequency waves: Alpha, Beta, Theta, Delta, and Gamma. Delta waves, which have a frequency of less than 3.5 Hz, are typically observed during sleep. Theta waves, with frequencies between 4 to 8 Hz, indicate drowsiness or the onset of sleep [171, 261, 278]. Alpha waves, ranging from 8 to 13 Hz, indicates relaxed states of a person. Beta waves range from 14 to 30 Hz and are associated with nervousness or active thinking. Gamma waves have frequencies above 30 Hz, indicates heightened alertness and emotional arousal.

Table 2.3 lists the different wearable devices for recording physiological signal responses, including several EEG headsets. For recording EEG, there are various EEG devices available, including Emotiv EPOC [5], NeuroSky MindWave [9], Emotiv

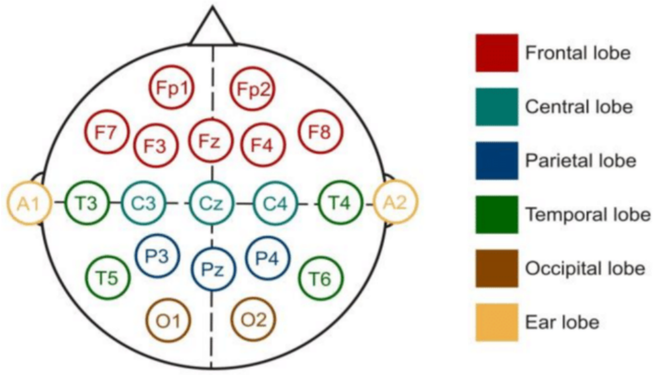


Figure 2.6: 19 Channel EEG headset montage mapping over brain regions.

Insight [116], DSI-24 [18], Neuphony Headband [19], and CGX Quick-20r V2 [17] which are placed over the head with various number of electrodes or channels ranging from 1 to 32 channels which touch to the scalp to record the signals sent by the brain to measure emotions. Figure 2.6 shows a montage to present a 19-channel EEG headset placed overhead, covering various brain regions. The frontal lobe is covered with Fp1, Fp2, F7, F3, Fz, F4 and F8, the central lobe is covered with C3, Cz and C4, the temporal lobe is covered with T3, T5, T6 and T4, the parietal lobe is covered with P3, Pz and P4, and the occipital lobe is covered with O1 and O2 [131].

2.3.1.1 Significance of EEG for Emotions

EEG measures electrical activity in the brain and is valuable for understanding emotional states. For instance, happiness is traced with increased alpha wave activity in the left frontal region of the brain, indicating a state of relaxed alertness and positive emotional processing [132, 232, 285]. Conversely, sadness corresponds to heightened activity in the right frontal region of the brain, reflecting emotional distress and negative affect [232]. Fear is witnessed with increased beta waves in the frontal and parietal regions of the brain, linked to heightened alertness and anxiety [162, 232]. During relaxation, there is an increase in alpha waves, particularly in the occipital and parietal regions of the brain, indicative of a calm and restful state [232, 285].

2.3.1.2 Features from EEG signals

- **Time-Domain:** Time domain features focus on the amplitude and temporal characteristics of the signal. It includes the following features for EEG [216]:
  1. ***Mean:*** The average value of the EEG signal over time.
  2. ***Median:*** The middle value of the EEG signal over time.
  3. ***Variance:*** Measures the variation of the EEG signal around the mean.
  4. ***Standard Deviation:*** The square root of the variance indicates the signal's dispersion.
  5. ***Hjorth Parameters:*** These capture essential features of the signal, including frequency content, signal strength and complexity.
     - *Activity*: It represents the variance of the signal. It quantifies the amount of information or the signal's power as Low or High [40, 223].
     - *Mobility:* The square root of the first derivative's variance is divided by the signal's variance. It essentially measures the signal's frequency content [40, 223].
     - *Complexity:* The ratio of the mobility of the first derivative of the signal to the mobility of the signal itself. It measures the variation in frequency content over time, indicating the intricacy of the signal's shape [40, 223].
  6. **Fractal Dimension (FD):** It is a feature that quantifies the complexity and self-similarity of EEG signals, particularly useful for analyzing non-linear and irregular patterns found in the EEG signals.

- **Frequency-Domain:** The time-domain signal is converted into the frequency domain using Fast Fourier Transform (FFT). The frequency domain features are derived via the signal's power spectrum, reflecting the distribution of power across different frequency bands. It includes the following features for EEG signal:
  1. ***Power Spectral Density (PSD):*** Gives a measure of the power distribution over frequency [179, 210, 233] and is calculated using the Welch method [233].

Table 2.4: Description of EEG signal features in time and frequency domain.

Domain	Features	Description
Time	Mean, Median, Standard Deviation, Variance	Average value, median, standard deviation, variance, change between consecutive values the EEG signal.
	Hjorth Activity	This parameter measures the overall energy or power of EEG signal.
	Hjorth Mobility	This parameter measures the rate of change of frequency content of EEG signal.
	Hjorth Complexity	This parameter is a combination of Hjorth Activity and Mobility measuring the complexity of EEG signal.
	Fractal Dimension	It is a feature that quantifies the complexity and self-similarity of EEG signal.
Frequency	Power Spectral Density	Measure of power distribution over frequency of the EEG signal, calculated using Welch method.
	Spectral Entropy	Measure of the distribution or randomness of the power spectrum of a signal
	Bandpower (alpha, beta, theta, delta)	This feature gives the power within specific frequency bands.

2. ***Spectral Entropy:*** A measure of the signal's complexity and randomness of brain activity in the frequency domain [31, 234]. It provides a single metric that reflects how power is distributed across frequencies, offering insights into brain states and conditions.

3. ***Band Power:*** It gives the power within specific frequency bands, including delta (0–3.5 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (14–30 Hz) and gamma (>30 Hz) [293].

Table 2.4 lists all the features from EEG signals in time and frequency domains.

2.3.1.3 Related Works of EEG for Emotion Recognition

Li et al. [269] investigated EEG features for cross-subject emotion recognition, examining different channels, brain regions, rhythms, and feature types. The authors identified the Hjorth mobility parameter in the beta rhythm as the most effective feature, achieving the highest mean recognition accuracy. Additionally, the study conducted a preliminary correlation analysis on 62 EEG channels, exploring highly correlated features for their potential to distinguish emotions across subjects. Khateeb et al. [154] improved emotion classification accuracy by extracting EEG features from the time and frequency domains. The authors used a grid search methodology to identify four optimal electrodes (FP1, FP2, F3, and C4) from a set of 32 electrodes for the

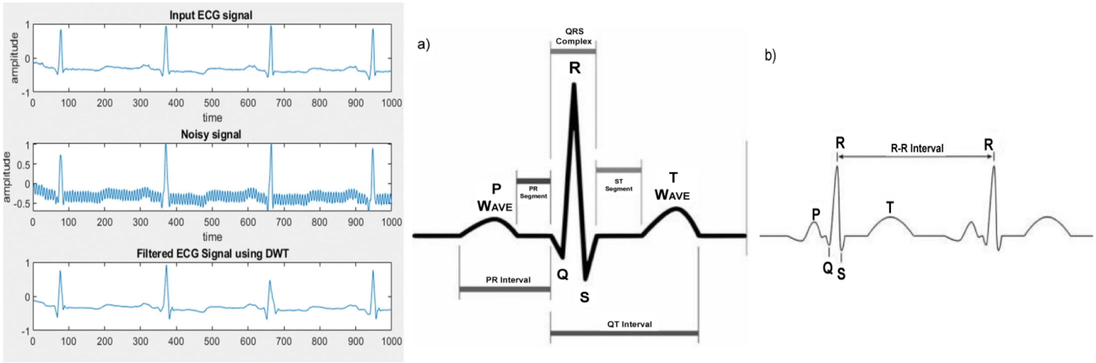


Figure 2.7: ECG original signal waveform and its filtered signal [201]

classification task. Gao et al. [275] employed a spatiotemporal attention mechanism to capture significant sequential and spatial information from EEG signals. The authors effectively represented diverse EEG activation patterns for different emotional states. The study concluded that EEG is essential for emotion identification.

2.3.2 Electrocardiographic (ECG) Signals

Electrocardiographic signals are electrical signals recorded to monitor the activity of the human heart [81, 292] as shown in Figure 2.7) [89]. These are essential for diagnosing various cardiac conditions and are used to understand heart rate variability and other aspects of cardiovascular health. The contraction and relaxation of the heart muscle upon electrical stimulation are represented numerically by the signals, which also indicate the potential fluctuations communicated to the skin surface due to the heart's electrical activity. The height of the ECG waves, measured in millivolts (mV), is known as amplitude.

Figure 2.7 illustrates the ECG signal waveform. It illustrates the different components of ECG waveform and is described as follows: (1). P Wave represents atrial depolarization, which is the electrical activity associated with the contraction of the atria. (2). QRS Complex represents ventricular depolarization, which is associated with the contraction of the ventricles. It is the most prominent part of the ECG waveform. (3). The T Wave represents ventricular repolarization when the ventricles return to their resting state [62]. (4). The RR interval is the time between successive R-wave peaks in the QRS complex of the ECG waveform [229, 292]. It measures the duration of one cardiac cycle, representing the time between two heartbeats [201].

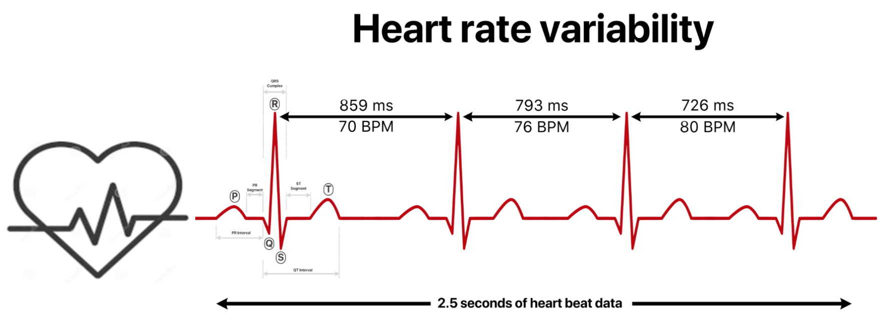


Figure 2.8: Heart Rate and Heart Rate Variability relation.

2.3.2.1 Significance of ECG for Emotions

ECG provides insights into cardiovascular responses associated with different emotions. For example, happiness is observed with a moderate increase in heart rate and higher Heart Rate Variability (HRV), signifying a flexible and adaptive autonomic nervous system response [107, 224, 240]. Sadness is characterized by lower HRV and a slightly reduced heart rate, indicating a state of reduced physiological arousal and autonomic inflexibility [94, 215]. Fear or anxiety causes a significant increase in heart rate and lower HRV, reflecting a heightened state of arousal and stress [58]. In contrast, relaxation is marked by a lower heart rate and higher HRV, denoting a state of calm and autonomic balance.

2.3.2.2 Heart Rate (HR) from ECG Signal

It is defined as the average number of times the muscle of the heart contracts or beats over a certain period, generally one minute [80, 198]. External bodily responses, including a smile or anger, and more physiological responses, such as an increase in heart rate, are regularly used by humans to communicate their emotions and current condition [57]. Unintentional body reactions naturally depict various emotions triggered by various physical reactions [80].

HR can be measured using various devices, including ECG monitors, wearable fitness trackers, and pulse oximeters [58]. Heart Rate reflects how many times the heart beats in a given period and is a fundamental measure of cardiovascular health. It is calculated by measuring the time between successive R-wave peaks (as shown in Figure 2.8) in the ECG (RR intervals) and converting this into Beats Per Minute (BPM) [29, 203]. The variation in time between successive heartbeats is the HRV [203,

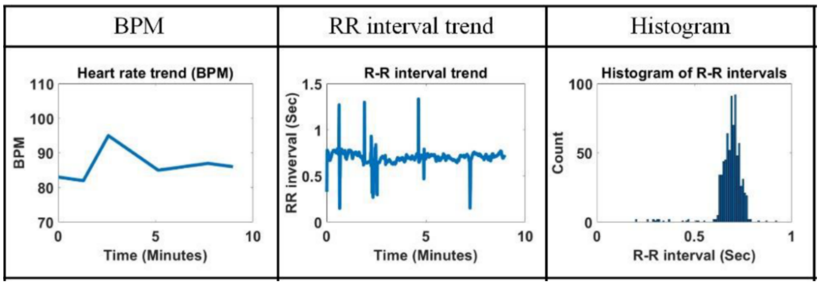


Figure 2.9: Heartbeats, RR interval, Histogram of RR interval [257] for a subject in neutral state.

239]. It reflects the variability in the RR intervals and indicates the autonomic nervous system's regulation of the heart. HRV can be analyzed using various methods, including time-domain, frequency-domain, and non-linear methods.

2.3.2.3 Significance of HR for Emotions

Heart rate and HRV are significant indicators of emotional states. Elevated heart rate is often linked to stress, anxiety, and excitement due to increased sympathetic nervous system activity. In contrast, a decreased heart rate generally reflects relaxation and contentment, associated with parasympathetic nervous system dominance [57]. HRV, which measures the variation in time between successive heartbeats, offers insights into emotional regulation; higher HRV indicates better emotional resilience and regulation, whereas lower HRV can be associated with negative emotional states like depression and anxiety. Analyzing these patterns can enhance emotion recognition and support applications like biofeedback for emotional management.

2.3.2.4 Features from ECG Signal

- **Time-Domain Features:** Directly analyze the ECG waveform, providing information on heart rate, intervals, and waveform amplitudes. These are described as follows:

  1. ***Mean Heart Rate:*** Average number of heartbeats per minute, calculated from the RR intervals [89, 292].
  2. ***Standard Deviation of Heart Rate (Std HR) :*** The standard deviation of heart rate values over a recording period [203].

3. ***Maximum, Minimum Heart Rate (Max, Min HR):*** The highest and lowest heart rate observed during a specific recording period [203].

4. ***RR Interval:*** Time interval between two consecutive R-wave peaks in the ECG signal. It reflects the heart rate variability (HRV) [125, 187, 203, 292].

5. ***Root Mean Square of Successive Differences (RMSSD):*** A measure of beat-to-beat variability in heart rate [60]. It is calculated from the square root of the mean of squared difference in successive RR intervals. It is an indicator of parasympathetic activity.

6. ***NN50 Count:*** Number of successive RR intervals differing by more than 50 ms, another measure of HRV [235, 276].

7. ***Triangular Index:*** It is calculated as the total number of all RR intervals divided by the height of the histogram's peak (as shown in Figure 2.9). It indicates the overall variability in heart rate [166, 203].

8. ***Triangular Interpolation of NN Interval Histogram (TINN):*** It is the baseline width of the RR interval histogram (as shown in Figure 2.9), measured by interpolating the area under the histogram to form a triangle [166, 203].

9. ***Coefficient of Variation of Successive Differences (CVSD):*** The ratio of the root mean square of successive differences (RMSSD) to the mean RR interval [203].

10. ***Coefficient of Variation of NN Intervals (CVNNI):*** The ratio of the standard deviation of RR intervals to mean RR interval, expressed as percentage [83].

11. ***Sample Entropy (SampEn):*** A measure of the complexity and irregularity of the RR interval time series. It is the negative logarithm of the conditional probability that sequences of similar patterns remain similar at the next point [129].

- **Frequency-Domain Features:** Examine the distribution of power across different frequency bands, offering insights into heart rate variability. The time domain ECG signal is converted into frequency components via FFT.

1. ***Power Spectral Density (PSD):*** Distribution of power across different frequency bands, providing information about energy in various frequency ranges of the ECG signal via Welch method [187].

2. ***Low-Frequency Power (LF, 0.04–0.15 Hz):*** Reflects sympathetic and parasympathetic activity [60, 193, 292].

3. ***High-Frequency Power (HF, 0.15–0.4 Hz):*** Mainly represents parasympathetic activity, often related to respiratory sinus arrhythmia [60, 193, 292].

4. ***LF/HF Ratio:*** The ratio of low-frequency to high-frequency power, indicating balance between sympathetic and parasympathetic nervous systems [60, 292].

5. ***Low-Frequency Power in Normalized Units (LFnu):*** The power in the low-frequency range (0.04–0.15 Hz) expressed as a percentage of total power (minus the VHF component). It is associated with parasympathetic activity [125].

6. ***High-Frequency Power in Normalized Units (HFnu):*** The power in the high-frequency range (0.15–0.4 Hz) expressed as a percentage of total power (minus the VLF component). It is associated with parasympathetic activity [125].

7. ***Total Power:*** The overall power in the ECG signal, indicating the total energy across all frequency bands [292].

Table 2.5 lists all the features from the ECG signal in the time and frequency domain.

2.3.2.5 Related Works of ECG for Emotion Recognition

Valenza et al. [89] focused on short-term emotion recognition using ECG signals. The authors characterized linear features of the Inverse Gaussian (IG) probability distribution, including the mean, standard deviation (STD), and power in low and high-frequency bands from ECG signals. Tarvainen et al. [174] created 'Kubios,' a framework for HRV analysis. This user-friendly software accepts various ECG and RR interval data formats for artefact removal, sample selection, and trend elimination, using

Table 2.5: Description of ECG signal features in time and frequency domain.

Domain	Features	Description
Time	Mean HR, Std HR	Average value, standard deviation of heart rate values
	Max HR, Min HR	The highest and lowest heart rates observed.
	RMSSD	Measure of beat-to-beat variability in heart rate
	NN50	Number of NN intervals differ by more than 50 milliseconds
	Triangular Index	Measure of the overall shape of the RR interval histogram
	TINN	Triangular interpolation of NN intervals
	CVSD	Coefficient of variation of successive differences between NN intervals
	CVNNI	Coefficient of variation of NN intervals
	SampEn	Measure of the irregularity or complexity of the time series ECG Signal
Frequency	Power Spectral Density	Measure of power distribution over frequency of the ECG signal, calculated using Welch method.
	LF	Power in the low-frequency (0.04 to 0.15 Hz) range of the ECG signal
	HF	Power in the high-frequency (0.15 to 0.4 Hz) range of the ECG signal
	LF/HF Ratio	Ratio of LF power to HF power
	LFnu	Normalized low-frequency power
	HFnu	Normalized high-frequency power
	Total power	The overall power in the ECG signal

optimal algorithms to calculate common frequency and time domain HRV features. It generates reports in text (ASCII), Matlab (.mat), or PDF formats. Subramanian et al. [221] conducted R-peak detection on ECG signals to calculate inter-beat intervals (IBI) by measuring the time difference between consecutive R-peaks. The authors extracted the HR, HRV features and PSD in low-frequency bands and computed a total of 32 features. In a pilot study, Guo et al. [94] used a movie clip approach to elicit five different emotional states. They recorded a 90-second ECG signal post-stimulus and used time, frequency domain, and statistical analysis to obtain HRV features. They computed 13 features, selecting five to categorize two distinct emotional states (positive and negative) and five emotions: sad, angry, fear, glad, and relaxed [94]. Costa et al. [112] developed EmotionCheck, a wearable device that detects users' heart rates and reduces anxiety levels through artificial feedback. Hasnul et al. [167] employed Neurokit and AuBT toolboxes for processing and feature engineering ECG signals. The authors detected R peaks and extracted 56 heart rate and heart rate variability features.

2.3.3 Galvanic Skin Response (GSR)

GSR (Galvanic Skin Response) is the fluctuation in sweat gland activity, reflecting the intensity of an individual's emotional arousal [100]. The emotional arousal changes in

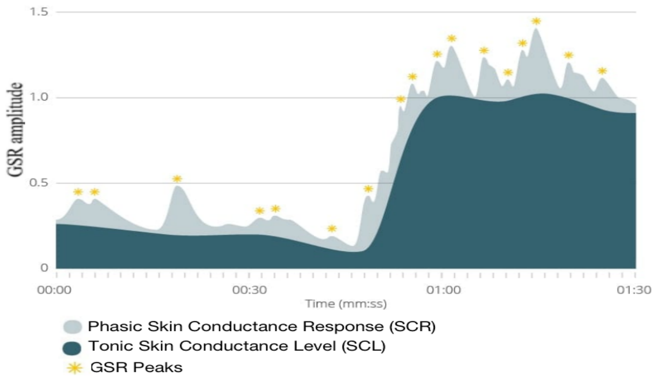


Figure 2.10: GSR signal and its components [203].

response to the surroundings, whether due to fear, joy, or other emotionally significant events, leads to more sweat. The GSR measures skin resistance by passing a tiny amount of current or voltage through the body and monitoring the variations between two sensor endpoints placed on the fingers [230]. It is also called Electrodermal Activity (EDA). Two electrodes are placed on the finger to test Skin Conductivity (SC) and Skin Resistance (SR), making the sensors on the fingertips essential. A tiny voltage is detected at the electrodes, providing data on the skin's electrical properties.

Figure 2.10 represents the GSR signal and its features. There are two components of GSR signals: (1) ***Phasic component*** presenting the peaks in GSR signals and also known as Skin Conductance Responses (SCRs). These represent transient increases in skin conductance that occur in response to specific stimuli or events [230, 282]. The height of the SCR peak indicates the magnitude of the response. Larger peaks usually reflect stronger or more intense stimuli. (2) ***Tonic component***, also known as Skin Conductance Level (SCL), represents the baseline level of skin conductance over time. It reflects the overall level of physiological arousal and can vary due to long-term changes in emotional or physiological states [178]. Changes in the tonic component can indicate gradual shifts in arousal levels or mood. For example, a sustained increase in SCL might suggest ongoing stress or anxiety [121]. The phasic component of the GSR signal is superimposed on the tonic baseline and reflects short-term variations in arousal.

2.3.3.1 Significance of GSR for Emotions

GSR measures changes in skin conductance due to sweat gland activity, which varies with emotional arousal. Happiness can be observed with a moderate increase in GSR, indicating a balanced state of arousal [153]. Sadness is associated with lower GSR, reflecting reduced arousal and emotional engagement [142, 178]. Fear and anxiety cause high GSR due to increased sweat gland activity, signifying heightened emotional and physiological arousal [155, 178]. During relaxation, GSR is low, indicating a state of calm and reduced arousal [155].

2.3.3.2 Features from GSR Signal

- **Time-Domain Features:** Time-domain features focus on the direct measurement and characteristics of the GSR signal over time.

  1. ***Mean Galvanic Skin Response (Mean GSR):*** The average value of the GSR signal over a specific period. Provides a baseline measure of skin conductance, reflecting the overall level of physiological arousal [243, 282].
  2. ***Variance of GSR (Var GSR):*** The variance of the GSR signal values over a specified period. Measures the degree of fluctuation or spread in the GSR signal [243].
  3. ***Skewness of GSR:*** A measure of the asymmetry of the GSR signal distribution about its mean. It indicates the direction of skew in the GSR data. Positive skewness suggests a distribution with a longer right tail, while negative skewness indicates a longer left tail [243, 282].
  4. ***Kurtosis of GSR:*** A measure of the sharpness of the GSR signal distribution. It reflects the presence of outliers or extreme values. Higher kurtosis indicates more outliers or extreme values, while lower kurtosis suggests a flatter distribution [243, 282].
  5. ***Standard Deviation of GSR (Std GSR):*** The standard deviation of the GSR signal values. It represents the amount of variation or dispersion from the mean GSR. Higher standard deviation indicates greater variability in skin conductance [243, 282].

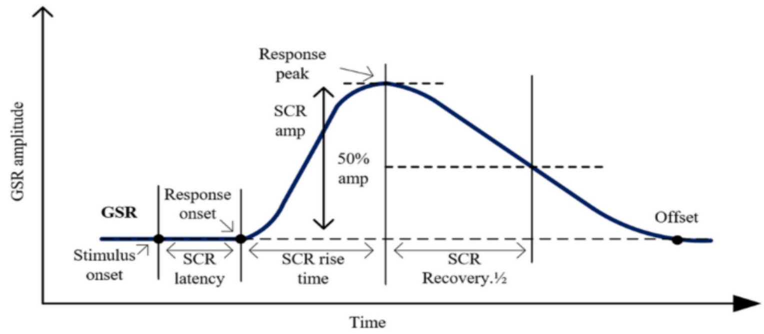


Figure 2.11: GSR signal and SCR components: Onsets, Peaks and Recovery [203]

6. ***Tonic Component:*** The slow, baseline-level changes in skin conductance are often obtained through low-pass filtering of the signal. It represents the general level of arousal over time [44, 133].

7. ***SCL (Tonic) Slope:*** The rate of change in the tonic component of the GSR signal is often calculated by fitting a linear model to the baseline (slow-changing) part of the signal. It Indicates the trend or direction of baseline skin conductance over time [44, 133].

8. ***SCR (Phasic) Peaks:*** The amplitude and timing of response peaks (as shown in Figure 2.11) in the phasic component of the GSR signal, which represents rapid changes in skin conductance in response to stimuli. It reflects the intensity and timing of skin conductance responses to specific stimuli or events. Higher peaks suggest stronger responses [44, 133].

- **Statistical Features Applied to SCR (Phasic) and SCL (Tonic)**

  1. ***Statistical Features for SCR (Phasic) [101, 148]:***

     – *Mean SCR Peak Amplitude:* The average amplitude of SCR peaks over a period.

     – *Variance of SCR:* Measures the average squared deviation of each SCR value from the mean SCR value.

     – *Standard Deviation of SCR Peak Amplitude:* The variability in the amplitude of SCR peaks.

Table 2.6: Description of GSR signal features in time and frequency domain.

Domain	Features	Description
Time	Mean GSR, Var GSR, Skew GSR, Kurtosis GSR, Std GSR	Average value, variance, skewness, kurtosis, standard deviation of the GSR signal
	Tonic Component	The slow, baseline-level changes in skin conductance
	SCL (tonic) slope	Slope of the tonic (slow-changing) component of the skin conductance level (SCL)
	SCR (phasic) peaks	The number of peaks representing the phasic (rapid-changing) component of the SCR
	Mean SCL, Var SCL, Std SCL, Mean SCR, Var SCR, Std SCR	Statistical features applied to SCR, SCL
Frequency	Power Spectral Density	Measure of power distribution over frequency of the GSR signal, calculated using Welch method.

2. ***Statistical Features for SCL (Tonic) [101, 148]:***

   – *Mean SCL Value:* The average value of the tonic component over a specified period.

   – *Variance of SCL:* Measures the average squared deviation of each SCL value from the mean SCL value.

   – *Standard Deviation of SCL:* The square root of the variance of SCL.

• **Frequency-Domain Features:** Frequency-domain features analyze the GSR signal in terms of its frequency components, providing insights into the underlying rhythms and periodicity.

1. ***Power Spectral Density (PSD):*** The distribution of power across different frequency bands of the GSR signal. Provides information about the frequency content and energy distribution in the GSR signal [125, 187, 233]. It is computed via the Welch method.

Table 2.6 lists all the features from the GSR signal in the time and frequency domain.

2.3.3.3 Related Works of GSR for Emotion Recognition

Nasoz et al. [84] used a SenseWear armband to collect the GSR signals for 29 participating subjects. The proposed study normalized the GSR signal data and fed it into the machine learning algorithms for emotion classification. Subramanian et al. [221] collected the GSR data from 58 university students while watching 36 video stimuli. The study extracted mean skin resistance, number of local minima in the

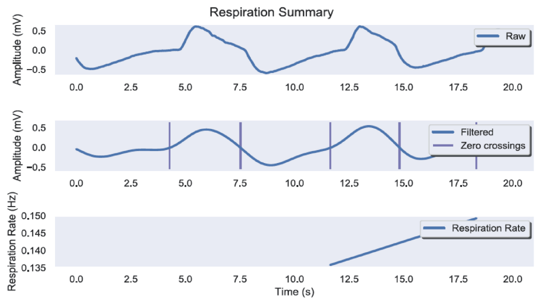


Figure 2.12: Raw and filtered Respiration signal [203].

GSR signal, average rising time of the GSR signal, spectral power in the [0-2.4] Hz band, zero crossing rate of skin conductance slow response ([0-0.2] Hz), zero crossing rate of skin conductance features from the GSR signals and then fed them to train the machine-learning models. Ragot et al. [159] also collected the EDA data from Empatica E4 [6] smartwatch for 19 French volunteers and extracted the SCL features from the EDA signal. The study then fed them into the machine learning models for further emotion recognition. Ali et al. [151] extracted statistical measures from Skin Conductance (SC), including the mean, standard deviation, maximum, minimum, root mean square, the means of the first and second derivations, and the GSR signal's negative slope. The authors extracted the response latency of the first significant Skin Resistance (SR), the sum of SR amplitudes, and the sum of SR areas. Perry et al.[121] extracted basic statistical variables, first-order differential variables, and second-order differential variables from EDA signals for emotion analysis. Tara Hassani [100] used GSR data in a federated learning environment. The author utilized GSR data from the CASE [134] dataset to propose a model with good accuracy for valence and arousal emotions and ensured privacy for the GSR data.

2.3.4 Respiration (RESP)

It measures a subject's breathing pattern (as shown in Figure 2.12), specifically how deeply and quickly they breathe. In the context of emotion recognition, rapid and deep breathing can signal high arousal states, such as anger, fear, or joy. On the other hand, deep and slow breathing denotes a calm, peaceful state, indicating a sign of passivity or passive states like grief or subtle emotions. The respiration signal measures changes in the volume of air entering and leaving the lungs. It reflects breathing patterns and

Table 2.7: Description of Respiration signal features in time domain.

Domain	Features	Description
Time	Mean RESP, Median RESP, Var RESP, Skew RESP, Kurtosis RESP, Std RESP	Average value, median value, variance, skewness, kurtosis, standard deviation of the Respiration signal

respiratory rate and is often used in monitoring and diagnosing respiratory health. The respiration signal measures changes in the volume of air entering and leaving the lungs. It reflects breathing patterns and respiratory rate and is often used in monitoring and diagnosing respiratory health.

2.3.4.1 Significance of RESP for Emotions

RESP provides a window into respiratory patterns associated with emotions. Regular, moderate breathing rate, signifying a balanced and positive emotional state, can be linked to happiness [74, 102]. Sadness is associated with slower, deeper breathing, reflecting a contemplative or melancholic state [102]. Fear and anxiety lead to rapid, shallow breathing, indicative of heightened arousal and stress [74, 153]. Relaxation, in contrast, is characterized by slow, deep breathing, denoting a state of calm and reduced physiological arousal [74, 153]

2.3.4.2 Features from RESP Signal

- ***Mean Respiratory Rate (Mean RESP):*** The average rate of respiration over a specified period, typically measured in breaths per minute (bpm). It provides a general measure of the average breathing rate [55].
- ***Median Respiratory Rate (Median RESP):*** The middle value of the respiratory rate when the data is ordered from smallest to largest. It offers a robust measure of central tendency, less affected by outliers compared to the mean [36, 233].
- ***Variance of Respiratory Rate (Var RESP):*** The measure of how much the respiratory rates deviate from the mean, calculated as the average squared deviation from the mean [36, 233].
- ***Standard Deviation of Respiratory Rate (Std RESP):*** The standard deviation of the respiratory rate values over a specified period. Measures the dispersion or variability of the respiratory rate around the mean [36, 233].

Table 2.8: Benefits and limitations of physiological signals.

Physiological Signals	Benefits	Limitations
EEG	Permits measurements of individuals with disabilities like paraplegia and facial paralysis.	It requires complicated installation and maintenance of equipment susceptible to movement artefacts.
ECG	Mobile measures (i.e. smart clothes with embedded sensors, smartwatch) are possible during cardiac examinations.	It allows movement artefacts in mobile systems, higher accuracy in stationary measuring.
EDA	It is a good emotional indicator and distinguishes between conflict and non-conflict situations	It only measures arousal, which is influenced by temperature and requires reference and calibration.
RESP	Installation is easy and simple and can signify panic, terror, focus, or depression.	It is tough to distinguish between the many emotional spectrums.

Table 2.7 gives the description of the features from Respiration signal in time domain.

2.3.4.3 Related Works of RESP for Emotion Recognition

Folschweiller et al. [231] reviewed recent findings on brain activity synchronized with respiration, focusing on emotional cognition. The review highlighted how the brain's respiratory rhythm regulates emotion and cognition. The authors emphasized that changes in breathing rhythm's frequency, regularity, and amplitude correspond to different emotional and arousal states, underscoring its adaptive role in supporting cognition based on emotional context [231]. They concluded by highlighting the evident bidirectional relationship between emotions and respiration. Hafeez et al. [103] analyzed and processed respiration rate data collected by impulse radio ultra-wideband (IR-UWB) from 35 participants. The authors extracted statistical features from the respiratory rhythm to classify three emotions: fear, disgust, and happiness. Kyle et al. [138] detected child emotion regulation and liability by observing variations in the physiologic stress response employing respiration rate fluctuations during a mirror-tracing activity.

2.4 Multi-Modal Physiological Signals

Single physiological signals include challenges related to signal quality and noise, sensor accuracy, individual variability, data interpretation, integration, and synchronization, as shown in Table 2.8. For example, long-term monitoring can result in signal drift, where baseline levels shift over time, affecting the accuracy of

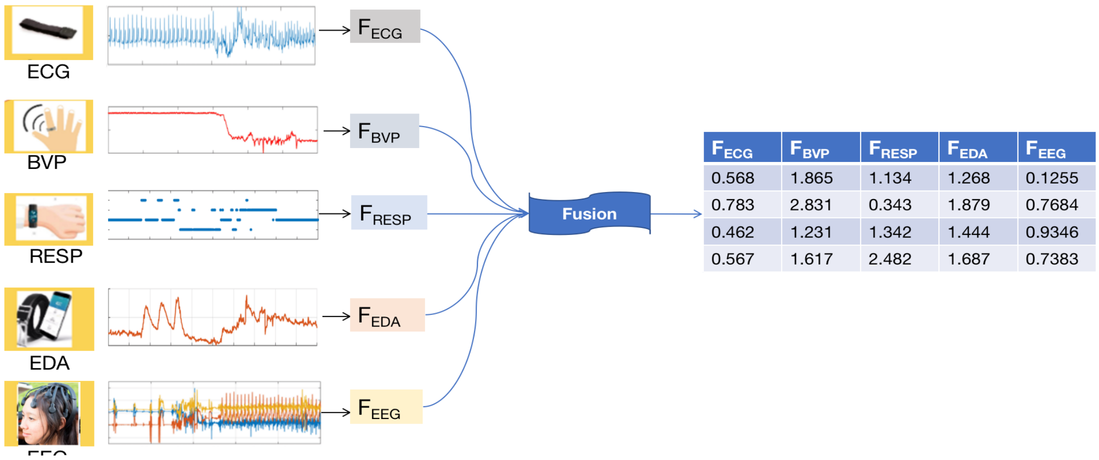


Figure 2.13: Illustration for multi-modal physiological signals fusion.

measurements. Hence, combining these signals is encouraged for a more accurate prediction rate [50, 92, 109].

The combining of different physiological signals is called multi-modal physiological signals. Multi-modal physiological signals refer to collecting and analyzing data from multiple physiological sensors, such as heart rate, skin conductance, body temperature, and blood pressure, to provide a comprehensive and integrated view of an individual's physiological and bodily responses. Combining information from various measurement modalities enables a more detailed and accurate assessment of health, stress, and emotional responses [92, 109]. This approach leverages the strengths of different sensors to overcome the limitations of single-modality measurements, leading to improved diagnostic accuracy and a deeper understanding of complex physiological interactions. Figure 2.13 gives an illustration of the fusion process. Physiological signals can be combined using fusion methods such as the following:

- ***Feature-Level Fusion (FLF):*** It integrates the features extracted from different physiological signals into a single feature set before making any prediction. This approach combines raw data or derived features from multiple modalities into a unified representation vector [50, 92, 109]. It involves steps such as (1) Data processing for the same frequency range, (2) Data Extraction, (3) Feature Integration, and (4) Feature Selection.

- ***Decision-Level Fusion (DLF):*** It combines the outcomes or decisions made by individual models that analyze each physiological signal separately. Instead of

Figure 2.14: Workflow for physiological data pre-processing steps.

merging raw data or features, this approach combines the results of separate analyses to reach a final decision [50, 92]. It involves steps such as (1) Individual Analysis and (2) Combining Decisions.

Recent studies [50, 92, 109, 135] evaluated coherence between multi-modal signals to allow feature-level fusion and discovered that it enhanced overall accuracy over decision-level fusion. Hence, this thesis uses the FLF approach for proposing the multi-modal FL-based ERS architectures.

2.5 Data Processing of Physiological Signals

Data processing is a broader term for collecting, storing, manipulating, and analyzing data. It encompasses data preprocessing and the various methods used to analyze and interpret the data. Figure 2.14 gives a workflow for the data Pre-processing steps.

2.5.1 Data Pre-Processing

In order to prepare raw data for analysis and modelling, data preprocessing is essential. It must be cleaned, transformed, and encoded to be fed into machine learning algorithms. Missing value management, data normalization or standardization, and data division into training and test sets are all included in this procedure. Data preprocessing enhances the effectiveness and reliability of subsequent analytical and predictive tasks by ensuring data is accurate, consistent, and in the required format. It can be done with a variety of tools and methods, as listed below:

2.5.1.1 Filtering

EEG, ECG, GSR, and other physiological signals require filtering at specific frequencies to improve data accuracy. Filtering either removes specific frequencies or retains others. If the subject moves during data collection, sensor readings can become inaccurate and contain noises. Data filtering methods such as *Butterworth filters* [61, 147] including *High-pass filters* [64], *Low-pass filters* [147], *Band-pass filters* [73], and *Elliptic filters* [61, 147, 192, 228] prevent disturbances and artefacts from the raw signal. Butterworth filter is a low-pass, high-pass, or bandpass filter with a maximally flat frequency response in the passband. A high-pass filter removes low-frequency noise, such as baseline drift, below a specific cutoff frequency. A low-pass filter removes high-frequency noise above a certain cutoff frequency. Bandpass filter passes signals within a specific frequency range and attenuates frequencies outside this range. Different physiological measurements require different filters; for example, a low-pass filter works best for ECG at 100 and 500 Hz. Additionally, Fourier frequency analysis splits raw EEG signals and uses a bandpass filter to remove detected noise. EEG signals are particularly prone to noise and artefacts due to the large number of electrodes and sensitivity to facial movements.

Bos et al. [73] used a bandpass filter from EEGLab for Matlab to remove noise and artefacts from EEG recordings. They applied a moving average filter to preprocess RESP and EDA signals [87]. Izard [64] used high-pass filters with frequencies of 0.1 Hz and 4 Hz for RESP and ECG data, respectively. To preprocess ECG signals and smooth GSR signals, the author employed Butterworth filter [61, 147, 192, 228].

2.5.1.2 Artefact Removal and Denoising

Artefacts such as muscle movements, eye blinks (in EEG), or motion artefacts (in EEG) from environmental, experimental, and physiological conditions can degrade signal quality and make parts unusable. It is crucial to eliminate these artefacts to ensure reliable analysis of physiological signals. The major techniques for this are:

1. **Independent Component Analysis (ICA):** is an analytical approach that works for source signals that are statistically independent and non-Gaussian. It then finds a linear transformation that maximizes the independence between the resulting components [42, 88]. It separates physiological signals from noise and artefacts, such as separating brain activity signals from eye movement artefacts in EEG data.

2. **Empirical Mode Decomposition (EMD):** is a signal decomposition method specifically designed to handle non-linear and non-stationary signals (largely for

EEG and ECG) [180, 258]. It's an adaptive method that decomposes a signal into a series of Intrinsic Mode Functions (IMFs), each representing oscillatory modes at different time scales. EMD allows for the separation of different frequency components that correspond to different brainwave frequencies, such as delta, theta, alpha, and beta, without the need for a predefined filter bank [246, 290]. The different IMFs may contain artefacts, which can be disregarded, while those without artefacts can be readily identified and selected for further experimentation and analysis. EMD has three extensions: Bivariate EMD for two channels, Trivariate EMD for three channels, and Multivariate EMD for up to 32 channels [34, 177].

3. **Principal Component Analysis (PCA):** is a dimensionality reduction technique that transforms data into a new coordinate system. Orthogonal (uncorrelated) principal components represent the transformed data. It helps filter out noise and irrelevant variations in the data by focusing on the principal components that capture the most variance.

Valenza et al. [88] used ICA to extract the Respiration Sinus Arrhythmias (RSA) feature from ECG signals. Bigirimana et al. [42] proposed an artefact removal approach combining ICA with Wavelet Transform (ICA-WT) for EEG signals, significantly improving recognition performance compared to standard ICA.

Zhang et al. [290] used EMD to convert the EEG signals into IMFs and then extracted the features from the IMFs using an autoregressive (AR) Model. Patel et al. [218] used EMD to remove eye blinks from EEG signals by decomposing them into IMFs. Mert et al. [34] used multivariate EMD (MEMD), which is an extension of EMD works for up to 32 EEG channels for extracting the IMFs [177].

Qiang et al. [206] applied PCA to select features and remove irrelevant or redundant features from EEG signals. Arjun et al. [46] employed PCA for dimensionality reduction, reducing 32-channel raw EEG data in the DEAP dataset to 8 dimensions. Alickovic et al. [75] compared three noise removal algorithms, Multiscale Principal Component Analysis (MSPCA), PCA, and ICA, concluding that MSPCA performs the best among the others.

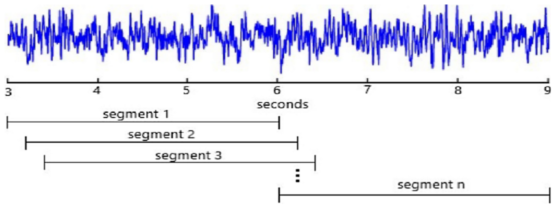


Figure 2.15: Overlapping windows on EEG signal [104].

2.5.2 Feature Extraction

Feature extraction from physiological signals is crucial in preparing data to be fed into Machine Learning (ML) and Deep Learning (DL) algorithms for emotion classification. Following preprocessing, the signals are segmented into temporal segments called windows. Figure 2.15 presents an illustration of an EEG signal with overlapping windows. Various features can be extracted from each of the segments to capture relevant information [40, 77, 210]. These features include statistical metrics such as mean and variance, temporal features such as peak intervals and signal amplitude changes, or spectral features such as frequency components and power spectral density [36, 190, 210]. The extracted features are then used as input to ML and DL algorithms, enabling the models to learn patterns and relationships within the data, ultimately facilitating accurate classification or prediction of physiological states or emotions [125, 135, 233]. This thesis uses the features extracted from the physiological signals EEG, ECG, GSR, and RESP as explained in earlier section 2.3.

2.6 Summary

An in-depth examination of emotions includes two primary models for categorizing and mapping emotions: discrete models and dimensional models (2D-VA and 3D-VAD models of emotions). The 2D-VA model maps emotions along Valence and Arousal axes, while the 3D-VAD model expands this by adding a third axis, Dominance. Emotion ground truth is gathered through self-assessment questionnaires such as SAM [119], PANAS [71], and POMS [220], with SAM being widely utilized. The 3D-VAD model enables a broader range of emotion classification (octal classification) compared to the 2D-VA model (binary classification). For accurate emotion mapping, physiological signals like EEG, ECG, and GSR are valuable as they reflect an individual's emotional state through bodily responses, capturing essential characteristics in both time and frequency domains. These signals undergo data processing techniques such as artefact removal, filtering, and feature extraction to enable emotion classification. Multi-modal physiological data achieves more accurate emotion classification than single-physiological signal data.

Chapter 3

Open Emotion Datasets

This chapter presents various physiological signals-based emotion recognition datasets, which are openly accessible [67, 195, 233, 299]. These datasets provide diverse physiological signal data, including EEG, ECG, GSR, and others, recorded using a variety of wearable biosensor devices such as EEG headsets, respiration belts, smart wristbands and watches. Each dataset is meticulously curated, encompassing a wide range of emotional states, enabling researchers to train and validate their models effectively. Moreover, these datasets facilitate actual comparisons between different methodologies' outcomes, serving as benchmarks for emotion recognition algorithms. Each dataset's unique characteristics and diverse emotional contexts highlight the need for robust models that can generalize across different settings.

FL has been applied to ERS, particularly using the DEAP [233], DREAMER [135], and CASE [134] datasets. However, these studies have typically utilized only peripheral signals from the DEAP [233] dataset or focused exclusively on EEG data from the CASE [134] and DREAMER [135] datasets without incorporating multi-modal physiological signals. In contrast, this thesis leverages multi-modal physiological signals from these datasets to provide a more comprehensive approach to emotion recognition.

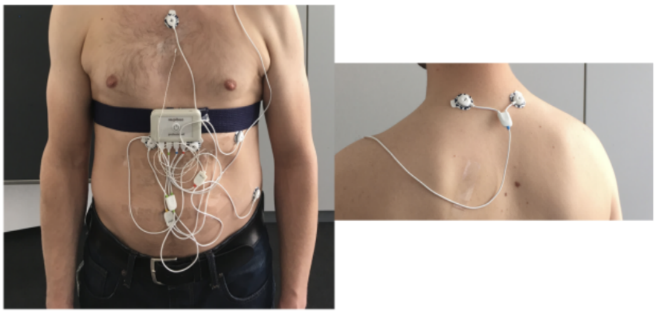


Figure 3.1: Experimental setup for WESAD dataset collection [195].

3.1 Wearable Stress and Affect Detection (WESAD) [195]

WESAD [195] is a comprehensive dataset that collected data from 15 subjects, including 12 males and 3 females. It collected the following multi-modal physiological signals:

- ECG, EMG using RespiBAN Professional [12].
- EDA, RESP, SKT using EMPATICA E4 [6].

WESAD [195] is collected for two scenarios: (1) subjects were shown 11 funny short videos (for a total of 392 seconds), and (2) subjects performed a Trier Social Stress Test (TSST) [54]. The subjects rated themselves with PANAS [71] and STAI [63] tests. Figure 3.1 presents the data collection setup for WESAD [195].

3.2 SJTU Emotion EEG Dataset (SEED) [299]

SEED [299] is a narrow EEG-based dataset collected EEG data in the BCMI laboratory. It consists of EEG data for 15 subjects, including 7 males and 8 females, using a 62-channel ESI NeuroScan System [7]. The data was collected while the subjects watched 15 Chinese film clips for 15 trials in one experiment. The subjects rated their emotional reactions to each film clip by completing the questionnaire immediately after watching each clip with positive, negative and neutral emotion labels [299]. This dataset has two limitations: (1) it only consists of EEG physiological data, and (2) the stimulus shown is only in the Chinese language, which is restricted to a specific population. Figure 3.2 presents the stimuli and data collection.

Figure 3.2: Experimental setup for SEED dataset collection [299].

Table 3.1: Details of the 18 film clips shown to each subject in DREAMER [135].

Video ID	Film Clip	Target Emotions
1	Searching for Bobby Fischer	Calmness
2	D.O.A	Surprise
3	The Hangover	Amusement
4	The Ring	Fear
5	300	Excitement
6	National Lampoon's VanWilder	Disgust
7	Wall-E	Happiness
8	Crash	Anger
9	My Girl	Sadness
10	The Fly	Disgust
11	Pride and Prejudice	Calmness
12	Modern Times	Amusement
13	Remember the Titans	Happiness
14	Gentlemans Agreement	Anger
15	Psycho	Fear
16	The Bourne Identity	Excitement
17	The Shawshank Redemption	Sadness
18	The Departed	Surprise

3.3 A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices (DREAMER) [135]

DREAMER [135] is a wide dataset collected data from Twenty-three subjects, including 14 males and 9 females, while watching 18 film clips (ranging from 65-393 seconds) as shown in Table 3.1. It collected the following physiological signal data:

- EEG using 14-channel Emotiv EPOC [5] at 128 Hz.
- ECG using wireless SHIMMER [15] sensor at 256 Hz.

The subjects rated their emotions using SAM [119] ratings for arousal, valence and dominance on a scale of 1 to 5.

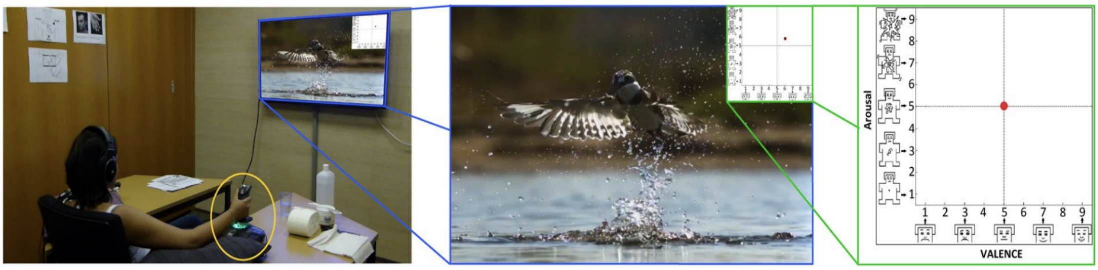


Figure 3.3: Experimental setup for CASE dataset collection [134]

3.4 Continuously Annotated Signals of Emotion (CASE) [134]

CASE [134] dataset collected data from 30 subjects, including 15 males and 15 females, while watching 20 videos (each of approx. one minute). It collected the following multi-modal physiological signal data:

- ECG, RESP, blood flow, EDA, and SKT at 1000Hz.

The subjects rated their emotions while watching video stimuli using SAM [119] for arousal and valence on a scale of 1 to 5. Figure 3.3 presents the experimental setup for data collection.

3.5 A Database for Emotion Analysis using Physiological signals (DEAP) [233]

DEAP [233] is a rich dataset comprising data from 32 subjects, including 16 males and 16 females, while watching 40 video stimuli, each lasting for one minute. It collected the following physiological signal data:

- EEG, ECG, EDA, SKT, and RESP using the Biosemi ActiveTwo system [3].

It recorded their ratings using SAM [119] ratings for arousal, valence and dominance on a scale of 1 to 9 and liking and familiarity on a scale of 1 to 5. Figure 3.4 presents the experimental setup for data collection.

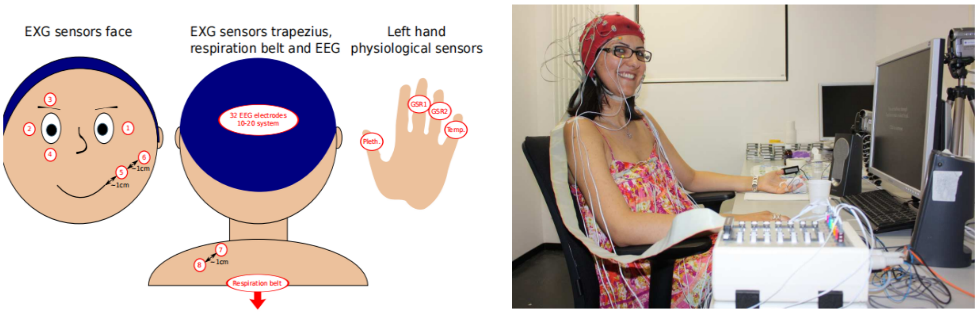


Figure 3.4: Experimental setup for DEAP dataset collection [233].

Table 3.2: Details of the 15 film clips shown to each subject in EMDB dataset [148].

Video ID	Film Clip	Target Emotions
1	On Golden Pond	Amusement
2	When Harry met Sally	
3	An Officer and a Gentleman	
1	The Champ	Sadness
2	The Killing Fields	
3	Witness	
1	Gandhi	Anger
2	Cry Freedom	
3	The Godfather	
1	Maria´s Lovers	Disgust
2	Pink Flamingos	
3	Modern Times	
1	Silence of the Lambs	Fear
2	Halloween	
3	Marathon Man	

3.6 BioVid Emo DB (EMDB) [148]

EMDB [148] database collected heart rate and skin conductance level using Nexus-32 amplifier device [10]. It collected data from 86 subjects while watching 15 film clips (detailed in Table 3.2). It recorded their ratings using SAM [119] for arousal, valence, amusement, sadness, anger, disgust and fear on a scale of 1 to 9. This dataset is limited to only two physiological signals, HR and SCL.

Table 3.3: Details of the 16 short videos shown to each subject in AMIGOS dataset [125].

Video ID	No. of Samples	Duration (in sec)	Videos
10	12225	96	August Rush
13	7229	57	Love Actually
138	15160	122	The Thin Red Line
18	10575	83	House of Flying Daggers
19	16106	126	Exorcist
20	8335	65	My Girl
23	14265	112	My Bodyguard
30	9717	76	Silent Hill
31	19886	155	Prestige
34	8417	66	Pink Flamingos
36	8698	68	Black Swan
4	11621	91	Airplane
5	14347	112	When Harry Met Sally
58	8181	64	Mr Beans Holiday
80	13047	102	Love Actually
9	9630	75	Hot Shots

Figure 3.5: Experimental setup for AMIGOS dataset collection [125].

3.7 A dataset for Multi-modal research of affect, personality traits and mood on Individuals and GrOupS (AMIGOS) [125]

AMIGOS [125] is a rich dataset comprising data recordings in two experimental settings. In the first one, 40 subjects, including 13 females and 27 males, watched 16 short emotional videos (57-155 seconds) given by Table 3.3. In the second one, the subjects watched 4 long videos (ranging from 14-23 minutes), some of them alone and the rest in groups. It collected the following physiological signal data:

- EEG using 14-channel Emotiv EPOC Neuroheadset [4]
- ECG, GSR using Shimmer 2R [14]

The subjects were asked to self-report their emotional experiences, providing annotations for various emotion dimensions such as arousal, valence, and dominance

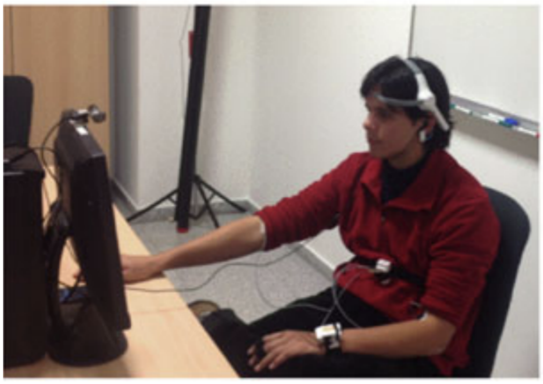


Figure 3.6: Experimental setup for ASCERTAIN dataset collection [221].

using SAM [119] ratings on a scale of 1 to 9. Figure 3.5 presents the experimental setup for data collection from alone and group scenarios.

3.8 A multi-modal databaASe for impliCit pERsonal-iTy and Affect recognitIoN (ASCERTAIN) [221]

ASCERTAIN [221] dataset collected physiological data from EEG using a single dry-electrode EEG device, ECG from two measuring electrodes placed at each arm crook, and GSR from two electrodes positioned on the middle and index finger phalanges. It collected data from 58 subjects, including 37 males and 21 females, while watching 36 movie video clips (ranging from 51-128 seconds). It recorded their ratings for arousal and valence for emotions on a scale of 1 to 7. This dataset has a limitation regarding emotion dimension, as it only rates emotions using valence and arousal. Figure 3.6 presents the experimental setup for data collection.

3.9 Chosen Emotion Datasets for Proposed FL-based Automated ERS

Table 3.4 presents the summary of all the described datasets. Out of these datasets, this thesis uses the benchmark datasets DEAP[1] [233], AMIGOS[2] [125],

[1] https://www.eecs.qmul.ac.uk/mmv/datasets/DEAP/download.html
[2] http://www.eecs.qmul.ac.uk/mmv/datasets/AMIGOS/index.html

Table 3.4: Summary of openly accessible datasets for ERS based on physiological signals.

Summary of Datasets				
Dataset	**No. of Subjects**	**Emotion Dimensions**	**Physiological Modalities**	**FL Applied**
WESAD [195]	15	Neutral, Stress, Amusement	EDA, ECG, RESP, SKT, EMG, PPG	×
SEED [299]	15	Neutral, Positive, Negative	EEG	×
DREAMER [135]	23	Arousal, Valence, Dominance	EEG, ECG	×
CASE [134]	30	Arousal, Valence, Dominance	EMG, ECG, EOG, Facial Video	✓
DEAP [233]	32	Arousal, Valence, Liking, Dominance, Familiarity	EEG, ECG, EDA, SKT, RESP, Facial Video, EMG, EOG	✓
EMDB [148]	32	Arousal, Valence, Dominance	SCL, HR	×
AMIGOS [125]	40	Arousal, Valence, Dominance, Familiarity	EEG, ECG, GSR	×
ASCERTAIN [221]	58	Arousal, Valence, Liking, Engagement, Familiarity	ECG, EDA, EEG, Facial Video	×

and DREAMER[3] [135], as they comprise multi-modal physiological data, majorly including EEG. These three datasets validate the proposed architectures F-MERS and AFLEMP as described later in Chapters 6, 7 and 8, respectively.

3.9.1 Data processing for Emotion Datasets (AMIGOS, DEAP, DREAMER)

The proposed F-MERS and AFLEMP architectures uses the pre-processed data given by the data owners as it has achieved state-of-the-art results. The steps are given below:

1. In the AMIGOS [125] dataset, Not a Number (NaN) values are detected in the ECG signals for the 28th subject for the 9th video, which is removed from the experimental setup.
2. The EEG, GSR, RESP and ECG data are downsampled to a common frequency of 128 Hz in order to fuse them in a common feature vector (For all three datasets) [125, 135, 233].
3. EEG data is filtered from 4.0 to 45.0 Hz with a bandpass filter (For all three datasets).
4. The ECG and GSR data are filtered with a cutoff frequency of 60Hz using a low-pass filter (For AMIGOS [125]).

[3] https://zenodo.org/record/546113#.ZEn3hi8RpQI

3.9.2 Data Clipping

1. ***AMIGOS [125]:*** The proposed *F-MERS* uses the 16-short video experiments since longer videos are less likely to elicit stable emotions [37]. Each video displayed to the subject in the experiment is of varying duration, from which we use the last 50 seconds of data of every video and clip the starting portion.
   - Data Matrix =[total subjects x videos (per subject) x samples (per video)]
   = [40 x 16 x 6400 (50 x 128)]

2. ***DEAP [233]:*** The proposed F-MERS uses the data from the last 60 seconds of every video stimuli shown to the subjects after removing the 3-second baseline.
   - Data Matrix = [total subjects x videos (per subject) x samples (per video)]
   = [32 x 40 x 7680 (60 x 128)]

3. ***DREAMER [135]:*** The proposed F-MERS uses the data from the last 60 seconds of every video stimulus shown to the subjects.
   - Data Matrix = [total subjects x videos (per subject) x samples (per video)]
   = [23 x 18 x 7680 (60 x 128)]

Table 3.5 presents the data description for all three datasets and their pre-processing, which is further used for validating the proposed architectures in Chapter 6, 7, and 8.

3.9.3 Data Labelling

Subjects from all three datasets use Self-Assessment Manikin (SAM) [119] to rate their valence, arousal, and dominance levels. On a scale of 1-9 (in AMIGOS [125], DEAP [233]) and 1-5 (in DREAMER [135]). The emotions of low and high valence, arousal, and dominance are defined by a threshold float value of 4.5 (in AMIGOS [125], DEAP [233]) and 3 (in DREAMER [135]), as shown in Table 3.6.

3.10 Summary

Various publicly accessible emotion recognition datasets exist based on physiological signals [67, 195, 233, 299]. These datasets include physiological signal data such as EEG, ECG, and GSR, recorded using devices like EEG headsets, respiration belts,

Table 3.5: Descriptions with pre-processing of the emotion dataset used in the thesis.

Dataset Description			
Content	DEAP [233]	AMIGOS [125]	**DREAMER**[135]
Subjects	32	40	23
Videos per subject	40	16	18
Video duration	63 sec	57-155sec	65-393 sec
Physiological Signals	EEG, GSR, RESP	EEG, ECG, GSR	EEG, ECG
Emotion Dimensions	VAD	VAD	VAD
Recorded Signal	EEG, GSR, RESP 512 Hz	EEG 128 Hz, ECG, GSR 256 Hz	EEG 128 Hz, ECG 256 Hz
Data Matrix (3D) (Subjects x videos x samples)	32 x 40 x 7680	40 x 16 x 6400	23 x 18 x 7680
Label Matrix (2D)	40 x 3	16 x 3	18 x 3
Emotion Assessment	Self-Assessment Manikin (SAM) [119]		
EEG Device	BioSemi Active Two [3]	Wireless EMOTIV EPOC [5]	
EEG Electrodes	32 channels	14 channels	
ECG, GSR Device	BioSemi Active Two [3]	Wireless SHIMMER [15]	
ECG Electrodes	-	ECG right and Left	
GSR Electrode		channel no. 17	-
Data processing			
Downsampling (at 128 Hz)	EEG, GSR, RESP	EEG, ECG, GSR	EEG, ECG
Bandpass filtering (from 4.0-45 Hz)	EEG		
Low-pass filtering (at 60 Hz)	GSR	ECG, GSR	ECG
Baselines	3 seconds	5 seconds	4 seconds
Data Clipping	Uses the last 60 seconds of every video experiment and clips the starting portion.	Uses the last 40 seconds of every video experiment and clips the starting portion.	Uses the last 60 seconds of every video experiment and clips the starting portion
Data Labelling	Subjects rated their emotions for VAD using the SAM on scale of 1-9 (4.5 as threshold)	Subjects rated their emotions for VAD using the SAM on scale of 1-9 (4.5 as threshold)	Subjects rated their emotions for VAD using the SAM on scale of 1-5 (3 as threshold)

Table 3.6: Emotions ratings for open emotion datasets using 3D-VAD

Ratings	**AMIGOS [125]**	**DEAP [233]**	**DREAMER [135]**
Low (Arousal/Valence/Dominance)	1 - 4.5	1 - 4.5	1 - 3
High (Arousal/Valence/Dominance)	4.5 - 9	4.5 - 9	3 - 5

and smart wristbands. Each dataset is carefully curated to represent a broad spectrum of emotional states, offering the ability to train and validate classification models effectively. Additionally, these datasets serve as valuable benchmarks, allowing for direct comparisons between different emotion recognition methodologies.

Chapter 4

Machine Learning for Automated ERS using Physiological Signals

A recent upsurge in research focuses on developing automated emotion recognition technology by leveraging machine learning and deep learning algorithms trained on physiological signal data. This chapter presents the various machine learning and deep learning methodologies and approaches for classifying emotional states.

4.1 Machine Learning Models for Emotion Recognition using Physiological Signals

Machine learning approaches enable a trained model to identify patterns in raw data and extract meaningful insights. The training phase is crucial for ensuring the accuracy of the analysis and requires extensive and diverse datasets to achieve a high-quality model [247]. The process involves collecting data from various sources, consolidating it in a central repository, and then using it to train the model. This centralization can lead to privacy issues, as data from different sources may be accessible to others, potentially violating data privacy regulations and increasing the risk of misusing sensitive information. These challenges are crucial when managing sensitive physiological data containing emotional information. This issue of privacy concern is an obstacle to advancing machine learning techniques in healthcare [76]. Machine learning model maps physiological signal characteristics to their labels using

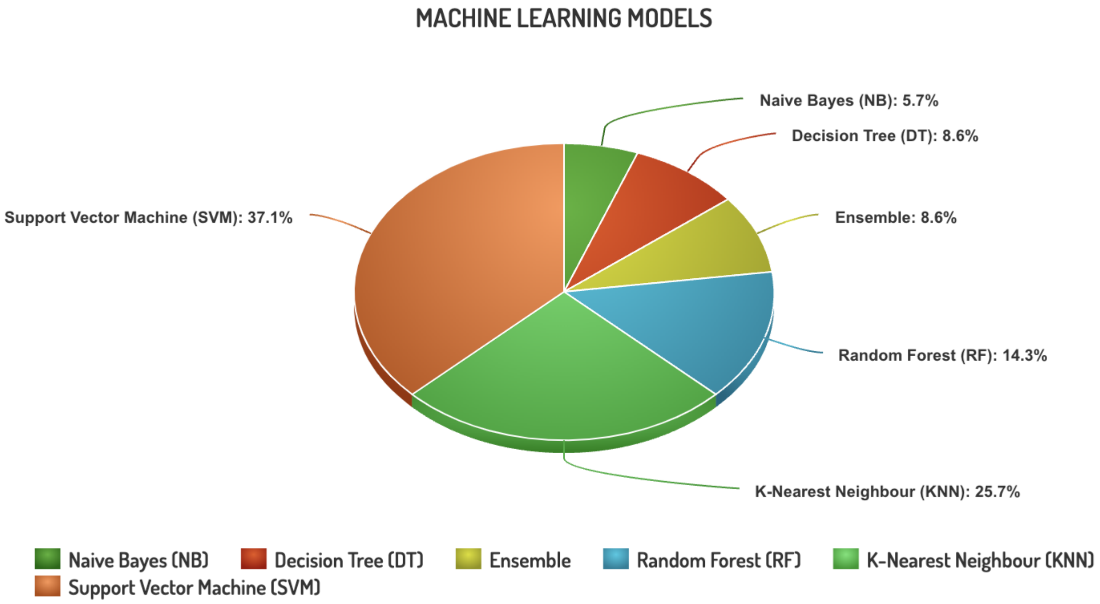


Figure 4.1: Distribution of ML models for automated motion Recognition System (ERS).

a training set [125, 135, 145]. SVM [252], K-NN [84], Random Forest (RF) [185], and Decision Tree (DT) [111] are some of the examples of Supervise Learning algorithms. The SVM classifier is the most widely used classification model in the literature(as shown in Figure 4.1) [111, 194, 206, 252].

Related Works: Multiple ML models have been implemented for emotion recognition using single physiological modalities, such as Cheng et al. [111] classified emotions based on EEG data using SVM and DT classifiers, achieving average accuracies of 87.11% and 75.89%, respectively, for arousal, valence, and dominance from the DREAMER [135] dataset. Tuncer et al. [252] introduced LEDPatNet19, a novel SVM-based network for EEG data from the DREAMER [135] dataset. The proposed LEDPatNet19 network featured multilevel feature fusion and achieved accuracies of 94.44%, 94.58%, and 92.86% for valence, arousal, and dominance, respectively. Sarma et al. [194] classified emotional states (happy, sad, and neutral) using EEG signals from the SEED [299] dataset with SVM, KNN, and RF classifiers, achieving average accuracies of 81.90%, 89.13%, and 87.30%, respectively. Gao et al. [206] proposed a classification model for neutral, happy, and sad emotions using feature fusion of power spectrum and wavelet energy entropy from EEG signals of SEED [299] and DEAP [233] datasets. The authors achieved accuracies of 89.17% with SVM and

91.18%. Hasnul et al. [167] proposed a uni-modal emotion recognition system using ECG signals and four ML algorithms (KNN, SVM, RF, DT). The authors validated the proposed model on the AMIGOS [125], DREAMER [135], and a self-collected dataset (A2ES), achieving an average accuracy of over 90% for arousal and valence.

Some works on multi-modal physiological signals combining two or more physiological signal data have also been implemented using ML models, such as Sharma et al. [145] developed KNN with K=3 and SVM with RBF kernel for EEG and ECG signals. The authors showed that KNN outperforms SVM in terms of accuracy on the DREAMER [135] and AMIGOS [125] datasets. Younis et al. [77] introduced a Stacking classifier for emotion recognition with a self-collected dataset, incorporating multi-modal physiological signals including SKT, EDA, and HR, achieving accuracies of 93%, 94%, 97.50%, 97%, and 98.20% using KNN, SVM, RF, DT, and stacking, respectively. Awaan et al. [267] proposed an ML-based emotion recognition classifier using SVM, RF, and LSTM stacking, combining EEG, ECG, and GSR signals. They validated the model on the AMIGOS [125] dataset with K-fold cross-validation (K=10), achieving an average accuracy of 94.5%. Table 4.1 summarises the ML approaches used with various physiological signals for automated emotion recognition.

4.2 Deep Learning Models for Emotion Recognition using Physiological Signals

Numerous studies have utilized deep learning methods independent of specific features, such as Attention-based Convolutional Recurrent Neural Networks (ACRNN) [266], Convolutional Neural Networks (CNN)[173], Dynamical Graph Convolutional Neural Networks (DGCNN)[251], Long Short-Term Memory (LSTM) networks [98], Multi-layer Perceptrons (MLP)[96], and Recurrent Neural Networks (RNN)[287] for emotion classification tasks. CNNs are deep, feed-forward neural networks that generate weights and exhibit high translation invariance, and is the most widely adopted neural network (as shown in Figure 4.2) for emotion classification.

Table 4.1: Summary of ML models for automated ERS.

Reference	Year	Dataset	Physiological Signals	Classification Model	Average Accuracy			Modality	ED[1]	DP[2]
					Arousal	Valence	Dominance			
Nasoz et al. [84]	2003	Self-collected dataset	HR, GSR	KNN	Avg - 72%			Dual	2D	No
Koelstra et al. [233]	2011	DEAP [233]	EEG	GNB	62%	57.60%	-	Single	2D	No
Subramanian et al. [221]	2016	ASCERTAIN [221]	GSR	NB	66%	68%	-	Single	2D	No
			ECG		59%	60%	-			
			EEG		61%	60%	-			
			EEG, ECG, GSR		67%	69%	-	Multi		
Ferdinando et al. [93]	2016	AMIGOS [125]	ECG	KNN - 10 fold CV	59.07%	55.80%	-	Single	2D	No
				KNN + LOSO	58.70%	59.20%	-			
Katsigiannis et al. [135]	2018	DREAMER [135]	EEG	SVM - RBF	62.17%	62.49%	61.84%	Single	3D	No
			ECG		62.37%	62.37%	61.57%			
			EEG, ECG		62.32%	61.84%	61.84%	Dual		
Zheng et al. [300]	2017	SEED [299]	EEG	SVM	Avg - 79.28% (positive, negative, neutral)			Single	2D	No
		DEAP [233]			69.67% (Arousal, Valence)		-			
Cristian et al. [256]	2017	Mahnob - HCI [161]	EEG	SVM	Avg - 65.29% (Arousal, Valence)			Single	2D	No
Wiem et al.[168]	2017	Mahnob - HCI[161]	ECG	SVM -Linear	51.40%	44.05%	-	Single	2D	No
			GSR		47.36%	48.93%	-			
			RESP		47.36%	53.19%	-			
			TEMP		40.14%	43.30%	-			
Li et al.[269]	2018	DEAP[233]	EEG	SVM + LOSO	Avg - 59.06%			Single	2D	No
		SEED[299]	EEG	SVM + LOSO	Avg - 83.33% (positive, negative, neutral)			Single	2D	No
Cheng et al.[111]	2020	DREAMER[135]	EEG	DT	75.74%	75.53%	76.40%	Single	3D	No
				SVM	87.03%	87.14%	87.18%			
Bota et al.[185]	2020	WESAD [195]	EDA	RF	85.78%	92.86%	-	Single	2D	No
			ECG		85.75%	92.86%	-			
			BVP		85.78%	94.39%	-			
			RESP		85.78%	92.86%	-			
Tuncer et al.[252]	2021	DREAMER[135]	EEG	SVM	94.58%	94.44%	92.86%	Single	3D	No
Galvão et al.[82]	2021	DREAMER [135]	EEG	KNN (K=1)	93.72%	92.16%	-	Single	2D	No
				RF	93.79%	93.65%	-			
Sharma et al.[145]	2021	DREAMER[135]	EEG, ECG	KNN (K=3)	92.06%	92.3%	92.38%	Dual	3D	No
				SVM (RBF)	92.01%	91.53%	92.29%			
		AMIGOS [125]	EEG, ECG	KNN (K=3)	92.01%	85.26%	89.65%	Dual	3D	No
				SVM (RBF)	89.87%	81.97%	87.31%			
Khateeb et al. [154]	2021	DEAP[233]	EEG	SVM - 10 fold CV	Avg - 65.72%			Single	2D	No
				SVM + LOOCV	Avg - 65.92%					
Sarma et al. [194]	2021	SEED[299]	EEG	SVM	Avg - 81.90% (positive, negative, neutral)			Single	3D	No
				RF	Avg - 87.30%					
				KNN	Avg - 89.13%					
				RF	81.31%	81.52%	-			
		DREAMER [135]		KNN	88.31%	86.40%				
				RF	84.69%	84.73%	-			
Rupal et al.[212]	2022	DEAP [233]	EEG + Peripheral	KNN	86.64%	86.40%	-	Multi	2D	No
		SEED [299]		Ensemble (DT)	Avg- 95.7% (positive, negative, neutral)					
		DEAP [233]			84.3%	83.9%	-			
Gao et al.[275]	2022	SEED [299]	EEG	Graphical CNN (RF)	Avg - 70.88% (positive, negative, neutral)			Single	2D	No
		DEAP [233]			66.37%	63.93%	-			
Younis et al. [77]	2022	Self-collected dataset	SKT, EDA, HR	KNN(K=5)	Avg - 93%			Multi	3D	No
				SVM(RBF)	Avg - 94%					
				RF	Avg - 97.50%					
				DT	Avg - 97%					
				Stacking	Avg - 98.20%					
Awaan et al. [267]	2022	AMIGOS [125]	EEG, ECG GSR	Ensemble (SVM+RF+LSTM)	Avg - 94.5% (Arousal, Valence)			Multi	2D	No
Hasnul et al. [167]	2023	AMIGOS [125], DREAMER [135] Self-collected(A2ES)	ECG	SVM(RBF) KNN(K=3) RF DT	Avg - 90% (Arousal, Valence)			Single	2D	No

[1]Emotion Dimension(ED)

[2]Data Privacy (DP)

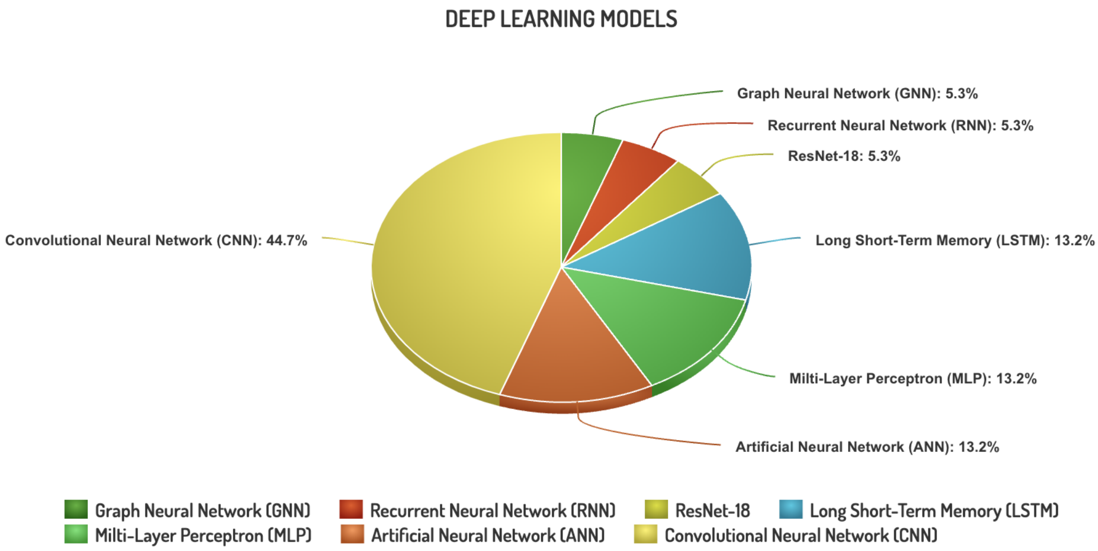


Figure 4.2: Distribution of DL models for automated Emotion Recognition System (ERS).

Related Works: Santamaria-Granados et al. [143] used CNN to extract features from ECG and GSR data, classifying arousal and valence emotional states using the AMIGOS [125]. Martinez et al. [101] effectively employed a Deep Convolution Neural Network (DCNN) on Skin Conductance and BVP to differentiate four emotional states: fun, excitement, relaxation, and anxiety. Topic et al. [40] used CNN on EEG data of the DREAMER [135] dataset and classified three emotional states with accuracies of 85.10%, 81.25%, and 85.10%, for arousal, valence and dominance respectively. The authors proposed a novel method based on feature maps for emotion recognition. They create feature maps based on the topographic (TOPO-FM) and holographic (HOLO-FM) representations of EEG signals [40]. The proposed approach tested four datasets, among which DREAMER [135] outperforms the others in terms of accuracy.

As evidenced by multiple examples in the present literature, CNN outperformed traditional ML algorithms for emotion classification and is directly proportional to emotion prediction performance [31, 266]. CNN is combined with other classifiers to improve the model's overall performance for classifying different classes of emotions [52].

Yang et al. [287] propose a hybrid network combining CNN and RNN to classify emotion states using a spatial-temporal representation of raw EEG data streams of the DEAP [233] dataset. The author also used LSTM to extract the textual information from participants' feedback. The network achieved accuracy for arousal and valence as 90.81% and 89.92%, respectively. Tao et al. [266] proposed a novel model named ACRNN based on CNN for extracting more discriminating features of EEG data. The authors used the publicly available DREAMER [135] dataset for classifying arousal and valence emotional states with accuracies of 93.38% and 93.72%, respectively. Iyer et al. [31] developed an Ensemble-hybrid model for emotion recognition using CNN, LSTM, a hybrid of CNN-LSTM, and a Stacking ensemble. Out of all, stacking achieves the best classification results. The proposed model is validated with two benchmark datasets, SEED [299] and DEAP [233], using EEG physiological signals, achieving an average accuracy of 97.16% and 65% for SEED and DEAP, respectively. Chakravarthi et al. [48] proposed a hybrid model using a CNN-based pre-trained ResNet-152 (Residual network) algorithm and LSTM. The authors employed EEG physiological signals to recognize humans' emotions and behavioural changes. The proposed model is validated on the SEED [299] dataset, achieving an average accuracy of 98% for positive, negative and neutral emotions.

Zali et al. [52] proposed a subject-independent human emotion recognition system. It uses multi-channel EEG signals from SEED [299] and DEAP [233] datasets. The authors employed different pre-trained models such as AlexNet, VGG19, ResNet18, Inception-V3, and EfficientNet-b0, among which Inception-V3 performs best with the SVM classifier. The proposed model achieved the highest average accuracy of 94.58% for SEED [299] and 90.7% for DEAP [233]. Singh et al. [172] proposed an emotion recognition framework using EEG physiological signals. The proposed model uses a deep convolution neural network (DCNN) for feature extraction and eight other classifiers (SVM, KNN, DT, NN, LSTM, AdaBoost, NB, RF) for classifying binary, quad, and octal emotion states. This framework validates two datasets, DREAMER [135] and DEAP [233], in two different scenarios, i.e. subject dependent and subject-independent. The best results are from subject-independent scenarios, with an average of 97.97% for DREAMER and 99.36% for DEAP.

Li et al. [263] proposed a novel model for emotion recognition using spatial and temporal streams of EEG physiological signals. The proposed model used MLP to classify emotions in valence and arousal states. It is validated in a subject-independent scenario with two datasets, DREAMER [135] and DEAP [233], achieving an average accuracy of 63.06% and 62.49% for DREAMER [135] and DEAP [233] respectively.

Table 4.2: Summary of DL models for automated ERS.

Reference	Year	Dataset	Physiological Signal	Classification Model	Average Accuracy			Modality	ED¹	DP²
					Arousal	Valence	Dominance			
Tang et al.[98]	2017	SEED[299]	EEG	Bimodal - LSTM		Avg - 93.97%		Single	3D	No
Song et al.[251]	2018	DREAMER[135]	EEG	DGCNN	84.54%	86.23%	85.02%	Single	3D	No
Yang et al.[287]	2018	DEAP[233]	EEG	CNN - RNN	90.81%	89.92%	-	Single	2D	No
Santamaria et al.[143]	2018	AMIGOS[125]	EEG	DCNN	81%	71%	-	Dual	2D	No
			ECG		71%	75%	-			
Ali et al. [151]	2018	Self-collected dataset	ECG,GSR	CNN		Avg - 71.05%		Dual	3D	No
Liu et al.[264]	2019	SEED[299]	EEG	DCCA		Avg - 94.58%		Single	3D	No
Dar et al.[173]	2020	DREAMER[135]	EEG, ECG GSR	CNN+LSTM		Avg - 90.8%		Multi	3D	No
Bhattacharyya et al.[27]	2020	DREAMER[135]	EEG	DNN	79.95%	79.95%	79.95%	Single	3D	No
Cheng et al.[111]	2020	DREAMER[135]	EEG	DNN	88.95%	87.23%	88.20%	Single	3D	No
		DEAP[233]			97.53%	97.69%	-			
Sarkar et al.[193]	2020	DREAMER[135]	ECG	Self - Supervised CNN	85.90%	85%	-	Single	2D	No
		WESAD[195]				95% (Affect state)				
Dar et al.[173]	2020	AMIGOS[125]	EEG, ECG GSR	CNN+LSTM	Avg - 99.0%			Multi	3D	No
Tao et al.[266]	2020	DREAMER[135]	EEG	ACRNN	93.38%	93.72%	-	Single	2D	No
Nath et al.[70]	2020	DEAP[233]	EEG	CNN+PCA	81.20%	84.30%	-	Single	2D	No
Topic et al.[40]	2020	DREAMER[135]	EEG	CNN	85.10%	81.25%	85.10%	Single	3D	No
Zhang et al.[255]	2020	CASE[134]	EDA	CorrNet	74.03%	76.37%	-	Single	2D	No
Yang et al. [291]	2020	Self-collected dataset	ECG	CNN		Avg - 75.4%		Single	3D	No
Deng et al.[268]	2021	DEAP[233]	EEG	3D-CNN	91.94%	92.49%	-	Single	2D	No
		SEED[299]				Avg - 99.19%				
Tan et al.[59]	2021	Mahnob-HCI[161]	EEG	CNN	79.39%	72.12%	-	Single	2D	No
Zhang et al.[288]	2021	Mahnob-HCI[161]	EEG	CNN	70.17%	71.38%	-	Single	2D	No
				HFCNN	88.28%	89%	-			
Liakopoulos et al.[146]	2021	WESAD[195]	ECG	CNN+LOSO		Avg - 82.35%		Single	3D	No
Bhatti et al.[28]	2022	WESAD[195]	EDA, ECG, SKT, RESP	CNN+LOSO		Avg - 89.57%		Multi	3D	No
		CASE[134]	EDA, BVP, SKT		71.0%	-	-		1D	No
		DREAMER [135]			83.04%	80.34%	82.50%		3D	
		DEAP [233]			78.44%	77.50%	79.38%			
Kumari et al.[176]	2022	AMIGOS [125]	EEG	CNN (EmotionCapsNet)	80.07%	79.06%	79.69%	Single	3D	No
		DREAMER [135]			91.51%	91.19%	-			
		DEAP [233]			99.58%	99.23%	-			
Anuragi et al.[26]	2022	SEED [299]	EEG	ANN		Avg- 78.1% (positive, negative, neutral)		Single	2D	No
		DEAP [233]			83.5%	82.1%	-			
Gao et al.[275]	2022	SEED [299]	EEG	Graphical CNN (GCNN)		Avg - 85.65% (positive, negative, neutral)		Single	2D	No
		DEAP [233]			81.95%	81.77%	-			
Iyer et al.[31]	2023	SEED [299]	EEG	CNN+LSTM (Stacking)		Avg - 97.16% (positive, negative, neutral)		Single	2D	No
		DEAP [233]				Avg - 65%(Arousal, Valence)	-			
Chakravarthi et al. [48]	2023	SEED [134]	EEG	CNN-LSTM+ ResNet-152 (Hybrid)		Avg - 98% (positive, negative, neutral)		Single	2D	No
Zali et al.[52]	2023	SEED [299]	EEG	CNN (Inception-V3)+ SVM		Avg - 94.58% (positive, negative, neutral)		Single	2D	No
		DEAP [233]			-	Avg - 90.7% (positive, negative)	-			
Singh et al.[172]	2023	DREAMER [135]	EEG	DCNN	97.64%	96.74%	99.55%	Single	3D	No
		DEAP [233]			99.58%	99.23%	99.28%			
Li et al.[263]	2023	DREAMER [135]	EEG	MLP+Self-Attention	64.25%	61.88%	-	Single	2D	No
		DEAP [233]			61.89%	63.10%	-			
Bagherzadeh et al. [227]	2024	SEED [134]	EEG	ResNet-18		Avg - 81.25% (positive, negative, neutral)		Single	2D	No

[1] Emotion Dimension(ED)

[2] Data Privacy (DP)

Deep learning algorithms eliminate the need for data pre-processing and feature extraction, which are the most time-consuming elements of the emotion classification system. Deep Learning reduces the extra dimensions and removes the noises in the data signals using methods like Autoencoders and Neural Networks, which have a high success rate for emotion recognition. According to past research, Deep Learning approaches are well-suited to dynamic simulation and improvised feature extraction. However, they are not well suited for human physiology-based recognition as these methods are limited to working as a black box [101]. They take extensive data to train and are exceedingly computationally costly. Table 4.2 summarises the Deep Learning approaches used with various physiological signals for automated emotion recognition.

4.3 Privacy Concerns in traditional ML and DL for ERS using Physiological signals

The automated ERS using traditional ML and DL methods requires complete raw data for training the model, resulting in a significant loss of privacy protection for sensitive physiological information [36, 45, 100]. However, feeding the physiological signals to the model directly allows the third party and data attackers to access the sensitive data, leading to a consequential loss of privacy. Conventional ML and DL models for ERS based on physiological signals allow multiple users to access each other's emotional state [59, 82, 145, 154]. Several potential challenges emerge as a result of privacy concerns with automated ERS that utilize sensitive physiological signals:

- ***Multiple data leakage*** incidents recently happened with the automated ERS [95, 164, 301]. In 2019 [25], millions of users' biometric data and records got leaked. Their fingerprint scans and face recognition records were publicly accessible.

- ***External hacking*** attempts leading to data breaches result in the unauthorized access or disclosure of sensitive physiological data utilized by the ERS [123, 270]. Such occurrences have the potential to significantly impact individuals' privacy and erode trust in the system.

- ***Inference attacks*** occur when adversaries derive sensitive information about individuals from seemingly innocuous data or system outputs [38, 115]. These attacks can infer individuals' emotional states by exploiting patterns in physiological signals.

- ***Model Inversion Attacks (MIA)*** involve reverse-engineering the ERS model to extract sensitive information about individuals from its outputs [23, 24]. This method allows adversaries to infer details about individuals' physiological responses or emotional states from the output predictions made by the ERS.

The acquisition and management of sensitive physiological data from wearable sensors in an automated ERS pose challenges of privacy considerations due to a centralized environment [36, 45, 100]. Conventional ML and DL models for automated ERS allow multiple users to access the raw physiological data while training, as it involves centralized data processing and predictions [82, 145, 154]. The possibility of external parties accessing the sensitive physiological data containing emotional information can lead to further emotional distress. In such circumstances, individuals become less inclined to disclose their personal sensitive physiological information. Hence, preserving the privacy of sensitive physiological data is essential so that some third party without authorization cannot access it [160].

4.4 Summary

Various ML and DL methodologies have been used for classifying emotional states using physiological signals. ML approaches focus on training models to identify patterns in raw data and extract meaningful insights, with SVM being the most commonly used model for ERS. In contrast, most studies leveraging DL techniques do so without relying on specific feature extraction, with CNNs being the most widely adopted neural architecture for emotion classification. The privacy concerns are reported, including data leakage, external hacking, inference attacks, and model inversion attacks in the traditional ML and DL models while recognizing emotions.

Chapter 5

Literature Review of Federated Learning Paradigm for ERS

This chapter comprehensively reviews the literature on the FL paradigm, particularly as it applies to Emotion Recognition Systems using physiological signals. It explores the evolution of FL, a distributed approach that enables machine learning models to be trained across decentralized data sources, thus preserving privacy by keeping sensitive data on client devices. This chapter critically analyzes various FL implementations and techniques used for ERS, examining their effectiveness in addressing key challenges, such as data privacy and model performance, and highlights gaps in the current literature.

5.1 Introduction to Federated Learning Paradigm

Federated Learning (FL) is a distributed paradigm for training machine learning models on datasets stored locally on client devices without collecting the complete raw data from the devices [272, 273]. Machine learning is used in physiological signals-based Emotion Recognition Systems (ERS) to help automate emotion recognition tasks. Traditional classifiers for automated ERS strive to achieve high accuracy, but they do not preserve user's sensitive information (physiological data) since they require access to the complete raw data. However, physiological data is personal and sensitive and requires developing a privacy-preserving environment. FL offers a solution to this

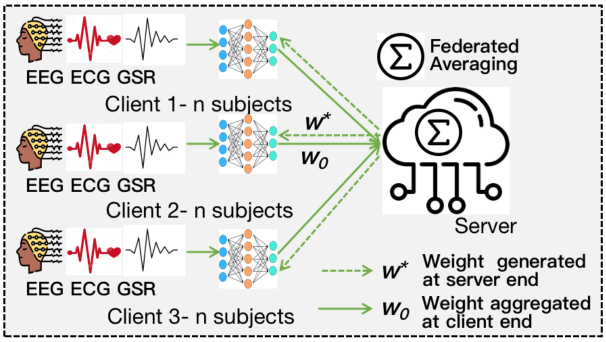


Figure 5.1: Federated Learning for Emotion Recognition System (ERS) with physiological signals.

problem by allowing the development of a privacy-preserving global aggregator [51] without gaining direct access to the user's raw, sensitive physiological data.

Figure 5.1 shows the Federated Learning setup for ERS using physiological signal data. Each client device stores wearable sensor physiological data locally. The machine learning model is trained on this local data at the client end, and only the model updates (such as gradients or weights) are sent to the central global server for further aggregation from all the client updates. The raw data never leaves the client end, reducing the risk of data breaches and unauthorized access. The client and server end are described in detail in the next subsection.

5.1.1 Types of Architecture of FL

5.1.1.1 Client-Server Architecture

In a client-server architecture, the data owners are known as clients. Users (or groups of users) with wearable sensors act as clients that run the local model at their end [36, 100, 149]. The local model at the client end generates the model updates (gradients) to be sent to the global server [69, 136, 214]. The server then aggregates the updates (gradients) and sends an updated version to all clients. After receiving the updates, the clients use them for further training iterations on their local data [136]. Figure 5.2 (A) presents the Client-Server architecture of FL, showing the movement of model gradients between the client and server end.

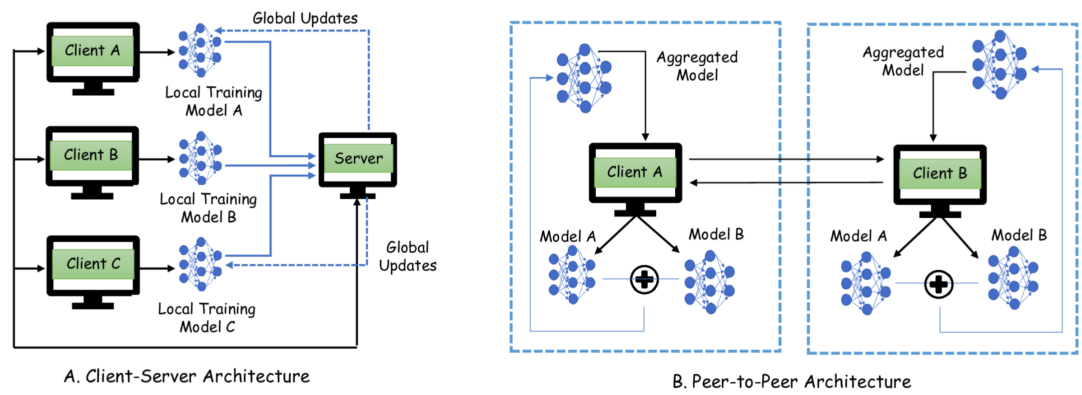


Figure 5.2: Architectures of Federated Learning: (A). Client-Server Architecture, (B). Peer-to-Peer Architecture

5.1.1.2 Peer-to-Peer Architecture

In a peer-to-peer architecture, multiple clients (peers) collaborate to train a machine-learning model without relying on a central server to coordinate the process. In this architecture, each client communicates directly with others to share model updates and collectively build a global model [280, 295]. Each client in this architecture has a local training model, which they aim to aggregate from nearby client model updates [32, 197, 253]. The participating clients in this architecture (as shown in Figure 5.2. B) do not require a third-party coordinator.

When the number of clients increases in a peer-to-peer architecture, coordination becomes complex as no central server exists [43, 280, 295]. Clients coordinate with each other, creating an extensive communication overhead. Also, disagreements or inconsistencies exist between clients in a peer-to-peer architecture, which slows down or even halts the training process. However, in client-server architecture, communication overheads are comparatively lower. Hence, this thesis adopts the client-server architecture to propose FL-based emotion recognition architectures.

5.1.2 Approaches of FL

Federated learning has three forms of data partitioning based on the statistical sharing among clients in characteristics and feature domains described below.

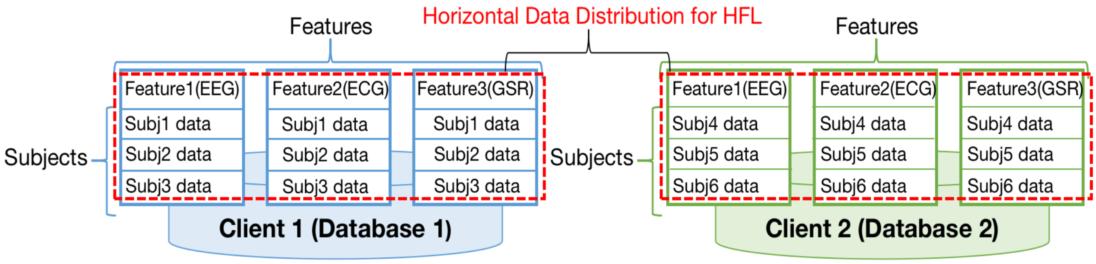


Figure 5.3: Horizontal data distribution of physiological data from two clients.

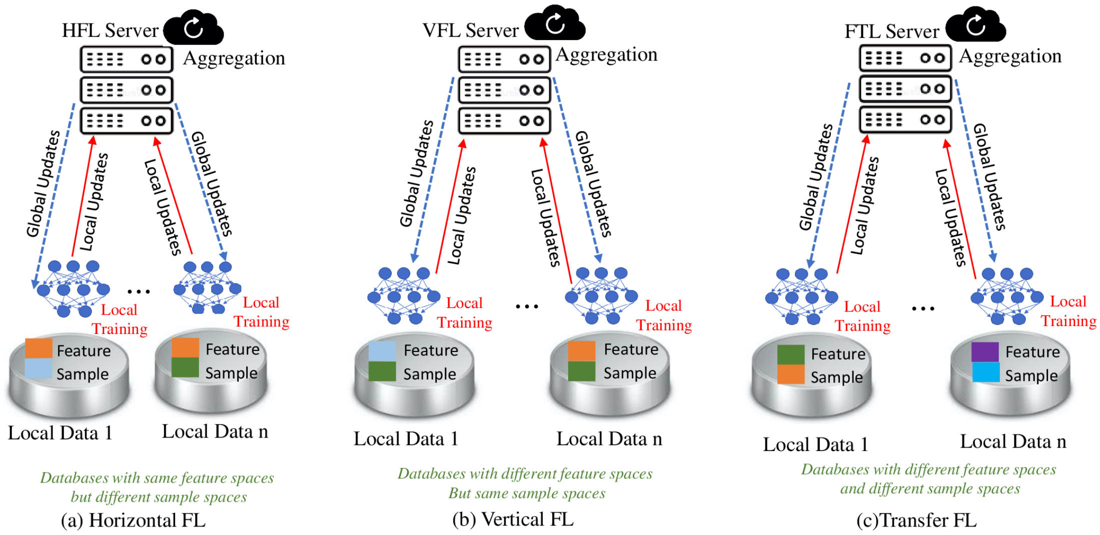


Figure 5.4: (a) Horizontal FL, (b) Vertical FL and (c) Federated Transfer Learning

5.1.2.1 Horizontal Federated Learning (HFL)

Horizontal Federated Learning (HFL) is a federated framework for horizontal data partitioning in a decentralized environment [51]. It involves collaboration among multiple clients with datasets having the same feature set and different samples [67]. Each participating client has data on different subjects with the same set of features [36, 45]. Figure 5.3 gives an illustrative example of horizontal data partitioning for physiological data from multiple subjects residing into two clients. It shows that different subjects have the same features at each client end, such as Subj1-Subj3 with Feature1-Feature3 on Client 1 and Subj4-Sub6 with the same features on Client 2. In horizontal data partitioning, data across different rows share the same features aligned by features. Figure 5.4 (a) presents how HFL works with horizontal data partitioning having the same feature space and different sample space.

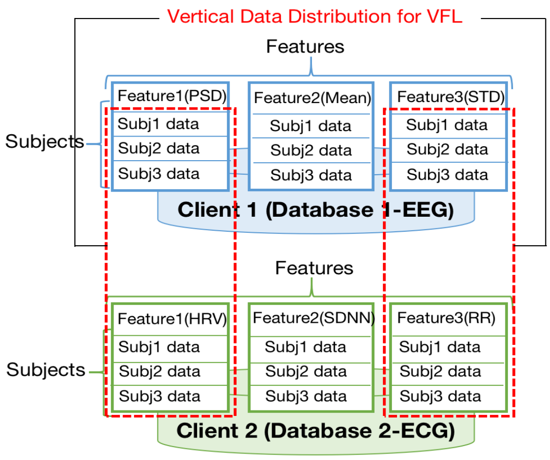


Figure 5.5: Vertical data distribution of physiological data from two clients.

5.1.2.2 Vertical Federated Learning (VFL)

On the other hand, Vertical Federated Learning (VFL) is a federated framework for vertical data partitioning in a decentralized environment. It involves collaboration among multiple clients with datasets with different feature spaces but overlapping samples [49, 211, 284]. In vertical data partitioning, each participating client has data on the same entity but with different types of information (features) [225, 249]. Figure 5.5 gives an illustrative example of vertical data partitioning for physiological data from multiple subjects residing into two clients. It shows that the same subjects overlap with different features at each client end, such as Subj1, Subj2 and Subj3 overlapping with different features from EEG signal data on Client 1 and ECG signal data on Client 2, respectively. In vertical data partitioning, data across rows is the same, but the features in columns are different [211, 281]. For example, two medical clinics wish to collaborate on training an ML model. The data from each clinic is different, but they have the same patients, and due to privacy and security issues, they cannot disclose their information.

5.1.2.3 Federated Transfer Learning (FTL)

Federated Transfer Learning (FTL) combines transfer learning principles with federated learning. It enables collaborative learning across clients with different datasets in terms of features and samples (as shown in Figure 5.4 (c)). Unlike traditional HFL or VFL, the datasets at different clients do not fully overlap in the features or the samples [207, 225, 236, 249]. Transfer learning involves reusing and fine-tuning machine learning models learned on a dataset from one domain (existing at one client) to solve a problem in a second domain (existing at another client). Liu et al. [279] introduced FTL as a new technique and framework for improving statistical modelling in data federation. Chen et al. [272, 273] created a framework FedHealth to collect data from various organizations and use FTL to give tailored healthcare services to the users.

This thesis uses HFL to propose FL-based emotion recognition architectures, as the physiological data from EEG, ECG, GSR, and RESP signals form a fused feature vector from different subjects, giving the same feature set to all subjects as columns. These subjects are partitioned into different clients with the same features to create a federated environment to propose a multi-modal emotion recognition architecture.

5.1.3 Aggregation Algorithm of FL

The aggregation algorithms explain the formation of a global aggregator by aggregating updates from the clients with their local servers engaged in the training session. A coordinator aggregates the model parameters centrally using an averaging algorithm in the client-server model (as shown in Figure 5.2. (A)). The different aggregation algorithms of FL are discussed in this section as follows:

- **Federated Averaging (FedAvg) [136]:** Google introduced the FedAvg algorithm to aggregate the model updates received from all the clients at the server via a client-server architecture of FL. FedAvg aggregates the gradients (at the server end) produced by Stochastic Gradient Descent (SGD) (at each client end) using the computation given by equation 5.1. Following aggregation, the local participating client receives the aggregated model updates, which repeat until the model converges or the number of iterations is complete. It is noticeable that the FedAvg [136] is the only algorithm used for emotion recognition [36, 45, 165].

$$w_t^g = \frac{1}{N_{total}} \sum_{i=1}^{N_{total}} w_{t,i}^l \tag{5.1}$$

  Where w_t^g is the global model created in time t, $w_{t,i}^l$ is the local model received from all clients in time t, and N_{total} is the total number of the local model received

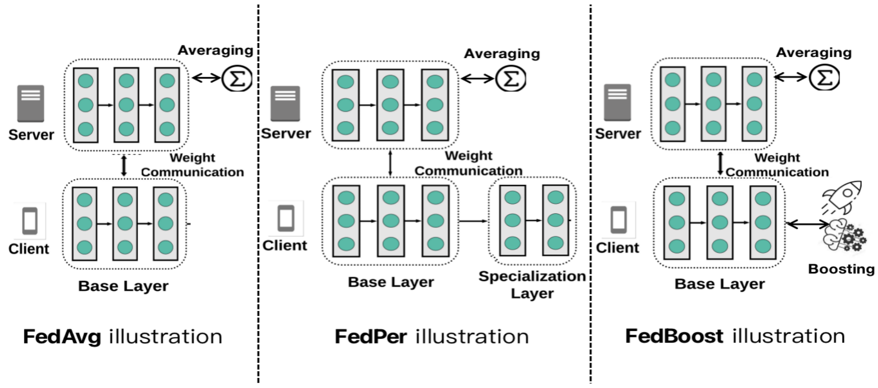


Figure 5.6: Illustration of different Federated Learning aggregation algorithms.

at the global server.

- **Federated Boosting (FedBoost) [114]:** It extends the traditional boosting technique to be applied in a federated learning setting. Boosting is an ensemble technique to build a robust classifier by combining weak classifiers sequentially. Fedboost allows boosting for each local model running at the client end. The local model updates from the client are then sent to the server for aggregation. It improves the overall model's performance by introducing an ensemble strategy for the training data at the client end. This algorithm provides computational speedups, faster convergence, along with data privacy.

- **Personalized Federated Learning (FedPer) [170]:** This FL method is for personalized model training in a privacy-preserving manner. The FedPer approach involves splitting the model into the base and personalized layers (as shown in Figure 5.6). In this method, only the base layers are transmitted to the server for aggregation, while the personalized layers remain undisclosed and are not shared. The rationale for this approach is that the base layers primarily concentrate on feature learning, facilitating efficient sharing among clients through aggregation. Conversely, the upper layers play a more significant role in classification or decision-making, providing customization that aligns with each client's local data.

- **Dynamic Weighted Federated Averaging (DWFA):** Chen et al. [110] proposed a novel dynamic weighted federated averaging algorithm for handing the imbalance in the data at the client end. It is an extension of the traditional

Table 5.1: Characteristics comparison between FL aggregation algorithms.

FL Methods → **Measures ↓**	**FedAvg** [136]	**FedPer** [170]	**FedBoost** [114]	**DWFA** [110]
Scalability	Yes	No	No	Yes
Communication Overhead	Yes	Yes	Yes	Yes
Aggregation	averaging weights and parameters	averaging Personalized weights	averaging Boosted weights	dynamically weighted averaging of weights
Privacy	Protects privacy of client data	Client data is exposed to central server	Client data is exposed to central server	Protects privacy of client data
Computational Complexity	Low	Higher	Higher	Medium
Use case	Homogeneous data distribution	Heterogeneous data distribution	Heterogeneous data distribution	Heterogeneous data distribution

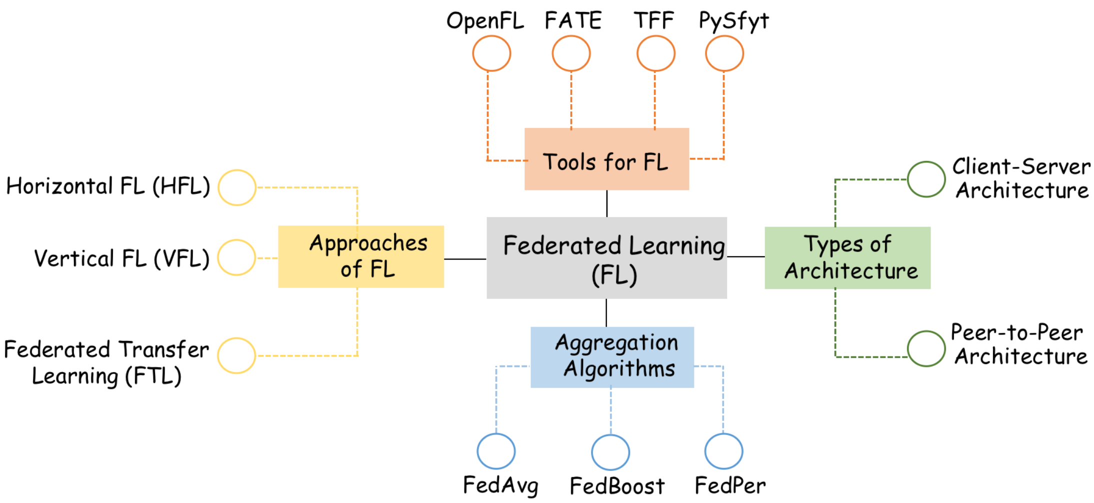


Figure 5.7: Components of Federated Learning

FedAvg algorithm aimed at improving the performance of federated learning environments that deal with heterogeneous data distributions across clients. DWFA adjusts the weights dynamically based on the significance of the client's contribution. For instance, a client with a larger dataset or better model accuracy might be assigned a higher weight during aggregation, allowing the global model to benefit more from valuable client models.

Table 5.1 compares the different federated learning aggregation algorithms based on their computation overheads, computational complexity, data privacy, and ability to address data heterogeneity. Figure 5.7 gives a visual representation of all four components of federated learning, including the type of architectures, different approaches of FL, different aggregation algorithms of FL and tools required for implementing FL.

Table 5.2: Tools for Federated Learning

Tool	Developer	Description	Data Partitioning
TFF[208]	Google	TensorFlow provides a friendly and adaptive framework for users to simulate distributed computing locally using TFF.	Horizontal
PySyft [250]	OpenMined	With PyTorch, it uses Federated Learning, Differential Privacy (DP), Model Predictive Control (MPC) to separate private data from model training.	Horizontal, Vertical
FATE [22]	Webank	FATE is part of the Federated AI Ecosystem, which creates safe computing protocol using homomorphic encryption and MPC.	Horizontal, Vertical
OpenFL [67]	Intel	Intel created Open Federated Learning to implement FL on sensitive data. It contains bash deployment scripts and uses certificates to secure communication.	Horizontal, Vertical
Flower [1]	ETH Zurich	Flower provides a high-level API that makes setting up and running FL experiments easy. It is developed by a team of researchers at ETH Zurich and is open-source and available on GitHub. It can be easily used with PyTorch and TensorFlow.	Horizontal, Vertical

5.1.4 Popular Tools for Federated Learning Setup

Various open-source frameworks and software options are available when exploring federated learning, including TFF [208], PySyft [250], FATE [22], OpenFL [67], and Flower [1]. The best option heavily relies on the use case's purpose and characteristics. This thesis uses TFF [1] version 0.19.0 for all the experimental setups for FL. Table 5.2 gives all other useful tools with their descriptions.

5.2 Research Methodology

The literature review aims to collect information for analyzing and evaluating all significant emotion recognition technologies with federated learning concepts based on physiological signals. It consists of four research papers for ERS based on physiological signals using FL.

5.2.1 Research Questions

This review delves into the following research questions (RQ) mentioned in Table 5.3. The research questions help define the keywords for the inclusion criterion for reviewing the articles.

Inclusion Criteria: These databases were used to search for articles to perform literature review (as Figure 5.8 presents): ***Elsevier, Google Scholar, Springer Link, ACM Digital Library,*** and ***PubMed***.

[1]https://github.com/google-parfait/tensorflow-federated

Table 5.3: Research questions

RQs	Questions
RQ1	What makes privacy a vital consideration for physiological signal-based emotion recognition systems?
RQ2	How does federated learning ensure data privacy in emotion recognition systems using physiological data?
RQ3	What evaluation measures can assess federated learning environment for emotion recognition?

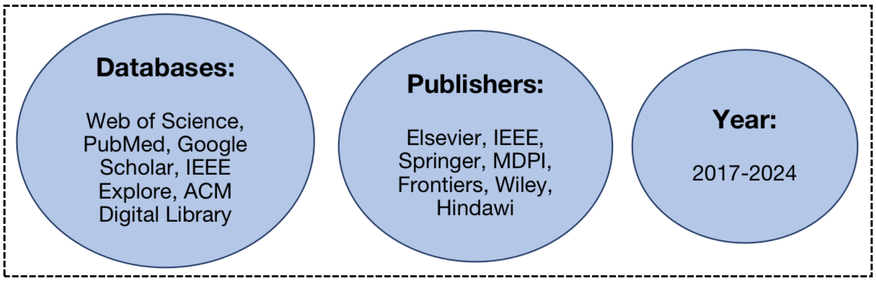


Figure 5.8: Search domains for reviewing.

Keywords: It uses the following keywords for searching: ***Emotion recognition, multi-modal emotion recognition, emotion recognition using physiological signals, bio-sensors, bio-signals, emotion recognition via wearable sensors, federated learning in smart healthcare, and emotion recognition using federated learning***. It considers the articles from journals (with indexing in *SCI, SCIE, and Scopus*) and good-ranking conferences published in English.

Exclusion Criteria: For reviewing the related work, articles not written in English are excluded. Also, it does not include articles based on the statistical evaluation of physiological signals.

5.3 Related work of FL in Emotion Recognition

Federated learning has been implemented to preserve the privacy of peripheral physiological data in emotion recognition systems. Only four research studies utilize federated learning to safeguard the privacy of sensitive physiological signals while performing emotion recognition by creating a decentralized framework. Table 5.4 summarizes these studies, giving the combination of physiological signals used, type of federated learning approach, experimental settings, and their achieved results.

Table 5.4: Recent studies on emotion recognition based on physiological signals using Federated Learning.

Reference	Dataset	Physiological Signal	Base Model	Average Accuracy	FL Approach	Tool	FL Algo	Modality
Nandi et al. [36]	DEAP[233]	EDA, RB	FFNN	81.92 %	Horizontal FL	TFF	FedAvg	Dual
Tara Hassani [100]	CASE[134]	GSR	CNN	79%	Horizontal FL	TFF	FedAvg	Single
Gu et al. [248]	DEAP[233], DREAMER [135]	EEG	CNN - LSTM	88.27%, 92.24%	Transfer FL	TFF	-	Single
Agrawal et al. [149]	DEAP[233]	EEG	CNN	69.77%	Horizontal FL	TFF	FedAvg	Single

Nandi and Xhafa [36] developed Fed-ReMECS, a real-time emotion classification framework using FL that leverages multi-modal physiological data from EDA and RESP wearable sensors. The authors classify nine emotional states from the DEAP [233] dataset using two domains of Arousal and Valence. Tara [100] employed the physiological signal GSR on the CASE [134] dataset to train a federated CNN-based model for emotion recognition. According to the author's findings, the proposed federated CNN architecture performs similarly to a centralized CNN architecture in terms of accuracy. However, its federated learning nature allows it to protect users' sensitive information. Gu et al. [248] proposed a federated transfer learning environment for ensuring data privacy while recognizing emotions in a teacher-student network. The authors used the EEG signals from the DEAP [233] and DREAMER [135] datasets. Agarwal et al. [149] developed a decentralized model FedCER using federated learning for recognizing emotions in three dimensions. FedCER validated the EEG data from the benchmark DEAP [233] dataset. A 2D convolution neural network was used for feature extraction at the client end for the EEG data. The model proved to be as accurate as the non-FL environment (baseline 2dCNN) while preserving the privacy of the client's data.

The literature demonstrates that deep learning models are the most extensively studied approaches for integrating federated learning with physiological data in emotion recognition. Their ability to generate model weights through neural networks, coupled with optimizers designed for deep neural architectures, ensures faster convergence and more precise weight updates.

5.4 Answers to RQs

Based on the literature on FL for ERS using physiological data, the following research questions are answered:

- **RQ1:** What makes privacy a vital consideration for physiological signal-based emotion recognition systems?

  ANS: Physiological data is inherently personal and highly sensitive, encompassing signals such as EEG, ECG, GSR, and Heart Rate. Data from these signals are unique to each individual, revealing a lot about their health status and potentially providing insights into their emotional state, behaviour, and habits. Hence, they are considered sensitive. Compromising the privacy of physiological data can have profound repercussions on an individual's life. It opens the door for data attackers and exposes them to the risk of data breaches, which, in turn, can lead to various threats [21]. These threats include the potential exposure of an individual's health status, emotional stability, and even their biometric identity based on physiological signals. Therefore, safeguarding privacy is crucial when handling physiological data in an ERS.

- **RQ2:** How does federated learning ensure data privacy in emotion recognition systems using physiological data?

  ANS: FL is a promising approach that creates a decentralized environment with the local model at the client end and a global aggregator at the server end [69]. It allows the local model updates to be sent to the central server, combining them to create a global aggregator [39, 99, 124]. This approach does not allow the global aggregator to access the raw data used for training (at the client end). Hence, it preserves the privacy of sensitive physiological data (stored on the client's end).

- **RQ3:** What evaluation measures can assess federated learning environment for emotion recognition?

  ANS: FL is a distributed environment creating a local model at the client end and a global aggregator at the server end. Classification accuracy with scalability for different data distributions at client end measures the predictive performance and effectiveness of the FL models for emotion recognition [36, 100, 149, 248].

5.5 Limitations of existing Federated Learning in ERS

The review of related work on FL for emotion recognition systems aims to assess the current state of the literature and identify existing challenges. It examines the latest FL studies for emotion recognition based on physiological signals and reports the following limitations:

1. **Limited Emotion Classification:** Existing research on FL for ERS did not focus on intricate emotions encompassing the dominance dimension of the 3D-VAD emotion model. They fail to classify a broad spectrum of emotions and cover only limited emotions. They could only classify into basic positive and negative emotions using arousal and valence dimensions.

2. **Limited Modality:** Prior studies using FL for ERS have primarily conducted experiments utilizing either a single physiological signal or integrating only peripheral signals. These studies lack a comprehensive approach incorporating multiple physiological modalities, majorly not including EEG for emotion classification.

3. **Absence of Subject-Independence:** No previous research on Federated Learning for ERS validates for independent subjects (or clients). The generalization of models across different clients is not thoroughly explored in the existing works of FL for emotion recognition models. The variability across subjects is a crucial challenge in ERS, and ensuring that a model trained in a federated environment can generalize well to new, unseen subjects is important for real-world applicability.

4. **Limited Research of FL in Emotion Recognition:** The use of federated learning for emotion recognition based on physiological signals is a less explored area of research. Since FL came in 2017, very limited work has been published on emotion recognition using the federated paradigm (as mentioned in Section 5.1.3). These research works focused on only one federated aggregation algorithm - FedAvg [136]. Other federated learning algorithms are described in Section 5.1.3 Table 5.1, which have not been explored for emotion recognition.

Different algorithms like FedBoost [114] and FedPer [170] may have better comparison outcomes regarding data privacy and communication computations.

5. **Lacking Communication Computation Discussion:** Communication is a crucial barrier in a federated environment, as it concerns the amount of data transfer, communication overhead, and data privacy. Communication between the central server and the client devices is often a major bottleneck in federated learning, which has not been discussed in any of the previous FL works for ERS.

6. **Data Heterogeneity Not Addressed:** The existing architectures of FL for emotion recognition using physiological signals have not addressed the issue of data heterogeneity. Imbalanced Data Heterogeneity can occur in a federated environment at the server end due to the uneven data distributions at the client end.

7. **Evaluation & Assessment:** The existing model of FL for ERS has used only classification accuracy with scalability for different data distributions at the client end to measure the predictive performance and effectiveness of their proposed FL models for emotion recognition [36, 100, 149, 248]. No classification performance other than accuracy and communication computation (in terms of times) has been discussed.

5.6 Evaluation Measures for Assessing Proposed FL Architectures

This section presents the various evaluation measures, lacking in the existing FL works for ERS. These measures assess the performance of the proposed federated learning-based emotion recognition architectures (described later in chapters 6, 7, 8), which perform both binary (low/high) and octal (eight emotions) classifications. Evaluating the effectiveness of the emotion recognition models is critical in determining their practical applicability, particularly for real-time applications that involve physiological signal data. The following sections describe the different performance measures used in this thesis, along with their relevance to emotion recognition tasks.

Table 5.5: Confusion matrix representation for binary classification

	Predicted Positive	Predicted Negative
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

5.6.1 Classification Performance Measures

The classification performance of the proposed federated learning-based emotion recognition architectures is evaluated using widely adopted metrics, such as Confusion Matrix, Accuracy, Precision, Recall, and F1-score. These measures provide insight into how well the models can distinguish between different emotional states, both in binary classification (low/high emotion) and multi-class classification (eight emotions). Each of these measures is crucial in analyzing the model's ability to handle imbalanced data, misclassification costs, and overall reliability.

5.6.1.1 Confusion Matrix

The *Confusion Matrix* is a useful metric for visualizing the performance of classification models. It provides a matrix representation of the true versus predicted labels, showing how well the model performs for each emotion. Table 5.5 displays the counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) predictions in a tabular format, making it easier to see where the model performs well and where it makes errors.

From the confusion matrix, the following metrics can be derived:

- **Accuracy**: Accuracy is derived from the confusion matrix as the ratio of correct predictions to the total predictions. It represents the overall proportion of correctly classified instances.
- **Precision**: The ratio of true positive predictions to the sum of true and false positives for each class.
- **Recall**: The ratio of true positive predictions to the sum of true positives and false negatives for each class.
- **F1-Score**: The harmonic mean of precision and Recall for each class.

5.6.1.2 Classification Accuracy

Classification Accuracy is the most commonly used performance metric. It represents the proportion of correct predictions made by the model out of the total number of predictions. The formula for accuracy is given by Equation 5.2 as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5.2}$$

The TP presents the predicted emotion label (low/high for binary and one of the eight emotions for octal), which matches the actual emotion label (low/high for binary and one of the eight emotions for octal). FP presents the predicted emotion label (low/high for binary and one of the eight emotions for octal) that does not match the actual emotion label (low/high for binary and one of the eight emotions for octal). TN presents the predicted emotion label correctly, which corresponds to the absence of the actual emotion label. FN presents the predicted emotion label that fails to capture the presence of the actual emotion label for each dimension (valence, arousal, dominance) in a given instance.

While accuracy is useful for providing an overall sense of model performance, it may be misleading in cases of imbalanced datasets, such as when certain emotional states (e.g., neutral emotions) occur far more frequently than others (e.g., extreme emotions). In such cases, additional metrics, like precision and Recall, become more valuable for understanding the model's performance. A higher accuracy value indicates better performance of the emotion classification model.

5.6.1.3 Precision

Precision, also known as positive predictive value, is a metric that reflects the proportion of positive predictions that are actually correct. It focuses on the reliability of the model in terms of classifying positive instances. The formula for precision is given by Equation 5.3 as follows:

$$Precision = \frac{TP}{TP+FP} \tag{5.3}$$

High precision indicates a low false positive rate, meaning that the model is conservative in labelling positive instances and only labels instances as positive when it has high confidence. In the context of emotion recognition, this is particularly important when distinguishing between highly similar emotional states, where false positives could lead to misinterpretations. In the case of octal classification, where eight different emotions are considered, precision can be calculated for each individual class (emotion) independently and then averaged.

5.6.1.4 Recall

Recall, also referred to as sensitivity or True Positive Rate (TPR), measures the model's ability to correctly identify all actual positive instances. It is crucial in scenarios where missing a positive instance (false negative) is more costly than incorrectly labelling a negative one. The formula for Recall is given by Equation 5.4 as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5.4}$$

In emotion recognition systems, Recall is significant when the goal is to ensure that all instances of a particular emotion are detected. For instance, if the system is used to detect high-arousal emotions in real-time applications, failing to identify such states could be detrimental. Recall can also be calculated for each class individually in multi-class classification (octal emotions).

5.6.1.5 F1-Score

The *F1-Score* is a composite metric that balances precision and Recall, providing a single score to evaluate the model's overall performance. It is particularly useful when there is an uneven class distribution or when the cost of false positives and false negatives are not the same. The F1-score is the harmonic mean of precision and Recall, and it can be calculated as given in Equation 5.5:

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{5.5}$$

A higher F1 score indicates that the model has a good balance between precision and Recall, making it effective in identifying true positive instances while minimizing false positives and negatives. In the context of multi-class emotion recognition (octal classification), the F1 score can also be computed for each class and then provided as an average.

5.6.2 Scalability Measures

This thesis evaluates the scalability of the proposed federated learning-based emotion recognition architectures, demonstrating their ability to efficiently handle various data partitioning strategies and models while maintaining appropriate training times across diverse client data distributions. The training time and model accuracy are used to assess the scalability of the proposed architectures, as detailed in the subsections below.

5.6.2.1 Training Time

It refers to the time taken by the local model to learn from the updated global model gradients at the client end after each averaging round. It is computed mathematically as in Equation 5.6.

$$T = \frac{t_{c_1} + t_{c_2} + t_{c_3} + ...t_{c_i}}{i} \tag{5.6}$$

Where,

- T is the average of all clients' training time.
- t_{c_i} is the time the i_{th} client takes for local model training.

5.6.2.2 Model Accuracy

In federated learning, scalability refers to the model's ability to maintain its performance, specifically accuracy, as the number of clients (data sources) increases. A key factor is how the model handles data distribution across clients and continues to converge effectively over multiple communication rounds. The model's accuracy should be maintained even as the number of clients increases. It is presented (with different client distributions and iterations).

5.6.3 Communication Computation Measures

FL is a distributed machine learning technique that leverages data residing in multiple local models and aggregates them at a central global server. The evaluation of the efficiency of communication in this process is by measuring the below measures:

5.6.3.1 Averaging Time

It is the time required for the server to aggregate the model updates received from the client end, which impacts the communication efficiency of the FL model and is computed at the server end. Longer averaging time leads to increased communication latency and a higher risk of update delay and network congestion.

$$T_{agg} = T_{C_1} + T_{C_2} + T_{C_3} + ...T_{C_i} \tag{5.7}$$

Where,

- T_{agg} is the total aggregation time at the server.
- T_{C_i} is the time taken by the server to average the weights received from client $C_i (i = 1, 2, 3...i)$.
- i is the total number of clients contributing their updates.

5.6.3.2 Convergence Speed

The convergence speed of an FL model refers to the speed at which the model achieves its optimal performance, i.e., the point at which the model can no longer improve with additional training rounds. One way to measure the convergence speed of an FL model is by obtaining the training and testing loss. The training and testing loss is computed for the local model at the client's end in an FL environment as it tracks how the training and testing losses are decreasing over time.

5.6.3.3 Binary cross-entropy

This function computes the training and testing loss for the proposed architectures while performing binary classification. The formula for binary cross-entropy is as given by Equation 5.8:

$$L_B = -(ylog(p) + (1-y)log(1-p)) \tag{5.8}$$

Where,

- L_B is the binary loss generated for binary classification.

Table 5.6: List of evaluation measures for proposed Federated Learning based emotion recognition architectures.

Evaluation Measures	**Measures**
Performance Measures	• Confusion Matrix • Accuracy • Precision • Recall • F1-Score
Scalability Measures	• Training Time • Accuracy
Communication Computation	• Averaging Time • Convergence Speed (Loss)

- y represents a binary indicator (0 or 1) for accurately classifying the class label, where 0 corresponds to ”low”, and 1 corresponds to ”high”.
- p denotes the predicted probability associated with the observations of class labels.

5.6.3.4 Categorical Cross-Entropy

This function computes the training and testing loss for the proposed architectures for multi-class classification (with eight emotion classes). It computes the loss by comparing the predicted probability distribution with the true class labels. The formula for categorical cross-entropy is given in Equation 5.9:

$$L_C = -\sum_{c=1}^{C} y_c \log(p_c) \tag{5.9}$$

Where,

- L_C is the categorical loss generated for octal classification.
- C is the number of classes. y_c is the binary indicator for class c (1 if the correct class, 0 otherwise).
- p_c is the predicted probability for class c.

Table 5.6 lists all the evaluation measures for proposed Federate Learning emotion recognition architectures.

5.7 Summary

The literature for the FL paradigm is reviewed, including FL architectures, approaches, aggregation algorithms, and tools. It gives the related work focusing specifically on its application to ERS using physiological signals. It traces the evolution of FL as a decentralized approach that allows machine learning models to be trained across distributed data sources, maintaining privacy by ensuring sensitive data remains on client devices. The review critically assesses different FL implementations and techniques used in ERS, evaluating their effectiveness and model performance. It identifies existing gaps and limitations in the current research of FL for ERS based on physiological signals. Different evaluation measures are also presented to assess the proposed FL architectures.

This chapter is based on the following work:

- **J1: Neha Gahlan**, and Divyashikha Sethia. "Federated Learning in Emotion Recognition Systems based on Physiological Signals for Privacy Preservation: A Review" **Multimedia Tools and Applications**: 1-69. June 2024. (**SCIE, Impact factor-3.6, Publisher: Springer**). Doi: https://doi.org/10.1007/s11042-024-19467-3.

Chapter 6

F-MERS for Privacy to Emotion Recognition System

One of the significant issues in keeping physiological data while identifying emotions is a significant privacy concern. Physiological data like EEG, ECG, HR and others are sensitive as they reveal information about a person's health status and biometrics data and can potentially provide insights into their emotional state, behaviour, and habits [122, 135, 233, 287]. Conventional ERS use wearable sensors to record sensitive, physiological data fed into a machine learning model for training and classification purposes. This traditional ERS uses a centralized environment for training and classification without any privacy standards, allowing data attackers easy access to sensitive personal data, resulting in data leaks [82, 145, 146, 154]. The breach of such sensitive data will destroy authentication mechanisms [95, 164, 301]. Therefore, it is essential to use physiological signal data for emotion recognition systems to protect people's personal sensitive information.

This chapter introduces a novel FL-based Multi-modal Emotion Recognition System (***F-MERS***) to address the data privacy concerns in ERS. The proposed *F-MERS* architecture creates a decentralized environment by creating a client end (with subjects' multi-modal physiological signal data) and a server end for aggregating the updates. It runs the local model update at the client end and then sends the model updates (gradients) to the central server for averaging them. It does not allow the central server to access the raw physiological data at the client end, thereby preserving the privacy of physiological data.

Following are the main contributions of this chapter:

1. ***Proposal of F-MERS architecture combining multiple physiological signals:*** EEG, GSR, ECG, and RESP using Feature-level Fusion (*Multi-modality*). It uses an MLP classifier as a base model for classifying the three dimensions of the VAD model of emotions, i.e., Valence, Arousal, and Dominance (*Additional Emotion Dimension*).

2. ***Performance evaluation of proposed F-MERS architecture:*** It validates the three benchmark datasets: AMIGOS [125], DEAP [233] and DREAMER [135] using two testing scenarios: Subject-Dependent and Subject-Independent, making it more generalized and robust. It is compared with the non-FL architecture by achieving accuracy comparable to the centralized MLP model (non-federated). This chapter provides the scalability (in terms of rounds and iterations) and reports the communication computation (in terms of training and averaging times) for the proposed *F-MERS* architecture.

6.1 Experimental Methodology

The proposed F-MERS architecture utilized the three dimensions: Valence Arousal and Dominance of the Mehrabian and Russell's 3D-VAD model [33, 127] (as described earlier in Chapter 2).

6.1.1 Data Processing

1. **Datasets Description**
   The proposed F-MERS architecture validates the three emotion benchmark datasets among the others, as they consist of multi-modal physiological data: AMIGOS [125], DEAP [233], and DREAMER [135]. Detailed descriptions of all the datasets with their data processing, including data clipping and labelling, are given in Section 3.9.1. Table 6.1 briefly describes all three datasets.

Table 6.1: Brief description of all the datasets for F-MERS.

Description/Dataset	AMIGOS [125]	DEAP [233]	DREAMER [135]
No. of subjects	40	32 (16 male, 16 female)	23 (14 male, 9 female)
Physiological Signals	EEG, ECG, GSR	EEG, GSR, RESP	EEG, ECG
Video Content	16 short videos, 4 long videos	40 videos	18 videos
Video Duration	57-155 seconds, 14-15 minutes	63 seconds	65 - 393 seconds
Label Matrix	16 x 3	40 x 3	18 x 3
Emotion Dimensions	Arousal, Valence, Dominance		
Emotion Assessment	SAM (Self Assessment Manikins)		
EEG electrodes	14	32	14
ECG electrodes	2	-	2
GSR electrodes	1	1	-
RESP electrodes	-	1	-

Table 6.2: Extracted features from ECG, EEG, GSR, RESP signals for F-MERS.

Physiological Signal	Domain	Features
EEG (238/544 Features)	**Time (10 x 14/32)**	Hjorth Complexity, Activity, Mobility, Fractal Dimension, Mean, Median, Variance, Standard Deviation (STD), 1st and 2nd difference.
	Frequency (7 x 14/32)	Spectral Entropy, SVD Entropy, Sample Entropy, Bandpower (alpha, beta, theta, delta)
ECG (64 Features)	**Time (25 x 2)**	Mean NNI, Median NNI, NNI_50, HRV Mean, SDNN, RMSSD, Mean HR, Max HR, Min HR, Std HR, TINN, CSI, CVI, CVSD, CVNNI, SampEn, SD1, SD2, Triangular Index
	Frequency (7 x 2)	Total Power, VLF, HLF, LF, HF, LFnu, HFnu, LF/HF Ratio
GSR (20 Features)	**Time (19)**	Mean GSR, Var GSR, Skew GSR, Kurtosis GSR, Std GSR, SCL (tonic) slope, SCR (phasic) slope, Mean SCL, Var SCL, Std SCL, Mean SCR, Var SCR, Std SCR
	Frequency (1)	Power Spectral Density
RESP (6 Features)	**Time (6)**	Mean RESP, Median RESP, Var RESP, Skew RESP, Kurtosis RESP, Std RESP

2. **Feature Extraction & Feature Fusion**

   The proposed F-MERS architecture extracts the features given in Table 6.2) from the physiological signals: EEG, ECG (Right and Left channels), GSR and RESP, which are detailed in earlier section 2.3. It applies a sliding window of 4 seconds with 50% overlap for the feature extraction methods for all physiological signal data [135, 233].

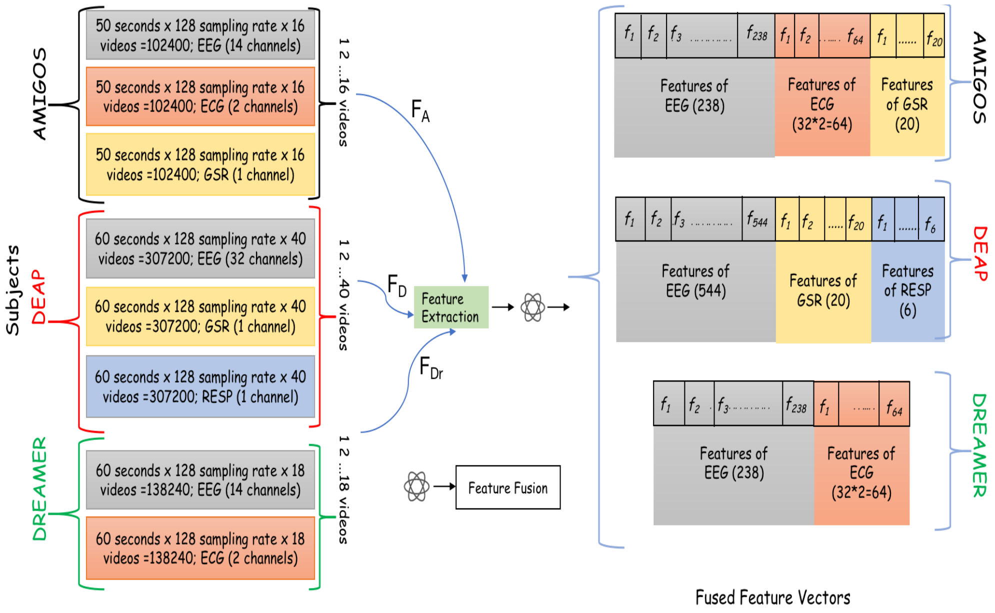


Figure 6.1: Feature Fusion for F-MERS.

(a) **Feature Fusion:** The proposed F-MERS architecture employs Feature-Level Fusion (FLF) to use rich data from distinct modalities. The FLF process involves performing feature extraction independently for each signal data. After feature extraction, the resulting feature vectors from each signal concatenate into a single fused feature vector. This fused feature vector represents the information extracted from all the signal modalities (EEG, ECG, GSR, and RESP) as shown in Figure 6.1 used in the training process. The final feature vectors obtained are:

For AMIGOS [125]:

$$F_A = f_{eeg}^{(14*17=238)} + f_{ecg}^{(32*2=64)} + f_{gsr}^{(20)} = f_{eeg+ecg+gsr}^{(322)} \tag{6.1}$$

For DEAP [233]:

$$F_D = f_{eeg}^{(32*17=544)} + f_{gsr}^{(20)} + f_{resp}^{(6)} = f_{eeg+gsr+resp}^{(570)} \tag{6.2}$$

For DREAMER [135]:

$$F_{Dr} = f_{eeg}^{(14*17=238)} + f_{ecg}^{(32*2=64)} = f_{eeg+ecg}^{(302)} \tag{6.3}$$

Where f_{eeg}, f_{ecg}, f_{gsr}, and f_{resp} present the number of features of EEG, ECG, GSR, and RESP signals, respectively. F_A gives the concatenated feature vector for the AMIGOS [125] dataset, and F_D gives for the DEAP [233] dataset, and F_{Dr} for the

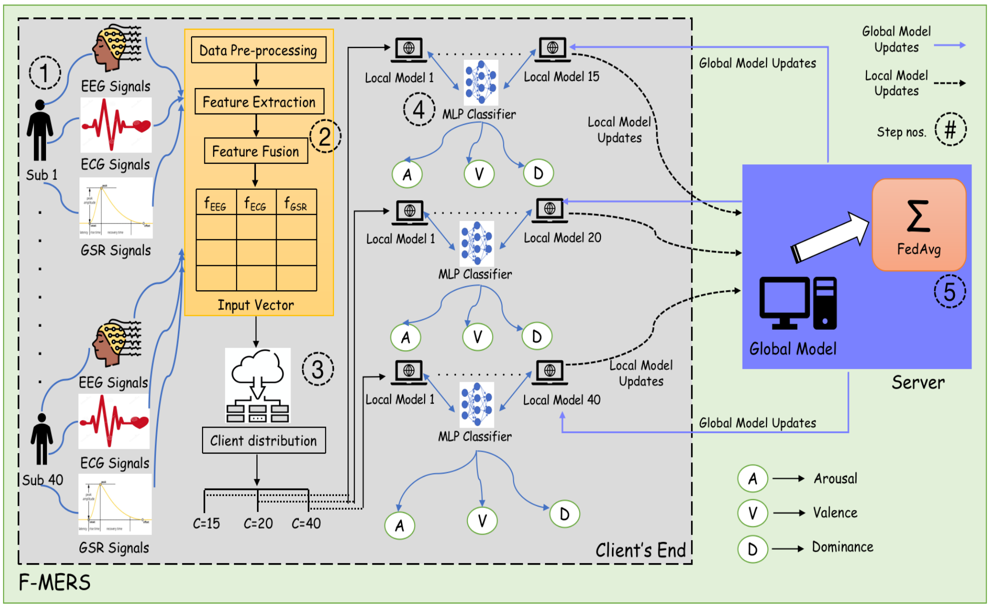


Figure 6.2: Architecture of proposed F-MERS with AMIGOS [125].

DREAMER [135] dataset. After concatenation, the fused feature vector consists of 322 features from AMIGOS [125] dataset, 570 from DEAP [233] dataset, and 302 for DREAMER [135] dataset presented by equations 6.1, 6.2, and 6.3 respectively.

6.1.2 Architecture of F-MERS

The architecture for the proposed ***F-MERS*** is discussed below stepwise in detail, and its illustration with marked respective steps is given by Figure 6.2.

- **Step 1: Data Collection and Preprocessing**

  Firstly, the proposed F-MERS architecture collects multi-modal physiological data from all the subjects. It includes physiological signals such as EEG, ECG, GSR, and RESP data. This data then goes for preprocessing, serving as input to the next step.

- **Step 2: Feature Extraction and Fusion**

  The architecture performs feature extraction on the collected physiological data for the Statistical, Time-domain, and Frequency-domain features. It fuses the extracted features into a single concatenated feature vector.

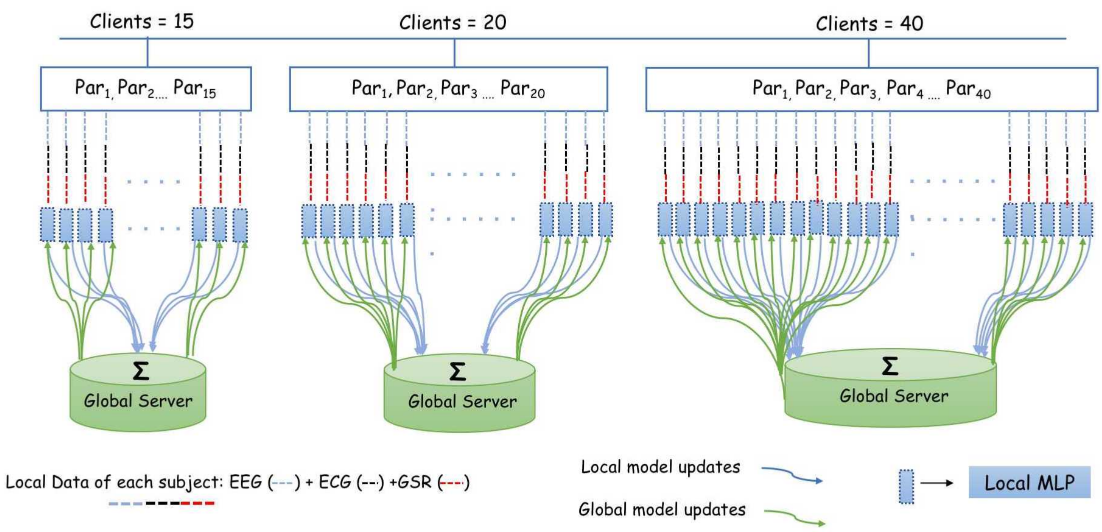


Figure 6.3: Data divisions of clients for F-MERS (with AMIGOS [125]).

Table 6.3: Partitioning subjects into clients for each dataset (1 Client = 1 Subject) for F-MERS.

AMIGOS [125] (40 Subjects)		
Client = 15	Client = 20	Client = 40
DEAP [233] (32 Subjects)		
Client=10	Client=16	Client = 32
DREAMER [135] (23 Subjects)		
Client=7	Client=11	Client = 23

- **Step 3: Data Division for FL Environment**

  This step defines the participating clients in the FL environment. For creating a federated environment, the architecture divides the datasets into three different client divisions, with each client consisting of one subject's multi-stream physiological data, followed similarly by two other datasets. An illustration is shown in Figure 6.3 for the AMIGOS [125] dataset, showing that the first experiment is with clients=15, the second with clients=20, and the third with clients=40. It works similarly for the other two datasets with different no. of client as given in Table 6.3.

- **Step 4: Model Selection and Training**

  As a baseline classifier, the proposed F-MERS architecture uses a Multi-Layer Perceptron (MLP) neural network in the federated learning environment. The MLP model takes the concatenated feature vector of the physiological signal data in the input layer. Table 6.4 gives the parameter for the proposed F-MERS

Table 6.4: Model parameters of the proposed F-MERS.

Layers	Output Shape	Activation Function	Hyperparameters
Input	None, 322/570/302	None	None
Dense	None, 512/1024	ReLU	Dropout=0.7
Dense	None, 256/512	ReLU	Dropout=0.7
Dense	None, 128/256	ReLU	Dropout=0.7
Dense	None, 256/128	ReLU	Dropout=0.7
Dense	None, 64	ReLU	Dropout=0.7
Output	None, 1	Sigmoid	None
Optimizer = SGD, Learning Rate =0.005			

Algorithm 1 **Federated Averaging (FedAvg [136])**

Server Execution:

initialize w_0
for each round r $=1,2,\ldots$ do
    $m \leftarrow max(int(C*m),1)$;
    $S_t \leftarrow$ (random set of m clients);
    for each client $k \in S_t$ **in parallel do**
        $w^k_{t+1} \leftarrow$ ClientUpdate(k,w_t) end
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w^k_{t+1}$

ClientUpdate(k,w)://Run on client k
$A \leftarrow$ (split P_k into batches of size B)
for each local epoch i from 1 to E do
    for batch $a \in A$ do
        $w \leftarrow w - \eta \Delta l(w;\, a)$

return w to server

architecture using multiple dense layers with an activation function.

- **Step 5: Creating FL Environment**

  TFF [208] is used as an FL tool in the proposed F-MERS architecture. After splitting the dataset, the architecture distributes it to multiple virtual clients. It creates a federated learning environment using TFF to produce FedAvg, a federated averaging algorithm [136]. TFF employs a distributed aggregation protocol [214] to collect and aggregate client model updates. Equation 6.4 gives the computation for FedAvg.

$$w_t^g = \frac{1}{N_{total}} \sum_{i=1}^{N_{total}} w_{t,i}^l \tag{6.4}$$

Table 6.5: Partitioning subjects into clients: Subject-Independent scenario for F-MERS.

Dataset = AMIGOS [125] (40 Subjects)			
1 Client= 1 subject	Client = 15	Client = 20	Client = 40
Training/Testing clients	12/3	16/4	32/8
Dataset = DEAP [233] (32 Subjects)			
1 Client= 1 subject	Client=10	Client=16	Client = 32
Training/Testing clients	8/2	13/3	26/8
Dataset = DREAMER [135] (23 Subjects)			
1 Client= 1 subject	Client=7	Client=11	Client = 23
Training/Testing clients	5/2	9/2	18/5

Where w_t^g is the aggregated weight at the global server in time t, $w_{t,i}^l$ are the weights received from all local models in time t, and N_{total} is the total number of the local model participating for aggregation. It employs the horizontal federated learning approach [27]. The Algorithm 1 describes the detailed steps for the algorithm for FedAvg as:

1. **Creation of Local Model (client end):** The local model is created using the multi-layer perceptron, initially taking the feature vector as input from each client's multi-stream physiological data. The models from each client are then sent to the server to generate a global aggregator.

2. **Creation of Global Aggregator (server end):** After receiving the local model from the clients, the server performs the federated averaging (FedAvg). It then sends the updates from the global server back to the clients.

3. **Local Model Updates (client end):** Following aggregation, the participating client receives updates from the global server. The process repeats until the model converges or completes the number of iterations. The experiments are performed with three rounds of iterations (Rounds = 100, 200, 500).

6.1.3 Experimental Setup

This chapter experiments with two perspectives: The non-federated learning environment (Non-FL) and the Federated learning environment (FL). It validates both of these approaches in Subject-dependent and Subject-independent scenarios. These approaches are employed to assess the effectiveness of the proposed F-MERS architecture. For both perspectives, this work utilizes Google Colab's Pro plus NVIDIA V100 GPU, CUDA version 11.2, Python 3.8, TensorFlow (TF version-2.6.0), TensorFlow Federated (TFF version-0.19.0) and Keras frameworks to run the models and trials on a MacBook Air with a 1.6 GHz dual-core Intel core i5. Each round of aggregation in the federated environment requires clients to train one epoch locally, with batch size 256 [173, 222].

6.1.4 Validation Scenarios of the architecture

This chapter validates the proposed FL approach in two scenarios: One is Subject-dependent, and the other is Subject-independent.

1. **The Subject-dependent scenario** uses data from each client for training and testing, i.e., 80% of each of the client's data for training and 20% of each of the client's data for testing.

2. **The Subject-independent scenario** uses different sets of training and testing clients, which are in the ratio of 80% (training) and 20% (testing). For example, the model uses 12 clients for training and the rest 3 clients for testing.

Table 6.5 presents the no. of subjects utilized for training and testing the experiment in the subject-independent scenario separately for each dataset, along with the details of the three experiments performed.

6.1.5 Evaluation Measures for F-MERS

The following measures evaluate the proposed F-MERS architecture (as detailed in Section 5.6). **Performance Measures:** Confusion Matrix, Binary Accuracy (Training Accuracy and Testing Accuracy) and F1-Score. **Scalability Measures:** Training Time and Model Accuracy with different client distributions and iteration rounds. **Communication Computation Measure:** Averaging time for different client distribution and iteration rounds.

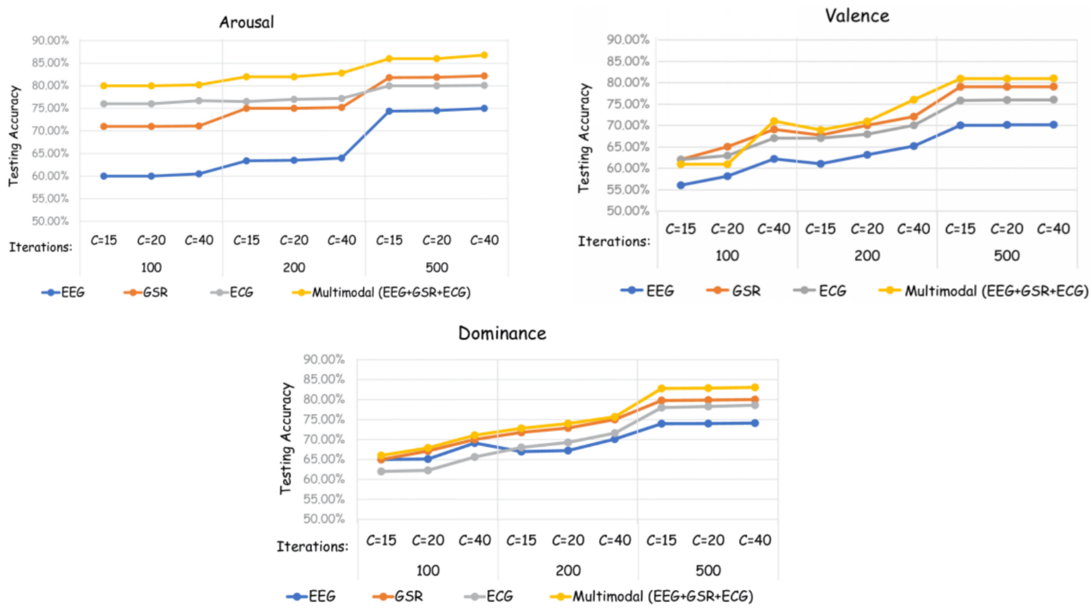


Figure 6.4: Testing accuracies for proposed F-MERS for AMIGOS [125] in **Subject-dependent** scenario.

6.2 Results and Discussion

This section presents the experimental results of the proposed F-MERS architecture for binary classification of emotion dimensions. The local models at the client end use the updates from the global server and perform the classification at their end, protecting user privacy. The global server does not have any access to the raw data. The following subsections present the experimental results for both Subject-dependent and Subject-independent scenarios.

6.2.1 Subject-Dependent Results

This section presents the experimental results of the proposed F-MERS architecture for the subject-dependent scenario with all three datasets: AMIGOS [125], DEAP [233] and DREAMER [135].

1. ***Testing Accuracies of F-MERS:*** Figures 6.4, 6.5, 6.6 shows the graphical representation of testing accuracy (average of all clients) scores of the proposed F-MERS architecture with all rounds of aggregation for all the modalities with AMIGOS [125], DEAP [233] and DREAMER [135] datasets, respectively in subject-dependent scenario. It clearly illustrates that the proposed F-MERS

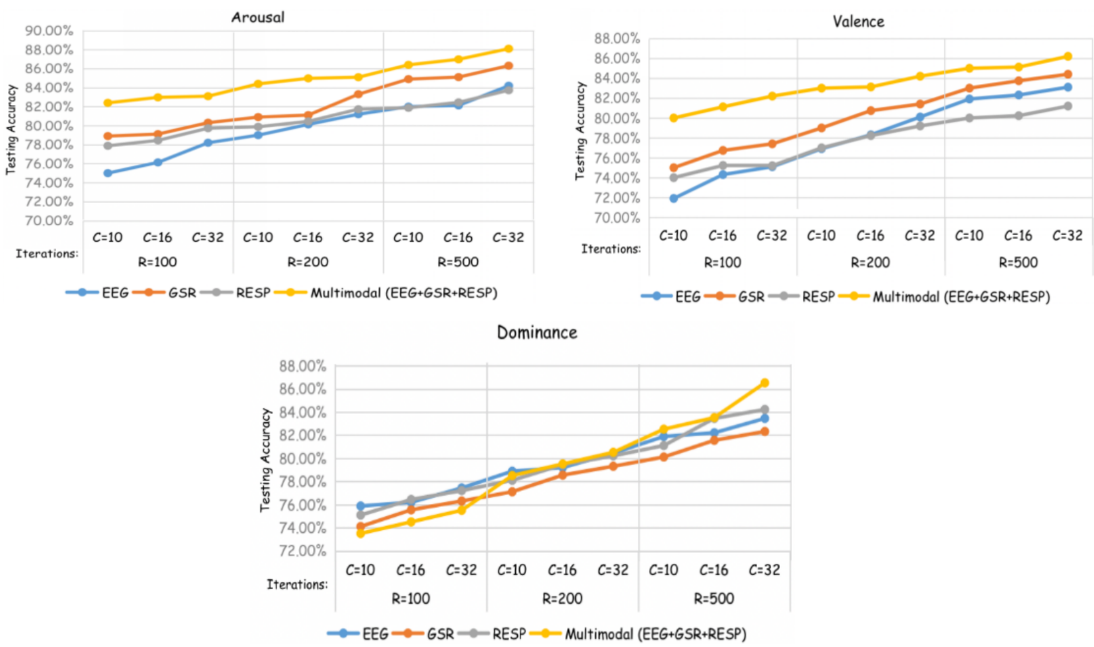


Figure 6.5: Testing accuracies for proposed F-MERS for DEAP [233] in **Subject-dependent** scenario.

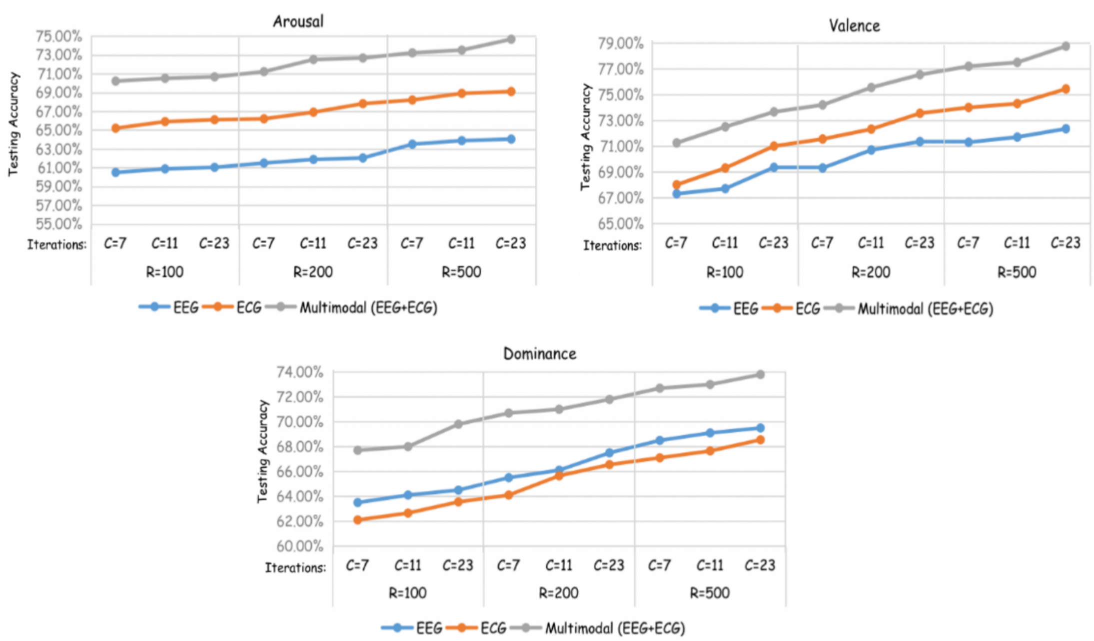


Figure 6.6: Testing accuracies for proposed F-MERS for DREAMER [135] in **Subject-dependent** scenario.

architecture performs better with multi-modal physiological signal data than single modalities of physiological signal data, validated by different client distributions. Hence, we present further results with 500 rounds of iterations.

Table 6.6: Training and testing accuracies for proposed F-MERS with 500 rounds in **subject-dependent** scenario.

Metrics	Training Accuracy			Testing Accuracy		
(AMIGOS [125])						
Physiological Signal/ Clients(C)	**C=15**	**C=20**	**C=40**	**C=15**	**C=20**	**C=40**
Arousal						
EEG	74.20%	75.13%	77.40%	74.40%	74.50%	75.00%
GSR	81.11%	82.23%	85.50%	81.80%	81.89%	82.20%
ECG	80.70%	81.40%	81.12%	80.00%	80.00%	80.10%
Multimodal EEG+GSR+ECG	**86.14%**	**87.11%**	**88.88%**	**86.00%**	**86.00%**	**86.80%**
Valence						
EEG	70.11%	71.11%	72.11%	70.00%	70.11%	70.15%
GSR	79.05%	80.55%	81.05%	79.00%	79.00%	79.03%
ECG	76.00%	76.11%	77.00%	75.80%	75.92%	76.00%
Multimodal EEG+GSR+ECG	**80.89%**	**81.40%**	**81.80%**	**80.90%**	**80.90%**	**80.98%**
Dominance						
EEG	73.22%	74.55%	75.44%	73.97%	74.00%	74.10%
GSR	79.90%	80.00%	81.55%	79.78%	79.89%	80.00%
ECG	79.00%	80.45%	80.22%	78.00%	78.26%	78.60%
Multimodal EEG+GSR+ECG	**82.10%**	**82.80%**	**84.11%**	**82.80%**	**82.88%**	**83.06%**
(DEAP [233])						
Physiological Signal/ Clients(C)	**C=10**	**C=16**	C=32	**C=10**	**C=16**	**C=32**
Arousal						
EEG	82.20%	83.78%	86.10%	82.00%	82.13%	84.20%
GSR	85.10%	85.78%	87.11%	84.90%	85.11%	86.31%
RESP	82.11%	83.11%	84.55%	81.88%	82.45%	83.73%
Multimodal EEG+GSR+RESP	**86.60%**	**87.10%**	**89.10%**	**86.40%**	**86.98%**	**88.10%**
Valence						
EEG	81.10%	83.11%	85.05%	81.91%	82.31%	83.10%
GSR	82.30%	84.44%	85.15%	83.00%	83.74%	84.40%
RESP	80.00%	80.55%	83.45%	80.00%	80.23%	81.20%
Multimodal EEG+GSR+RESP	**85.77%**	**86.45%**	**88.10%**	**85.00%**	**85.12%**	**86.20%**
Dominance						
EEG	81.89%	82.21%	83.45%	81.89%	82.21%	83.45%
GSR	80.12%	81.56%	82.32%	80.12%	81.56%	82.32%
RESP	81.11%	83.45%	84.22%	82.21%	83.45%	84.22%
Multimodal EEG+GSR+RESP	**82.52%**	**83.52%**	**86.52%**	**81.56%**	**82.32%**	**86.52%**
(DREAMER [135])						
Physiological Signal/ Clients(C)	**C=7**	**C=11**	**C=23**	**C=7**	**C=11**	**C=23**
Arousal						
EEG	61.50%	63.11%	65.10%	63.50%	63.88%	64.04%
ECG	69.10%	69.50%	71.25%	68.20%	68.90%	69.10%
Multimodal EEG+ECG	**73.10%**	**73.40%**	**75.12%**	**73.21%**	**73.50%**	**74.66%**
Valence						
EEG	70.10%	72.10%	73.45%	71.30%	71.70%	72.35%
ECG	73.80%	74.11%	76.21%	74.07%	74.30%	75.43%
Multimodal EEG+ECG	**76.16%**	**78.20%**	**79.05%**	**77.20%**	**77.50%**	**78.12%**
Dominance						
EEG	67.70%	69.60%	70.50%	68.50%	69.10%	69.50%
ECG	66.50%	68.10%	69.150%	67.10%	67.65%	68.55%
Multimodal EEG+ECG	**71.05%**	**71.30%**	**74.10%**	**72.70%**	**73.00%**	**73.80%**

2. ***Training and Testing Accuracy Comparison for F-MERS:*** Table 6.6 gives the average training and testing accuracies of the proposed F-MERS architecture (with 500 rounds of iterations) from all the client distributions for single and multi-modal physiological data. For example, when Clients =15, the table gives

Table 6.7: F1-Score results (during testing) for the proposed F-MERS with all datasets in a **subject-dependent** scenario

Datasets→	AMIGOS (EEG+ECG+GSR)			DEAP (EEG+GSR+RESP)			DREAMER (EEG+ECG)		
Emotions Dimensions↓	**C = 15**	**C = 20**	**C = 40**	**C = 10**	**C = 16**	**C = 32**	**C = 7**	**C = 11**	**C = 23**
Valence	0.801	0.810	0.811	0.854	0.865	0.864	0.736	0.748	0.761
Arousal	0.851	0.851	0.854	0.853	0.868	0.876	0.763	0.777	0.782
Dominance	0.825	0.829	0.833	0.819	0.834	0.861	0.711	0.732	0.735

an average of training accuracies achieved by all 15 models trained at each client end. The difference between training and testing results in Table 6.6 is hardly 1-2%, showing no overfitting in the proposed F-MERS architecture. It clearly illustrates that the multi-modal architecture performs better than single modalities, validated by different client distributions. The different client distributions represent the ability of the proposed F-MERS architecture to handle large amounts of data without compromising its performance, making it scalable. The proposed F-MERS architecture achieves an average testing accuracy of 88.10% (arousal), 86.20% (valence), and 86.52% (dominance) with the DEAP [233] dataset for clients = 32, 86.80% (arousal), 80.98% (valence), and 83.06% (dominance) with the AMIGOS [125] dataset for clients = 40, 74.66% (arousal), 78.12% (valence), and 73.80% (dominance) with the DREAMER [135] dataset for clients = 23, for the multi-modal physiological signal data.

3. ***F1-Score for F-MERS:*** Table 6.7 gives the F1-score values for the proposed multi-modal F-MERS architecture derived while testing (averaged from all the clients). The F1-score values being close to 1 demonstrate that the proposed F-MERS architecture ensures that the predicted labels (low/high) match closely to the actual class labels (low/high). It indicates a strong performance of the proposed F-MERS architecture in identifying correct positive instances with all three datasets.

4. ***Confusion Matrix for Binary Classification by F-MERS:*** Table 6.8 provides the confusion matrix for a better presentation of the classification task performed. It gives the values for each low/high-class label for binary classification performed for emotion dimensions arousal, valence, dominance. It shows the ratio of actual and predicted class is higher for true positive and true negative data samples.

Table 6.8: Confusion matrix for the proposed F-MERS in **subject-dependent** scenario.

AMIGOS [125]								
	Valence			**Arousal**			**Dominance**	
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.854	0.015	**Low**	0.895	0.018	**Low**	0.854	0.018
High	0.250	0.908	**High**	0.175	0.876	**High**	0.174	0.819
DEAP [233]								
	Valence			**Arousal**			**Dominance**	
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.942	0.059	**Low**	0.855	0.049	**Low**	0.842	0.059
High	0.121	0.878	**High**	0.128	0.869	**High**	0.122	0.878
DREAMER [135]								
	Valence			**Arousal**			**Dominance**	
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.894	0.145	**Low**	0.886	0.123	**Low**	0.775	0.121
High	0.109	0.889	**High**	0.129	0.862	**High**	0.142	0.862

Table 6.9: Testing accuracy for the proposed F-MERS and the Non-FL across all datasets in a **subject-dependent** scenario with maximum number of clients.

Environment	**Non FL**	**With FL**	**Non FL**	**With FL**	**Non FL**	**With FL**
Epochs(E)/Rounds(R)	**E = 500**	**R = 500**	**E = 500**	**R = 500**	**E = 500**	**R = 500**
AMIGOS [125] (with clients = 40)						
Physiological Signal	**(Arousal)**		**(Valence)**		**(Dominance)**	
EEG	75.10%	75.00%	70.35%	70.15%	74.15%	74.10%
GSR	82.30%	82.20%	79.11%	79.03%	80.11%	80.00%
ECG	80.11%	80.0%	76.0%	76.0%	78.55%	78.60%
Multimodal (EEG+GSR+ECG)	**86.81%**	**86.80%**	**81%**	**80.98%**	**83.11%**	**83.06%**
DEAP [233] (with clients = 32)						
Physiological Signal	**(Arousal)**		**(Valence)**		**(Dominance)**	
EEG	84.51%	84.20%	83.18%	83.10%	83.80%	83.45%
GSR	86.52%	86.31%	84.53%	84.40%	82.45%	82.32%
RESP	83.92%	83.73%	81.28%	81.20%	84.30%	84.22%
Multimodal (EEG+GSR+RESP)	**88.20%**	**88.10%**	**86.34%**	**86.20%**	**86.78%**	**86.52%**
DREAMER [135] (with clients = 23)						
Physiological Signal	**(Arousal)**		**(Valence)**		**(Dominance)**	
EEG	65.00%	64.04%	72.35%	72.15%	69.50%	69.12%
ECG	70.11%	69.10%	75.10%	75.20%	68.55%	68.23%
Multimodal (EEG+ECG)	**74.81%**	**74.66%**	**78.23%**	**78.10%**	**73.80%**	**73.43%**

5. ***Comparison of F-MERS with Non-federated MLP:*** This chapter also compares the efficacy of the proposed federated F-MERS architecture to the non-federated learning centralized MLP for emotion recognition based on physiological signal data via results from Table 6.9. The objective is to achieve comparability between the proposed F-MERS architecture and the non-federated centralized MLP model, along with the addition of data privacy considerations in the proposed F-MERS. The proposed multi-modal F-MERS architecture achieves an average accuracy of 86.94% (from valence, arousal and dominance), which is comparable with multi-

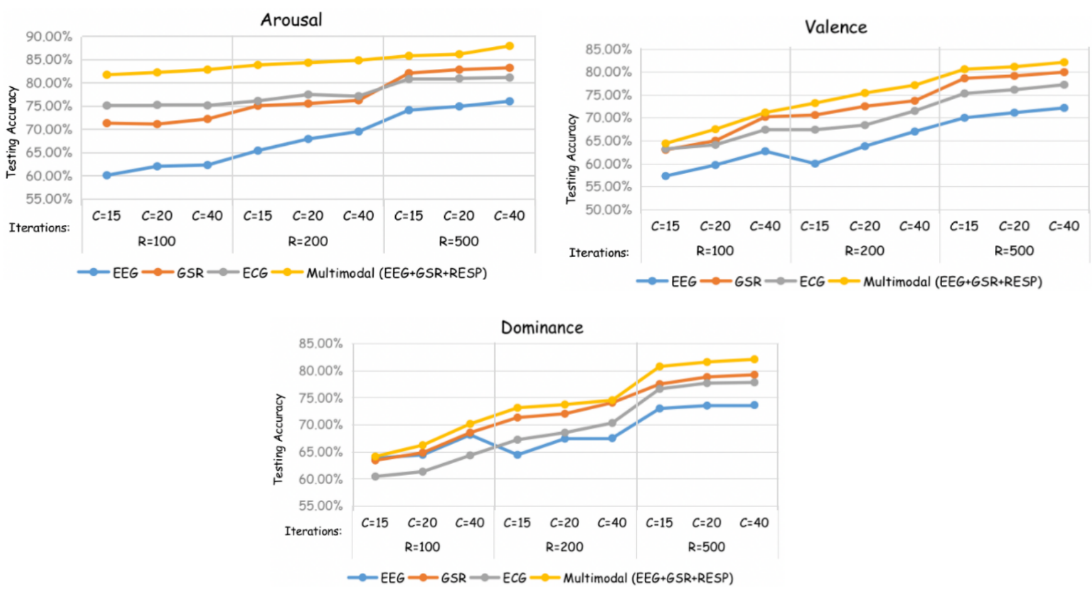


Figure 6.7: Testing accuracies for proposed F-MERS for AMIGOS [125] in **Subject-independent** scenario.

modal Non-FL achieving 87.10% (from valence, arousal and dominance) with the DEAP [233] dataset for client = 32. For the AMIGOS [125] dataset, the proposed multi-modal F-MERS architecture achieves an average accuracy of 83.61% (from valence, arousal and dominance), which is comparable with multi-modal Non-FL achieving 83.64% (from valence, arousal and dominance) with clients = 40. For the DREAMER [135] dataset, the proposed multi-modal F-MERS architecture achieves an average accuracy of 75.39% (from valence, arousal and dominance), which is comparable with multi-modal Non-FL achieving 75.61% (from valence, arousal and dominance) with clients = 23.

6.2.2 Subject-Independent Results

This section presents the experimental results of the proposed F-MERS architecture for the subject-independent scenario with the datasets: AMIGOS [125], DEAP [233] and DREAMER [135].

1. ***Testing Accuracies of F-MERS:*** Figures 6.7, 6.8, 6.9 shows the graphical representation of accuracy scores of the proposed F-MERS architecture with all rounds of aggregation for all the modalities with AMIGOS [125], DEAP [233] and DREAMER [135] datasets, respectively in subject-independent scenario. It shows that the proposed multi-modal F-MERS architecture performs best for all

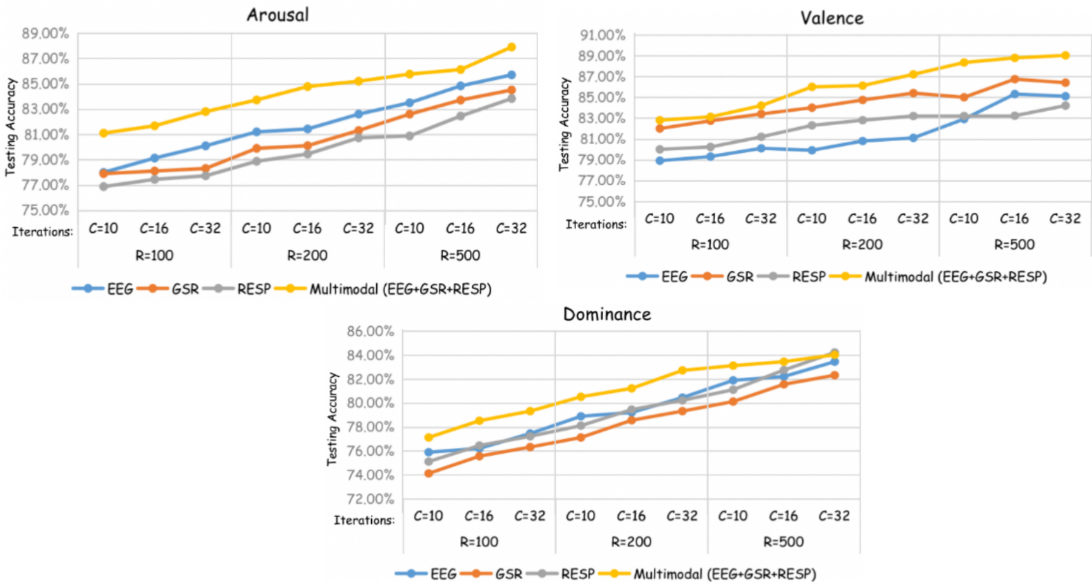


Figure 6.8: Testing accuracies for proposed F-MERS for DEAP [233] in **Subject-independent** scenario.

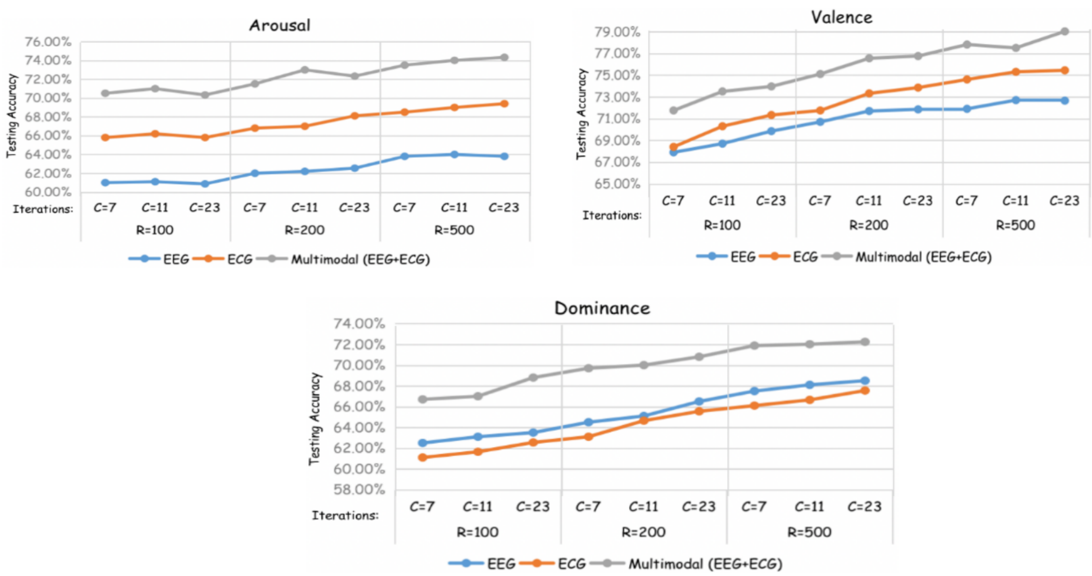


Figure 6.9: Testing accuracies for proposed F-MERS for DREAMER [125] in **Subject-independent** scenario.

three emotion dimensions when rounds = 500 among the rest of the rounds. It clearly illustrates that the proposed F-MERS architecture performs better with multi-modal physiological signal data than single modalities of physiological signal data, validated by different client distributions. Hence, we present further results with 500 rounds of iterations.

2. ***Training and Testing Accuracy Comparison for F-MERS:*** Table 6.10 gives the

Table 6.10: Training and testing accuracies for the proposed F-MERS with 500 rounds in **subject-independent** scenario.

Metrics	Training Accuracy			Testing Accuracy		
(AMIGOS [125])						
Physiological Signal/Clients (C)	**C=15**	**C=20**	**C=40**	**C=15**	**C=20**	**C=40**
Arousal						
Multimodal EEG+GSR+ECG	85.56%	87.10%	88.50%	85.76%	86.22%	87.90%
Valence						
Multimodal EEG+GSR+ECG	80.60%	81.52%	82.80%	80.58%	81.12%	82.10%
Dominance						
Multimodal EEG+GSR+ECG	81.10%	82.50%	83.10%	80.73%	81.56%	82.06%
(DEAP [233])						
Physiological Signal/Clients	**C=10**	**C=16**	**C=32**	**C=10**	**C=16**	**C=32**
Arousal						
Multimodal EEG+GSR+RESP	84.80%	86.10%	87.80%	85.22%	85.72%	86.50%
Valence						
Multimodal EEG+GSR+RESP	87.12%	88.18%	89.88%	88.34%	88.78%	89.02%
Dominance						
Multimodal EEG+GSR+RESP	81.10%	82.10%	84.88%	83.12%	83.45%	84.02%
(DREAMER [135])						
Physiological Signal/Clients	**C=7**	**C=11**	**C=23**	**C=7**	**C=11**	**C=23**
Arousal						
Multimodal EEG+ECG	72.80%	74.55%	75.10%	73.50%	74.01%	74.33%
Valence						
Multimodal EEG+ECG	76.60%	78.15%	79.88%	77.81%	78.52%	79.02%
Dominance						
Multimodal EEG+ECG	70.88%	72.30%	72.40%	71.88%	72.01%	72.24%

average training and testing accuracies of the proposed F-MERS architecture (with 500 rounds of iterations) from all the client distributions for single and multi-modal physiological data in the subject-independent scenario. The difference highlighted between training and testing results given in Table 6.10 is hardly 1-2%, showing that the proposed F-MERS architecture is not overfitted. Here, the multi-modal architecture performs better than single modalities, as validated by different client distributions. The different client distributions represent the ability of the proposed F-MERS architecture to handle large amounts of data without compromising its performance, making it scalable. In the subject-independent scenario, it achieves an average testing accuracy of 86.50% (arousal), 89.02% (valence), and 84.02% (dominance) with the DEAP [233] dataset for client = 32. 87.90% (arousal), 82.10% (valence), and 82.06% (dominance) with

Table 6.11: F1-Score results (during testing) for the proposed F-MERS with all datasets in **subject-independent** scenario

Datasets→	AMIGOS (EEG+ECG+GSR)			DEAP (EEG+GSR+RESP)			DREAMER (EEG+ECG)		
Emotions Dimensions ↓	C = 15	C = 20	C = 40	C=10	C = 16	C = 32	C = 7	C=11	C = 23
Valence	0.795	0.791	0.813	0.891	0.883	0.893	0.789	0.801	0.804
Arousal	0.805	0.882	0.891	0.864	0.878	0.888	0.702	0.749	0.768
Dominance	0.822	0.812	0.824	0.841	0.833	0.854	0.722	0.744	0.761

Table 6.12: Confusion matrix for the proposed F-MERS in **subject-independent** scenario.

AMIGOS [125]								
	Valence			Arousal			Dominance	
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.854	0.085	**Low**	0.891	0.145	**Low**	0.754	0.168
High	0.094	0.967	**High**	0.076	0.888	**High**	0.112	0.966
DEAP [233]								
	Valence			Arousal			Dominance	
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.842	0.151	**Low**	0.854	0.149	**Low**	0.812	0.169
High	0.159	0.848	**High**	0.089	0.908	**High**	0.191	0.828
DREAMER [135]								
	Valence			Arousal			Dominance	
Class	**Low**	**High**	**Class**	**Low**	**High**	**Class**	**Low**	**High**
Low	0.894	0.129	**Low**	0.806	0.101	**Low**	0.917	0.15
High	0.128	0.849	**High**	0.109	0.984	**High**	0.13	0.850

the AMIGOS [125] dataset for client = 40. 74.33% (arousal), 79.02%(valence), and 72.24% (dominance) with the DREAMER [135] dataset for clients = 23, for the multi-modal physiological signals.

3. ***F1-Score for F-MERS:*** Table 6.11 gives the F1-score values for the proposed multi-modal F-MERS architecture when tested in the subject-independent scenario averaged from all the clients. The F1-score values are near to 1, showing that the proposed F-MERS architecture ensures that the predicted labels (low/high) match closely to the actual class labels (low/high), indicating correct positive instances classification.

4. ***Confusion Matrix for Binary Classification by F-MERS:*** Table 6.12 presents the confusion matrix for a better presentation of the classification task performed. Each matrix representing the classification ratios breaks down the recognition performance for the emotion dimensions Valence, Arousal, and Dominance across Low and High classes. It shows that the ratio of actual and predicted class is higher for true positive and true negative data samples.

Table 6.13: Testing accuracy comparison for the proposed F-MERS and the Non-FL across all datasets in a **subject-independent** scenario with maximum number of clients.

Environment	**Non FL**	**With FL**	**Non FL**	**With FL**	**Non FL**	**With FL**
Epochs(E)/Rounds(R)	**E=500**	**R=500**	**E=500**	**R=500**	**E=500**	**R=500**
AMIGOS [125] (with clients=40)						
Physiological Signal	**(Arousal)**		**(Valence)**		**(Dominance)**	
EEG	76.21%	75.09%	71.15%	70.95%	73.76%	73.15%
GSR	82.45%	82.20%	79.91%	79.46%	80.11%	80.00%
ECG	81.56%	81.14%	76.35%	76.20%	77.86%	77.50%
Multimodal (EEG+GSR+ECG)	**88.12%**	**87.90%**	**81.81%**	**82.10%**	**82.23%**	**82.06%**
DEAP [233] (with clients=32)						
Physiological Signal	**(Arousal)**		**(Valence)**		**(Dominance)**	
EEG	84.89%	84.45%	85.67%	85.53%	84.67%	84.55%
GSR	85.25%	85.16%	85.53%	85.89%	81.45%	81.29%
RESP	82.12%	82.24%	82.12%	82.06%	80.56%	80.34%
Multimodal (EEG+GSR+RESP)	**86.88%**	**86.50%**	**89.12%**	**89.02%**	**84.16%**	**84.02%**
DREAMER [135] (with clients=23)						
Physiological Signal	**(Arousal)**		**(Valence)**		**(Dominance)**	
EEG	66.10%	64.78%	74.53%	74.21%	68.87%	68.43%
ECG	71.05%	70.10%	77.86%	77.68%	67.88%	67.10%
Multimodal (EEG+ECG)	**75.02%**	**74.33%**	**79.23%**	**79.02%**	**72.68%**	**72.24%**

5. ***Comparison of F-MERS with Non-federated MLP:*** This chapter compares the efficacy of proposed F-MERS architecture to non-federated learning (centralized MLP) for emotion recognition based on physiological signal data in subject-independent scenarios. Table 6.13 gives an average accuracy of 86.51% (from valence, arousal and dominance), which is comparable with Non-FL achieving 86.72% (from valence, arousal and dominance) for multi-modal data with the DEAP [233] dataset for clients = 32. For the AMIGOS [125] dataset, the proposed multi-modal FL architecture achieves an average accuracy of 84.02% (from valence, arousal and dominance), which is comparable with multi-modal Non-FL achieving 83.45% (from valence, arousal and dominance) with clients = 40. And, for the DREAMER [135] dataset, the proposed multi-modal FL architecture achieves an average accuracy of 75.19% (from valence, arousal and dominance), which is comparable with multi-modal Non-FL achieving 75.64% (from valence, arousal and dominance) with clients = 23.

Table 6.14: Training and averaging time (in seconds) for proposed F-MERS in subject-independent scenario. Clients (C), Rounds (R).

(AMIGOS [125])									
Measures	C=15			C=20			C=40		
(in seconds)	R=100	R=200	R=500	R=100	R=200	R=500	R=100	R=200	R=500
Training Time	99.157	194.701	508.756	99.427	198.958	510.943	103.376	215.104	520.434
Averaging Time	76.458	152.774	398.618	84.924	155.722	410.698	90.293	162.842	419.394
(DEAP [233])									
Measures	C=10			C=16			C=32		
(in seconds)	R=100	R=200	R=500	R=100	R=200	R=500	R=100	R=200	R=500
Training Time	130.412	220.011	550.123	145.411	240.812	568.132	158.631	260.421	596.287
Averaging Time	90.812	182.434	410.509	100.555	198.231	450.918	115.342	190.412	470.314
(DREAMER [135])									
Measures	C=7			C=11			C=23		
(in seconds)	R=100	R=200	R=500	R=100	R=200	R=500	R=100	R=200	R=500
Training Time	110.609	200.354	518.145	120.512	220.989	540.456	125.39	235.799	550.422
Averaging Time	77.512	155.445	400.867	86.432	160.254	415.821	92.367	170.256	421.443

6.2.3 Communication and Scalability Measures

This thesis is the first to report the training and averaging times for a federated learning architecture for emotion recognition using physiological data. Table 6.14 assesses the scalability and communication computation of the proposed F-MERS architecture. For measuring scalability, training time (in seconds) and classification accuracies with different data distributions is computed, which implies that the proposed approach can handle different data distributions with acceptable training times, making the proposed F-MERS architecture scalable. Furthermore, averaging time (in seconds) is computed for aggregation in the FL environment to measure the communication computation of the global server reflected by different rounds of aggregations, which converge at round = 500, giving optimal performance. Table 6.14 presents the mean of training and averaging times for subject-dependent and subject-independent scenarios. Federated learning requires more communication overhead due to aggregation [158, 265]. Table 6.14 provides the maximum training time of 596.287 seconds (taken by the local model) and maximum averaging time of 470.314 seconds (taken by the global server), which are manageable. The training and averaging times are low enough to conclude that the proposed F-MERS architecture can be feasibly deployed in real-time emotion recognition scenarios, where frequent model updates are needed without sacrificing privacy.

6.2.4 Discussion

1. **Proposed FL architecture in both subject-dependent and subject-independent scenario:** This chapter aims to validate the proposed FL architecture F-MERS through experiments conducted in two different scenarios: Subject-Dependent (SD) and Subject-Independent (SID). The results obtained in these scenarios show that the performance of the F-MERS architecture is slightly better (1-2%) in the Subject-independent scenario compared to the Subject-dependent scenario. However, it is noteworthy that the performance difference between the two scenarios is very minute, validated by all three datasets. Results in Tables 6.6, 6.9, 6.10, 6.13 indicates that the proposed F-MERS architecture performs well in both scenarios and proves to be robust, scalable and generalized.

2. **Proposed FL architecture comparison with existing FL work for emotion:** To compare the proposed F-MERS architecture with the existing works in FL for emotion recognition, the same evaluation grounds are included, i.e., the same datasets and validation approach. Table 6.15 compares the proposed F-MERS architecture with previous work based on FL for emotion recognition using physiological signals. The results in the table conclude that the proposed F-MERS architecture with multi-modal physiological signal data outperforms the previous works in both the validation scenarios of subject-dependent (SD) and subject-independent (SID) in terms of accuracy, robustness, and scalability.

6.2.4.1 Limitations

The proposed FL architecture, F-MERS, utilizes multi-modal physiological data, including signals from EEG, ECG, GSR, and RESP, as a combined vector. However, integrating these different signals together may result in variation data heterogeneity, which the proposed F-MERS architecture does not address. Additionally, F-MERS architecture uses uniformly distributed data across all clients, a scenario that does not align with the uneven data distribution often seen in real-world applications.

Table 6.15: Comparison of F-MERS with existing FL-based ERS.

Reference	Dataset	Physiological Signal	CM[1]	Tool	Algorithm	Avg. Accuracy	Modality	Validation	CC[2]	S[3]
Nandi et al.[36]	DEAP [233]	EDA+RB	FFNN	TFF	FedAvg	81.92% (VA)	Bi-Modal	SID[4]	×	✓
Tara Hassani [100]	CASE[66]	GSR	CNN	TFF	FedAvg	79% (V), 68% (A)	Single	SD[4]	×	×
Agrawal et al.[149]	DEAP [233] [72]	EEG	2D-CNN	TFF	FedAvg	66.99% (V), 70.10% (A), 72.22% (D)	Single	SD[4]	×	×
Proposed F-MERS	**DEAP [233]**	**EEG+GSR+RESP**	**MLP**	**TFF**	**FedAvg**	**89.02% (V)**, 86.50% (A), 84.02% (D)	**Multi-Modal**	SID[4]	✓	✓
	DREAMER [135]	**EEG+ECG**				**79.02% (V)**, 74.33% (A), 72.24% (D)				
	AMIGOS [125]	**EEG+ECG+GSR**				**80.10% (V)**, 87.90% (A), 81.06% (D)				
Proposed F-MERS	**DEAP [233]**	**EEG+GSR+RESP**	**MLP**	**TFF**	**FedAvg**	**86.20% (V)**, 88.10% (A), 86.52% (D)	**Multi-Modal**	SD[5]	✓	✓
	DREAMER [135]	**EEG+ECG**				**78.10% (V)**, 74.66% (A), 73.43% (D)				
	AMIGOS [125]	**EEG+ECG+GSR**				**80.98% (V)**, 86.80% (A), 83.06% (D)				

[1]Classification Model (CM), [2]Communication Computation (CC), [3]Scalability (S)
[4]Subject Independent (SID), [5]Subject Dependent (SD)

6.3 Summary

The proposed novel FL-based Multi-modal Emotion Recognition System (F-MERS) architecture successfully and accurately classify human emotions while protecting sensitive physiological information. It improves prior work in emotion recognition by generating a federated environment using federated averaging (FedAvg) at the server. The training and classification are performed at the client's end to protect data privacy from data breaches and sensitive information scenarios by not sharing the complete raw data (available at the clients' end) with other entities and the global server. The contributions of proposed F-MERS architecture are prominently evident from the results stating that the multi-modal F-MERS architecture outperforms single modalities, including EEG, ECG, GSR, and RESP.

The three datasets (AMIGOS [125], DEAP [233] and DREAMER [135]) validate the results for different iterations and varying rounds, concluding the model to be robust, scalable, and performs accurately. It disagrees with the prior works on emotion recognition in that they have not considered the privacy concerns for the user's physiological data. This chapter concludes that emotion recognition with a single modality is less accurate than multi-modal physiological signal data. Hence, the proposed FL-enabled multi-modal emotion recognition system can assist in better personalized emotional care with the security of personal data privacy while dealing with emotional distress.

This chapter is based on the following work:

- **J3: Neha Gahlan**, and Divyashikha Sethia. ”Federated learning inspired privacy sensitive emotion recognition based on multi-modal physiological sensors.” **Cluster Computing** (2023): 1-23. (**SCIE, Impact Factor: 4.4, Publisher: Springer**). Doi: `https://doi.org/10.1007/s10586-023-04133-4.`

Chapter 7

Attention for Variation Data Heterogeneity in multi-modal ERS

Combining multiple physiological signals in a multi-modal emotion recognition architecture results in Variation Data Heterogeneity (VDH) due to the distinct characteristics and variations of each signal. The existing works of FL for ERS did not address this data heterogeneity issue. The following chapter discusses addressing and reducing this issue of VDH in detail.

This chapter introduces a unique novel architecture *Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data* (**AFLEMP**) to create a decentralized environment by creating a client end (with subjects' multi-modal physiological signal data) and a server end for aggregating the updates. It runs the local model update at the client end and then sends the model updates (gradients) to the central server for averaging them. The proposed AFLEMP architecture resolves the challenge of VDH arising from variations in multi-modal physiological signal data at the client's end by incorporating an attention mechanism at the client's end. An attention mechanism is useful in this context as it selectively focuses on the most essential or relevant features from each signal, reducing the impact of VDH at the client end. The attention mechanism ensures that the local training model (at the client end) gives more weight to the essential features from the multi-modal signal data, thereby sending weights of only those essential features (after training at the client end) to the server. Hence, it reduces the communication overhead between the client and server as it only sends the weights from essential features rather than the weights from all the features. The proposed AFLEMP integrates three different attention mechanisms with an Artificial Neural Network (ANN) at the client end and evaluates them using two different datasets.

Previous literature of FL for emotion classification using physiological signals have used only the Valence and Arousal dimensions of the emotion model (2D-VA model of emotion), which do not cover a wide spectrum of emotions, as described earlier in Chapter 2, subsection 2.2.1. To fill this research gap, the proposed AFLEMP integrates Dominance along with Arousal and Valence (from the 3D-VAD model of emotion) together and performs binary and (Valence, Arousal, Dominance in low/high class) octal classification (Valence-Arousal-Dominance in eight emotion classes) of emotions.

Following are the main contributions of this chapter:

1. ***Proposal of AFLEMP architecture combining the physiological signals:*** It combines EEG, GSR, and ECG using Feature-level Fusion (*Multi-modality*). It utilizes an ANN classifier as a base model for performing binary (using Valence, Arousal and Dominance individually) and octal classification (using Valence-Arousal-Dominance together) of emotions using a 3D-VAD model of emotions at the client end. *AFLEMP* integrates three different attention mechanisms with ANN: **(a)** Generalized Attention, **(b)** Self-Attention, and **(c)** Transformer, to reduce the VDH at the client end.

2. ***Performance evaluation of proposed AFLEMP architecture:*** It validates the two benchmark datasets: AMIGOS [125] and DREAMER [135] using Leave-one-out cross validation (LOOCV) technique (*Validation*). The proposed AFLEMP architecture evaluates two classification scenarios: binary classification, where emotions are categorized into two distinct classes (e.g., high and low Valence), and octal classification, which expands the categorization into eight emotional states. This dual approach allows AFLEMP to demonstrate its effectiveness, providing insights into its performance for both simpler and more diverse emotions. The inclusion of octal classification (VAD together) highlights the ability of the proposed AFLEMP to handle more diverse emotions. It provides the scalability (in terms of rounds and iterations) and reports the communication

computation (in terms of training and aggregation times) for the proposed *AFLEMP* architecture.

7.1 Variation Data Heterogeneity (VDH)

There is a challenge in combining multi-modal physiological data for Emotion Recognition Systems (ERS) known as Variation Data Heterogeneity (VDH). This issue arises due to differences in the intrinsic characteristics of each physiological signal (such as EEG, ECG, and GSR) captured from different wearable sensors. Subjects exhibit variations in their physiological responses during emotion elicitation, and these differences lead to data heterogeneity, making it difficult to integrate and analyze such diverse data sources effectively. Equation 7.1 presents the fused feature vector (denoted as F) input consisting of multi-modal physiological data (EEG, ECG, GSR). This equation presents the VDH arising from combining the EEG, ECG, and GSR signal data.

$$F = [F_{\text{EEG}}, F_{\text{ECG}}, F_{\text{GSR}}] \tag{7.1}$$

The proposed AFLEMP architecture utilizes attention mechanisms to dynamically weigh the contribution of each physiological signal at the client end. The attention mechanisms assign weights to the subjects and sensor modalities based on relevance and quality. Equation 7.2 presents the computation of the attention weights as follows:

$$\alpha = \text{Attention}_{Trans}(F, \theta) \tag{7.2}$$
$$w = f_\theta(F, \alpha) \tag{7.3}$$

Where, the attention mechanism (Attention_{Trans}) computes attention weights (α) based on the input data (F) and the model parameters (θ). The local model is represented by f_θ. Equation 7.3 gives the computation of weights (w) by the local model f_θ, from the input data F, and the attention weights α (computed from Equation 7.2).

Table 7.1: Brief description of datasets for AFLEMP.

Description/Dataset	AMIGOS [125]	DREAMER [135]
No. of subjects	40	23 (14 male, 9 female)
Physiological Signals	EEG, ECG, GSR	EEG, ECG
Video Content	16 short videos, 4 long videos	18 videos
Video Duration	57-155 seconds, 14-15 minutes	65 - 393 seconds
Label Matrix	16 x 3	18 x 3
Emotion Dimensions	Arousal, Valence, Dominance	
Emotion Assessment	SAM (Self Assessment Manikins)	
EEG electrodes	14	14
ECG electrodes	2	2
GSR electrodes	1	-
RESP electrodes	-	-

7.2 Experimental Methodology

The proposed AFLEMP architecture utilize the three emotion dimensions: Valence, Arousal and Dominance from the Mehrabian and Russell's 3D-VAD model of emotions to classify the emotions [33, 127]. It performs binary classification for low/high Arousal, Valence and Dominance, and octal classification for eight different emotion classes using VAD (discussed later in detail in Subsection 7.2.1.3).

7.2.1 Data Processing

1. **Datasets Description**

   The proposed AFLEMP architecture validates the emotion benchmark datasets among the others, as they consist of multi-modal physiological data: AMIGOS [125] and DREAMER [135]. Detailed descriptions of all the datasets with their data processing, including data clipping and labelling, are given in section 3.9.1. Table 7.1 gives a brief description of all the datasets.

2. **Creation of Emotion Classification Labels:** The proposed AFLEMP architecture utilizes a dimensional emotion model (3D-VAD). For emotion evaluation, a threshold of 4.5 was applied for the AMIGOS dataset and a threshold of 3 for the DREAMER dataset. Based on these dimensions, two types of classification labels are outlined below (as detailed earlier in Chapter 2, subsection 2.2.1):

   (a) **Binary Classification:** In this case, emotions are grouped into two

Table 7.2: Mapping of emotions using 3D-VAD for AFLEMP.

Valence	Arousal	Dominance	Emotions
Low	Low	Low	Sorrow
Low	Low	High	Disgust
Low	High	Low	Fear
Low	High	High	Anger
High	Low	Low	Happiness
High	Low	High	Calm
High	High	Low	Surprise
High	High	High	Excitement

categories, ”low” or ”high,” depending on whether the values of Valence, arousal, or dominance are below or above the threshold [172]. The classification labels obtained here are Low Valence (LV) / High Valence (HV), Low Arousal (LA) / High Arousal (HA), and Low Dominance (LD) / High Dominance (HD).

(b) **Octal Classification:** In this case, the combination of all possible ”low” and ”high” values across the three dimensions results in eight different emotions, providing a more nuanced classification [172]. The classification labels obtained here for Valence-Arousal-Dominance (VAD) are HVHAHD / HVHALD / HVLAHD / HVLALD / LVHAHD / LVHALD / LVLAHD / LVLALD. Table 7.2 gives the mapping for these class labels into emotions.

3. **Feature Extraction:** After pre-processing the EEG, ECG and GSR signals from the datasets, an overlapping sliding window of 2 seconds with 50% overlap segments the signals [135]. Hence, the resulting segments will overlap by 1 second. Overlapping windows facilitate more reliable statistical estimation by increasing the number of data points used for calculations [182]. It reduces the risk of information loss at the boundaries of individual windows, as important signal characteristics that span adjacent windows are captured by overlapping regions [86]. Overlapping windows identify essential physiological phenomena, such as peaks in the case of ECG signals, and study their characteristics. After segmenting the signal, we extract a specific set of features, i.e. 219 features from the AMIGOS [125] dataset and 208 features from the DREAMER [135] dataset given in Table 7.3 and detailed in section 2.3.

4. **Feature Fusion:** The proposed AFLEMP architecture combines the features

Table 7.3: Extracted features from ECG, EEG, and GSR signals for AFLEMP.

Physiological Signal	**Domain**	**Features**
EEG (196 features)	**Time (6 x 14)**	Mean, Variance, Standard Deviation, Hjorth Mobility, Activity, Complexity
	Frequency (2 x 4 x 14)	Power Spectral Density, Band Power (alpha, beta, theta, delta)
ECG (12 features)	**Time (5 x 2)**	SDNN, RMSSD, SDSD, HRV, TINN
	Frequency (1 x 2)	Power Spectral Density
GSR (11 features)	**Time (2 x 5)**	Skin Conductance Level (SCL) , Skin Conductance Resistance (SCR)
	Frequency (1 x 1)	Power Spectral Density

extracted from the physiological signals data (EEG, ECG and GSR) via the Feature-Level Fusion (FLF) method. This process extracts features individually from each signal and then concatenates them together as a single vector. It preserves all the information from each modality without losing any essential features. The concatenation is expressed mathematically in Equations 7.4 and 7.5.

$$F_A = f_{eeg}^{(196)} + f_{ecg}^{(12)} + f_{gsr}^{(11)} = f_{eeg+ecg+gsr}^{(219)} \tag{7.4}$$

$$F_D = f_{eeg}^{(196)} + f_{ecg}^{(12)} = f_{eeg+ecg}^{(208)} \tag{7.5}$$

Where f_{eeg}, f_{ecg} and f_{gsr} presents the number of features of EEG, ECG and GSR signals, respectively. F_A gives concatenated feature vector for the AMIGOS [125] dataset and F_D gives for the DREAMER [135] dataset. After concatenating the feature vectors, the fused feature vector consists of many features, i.e., 219 for AMIGOS [125] and 208 for DREAMER [135].

7.2.2 Centralized ANN with Attention

The centralized environment uses the global dataset (data from all subjects) as input to the ANN integrated with attention mechanisms. This environment is created by splitting the global data into training and testing sets of 80% and 20%, respectively. In a centralized environment, clients upload their local datasets to a trusted central server, while in a decentralized, federated environment, clients maintain their private data locally. ANN is the base classifier for both the centralized environment and

Table 7.4: Model parameters of AFLEMP architecture at the client end.

Layers	**Output Shape**	**Activation Function**	**Regularization**	**Hyperparameters**
Input	(None, 219/208)	None	None	None
Attention Mechanism (Generalized, Self-attention, Transformer)	(None, 219/208)	Softmax	None	Attention Heads=6 (Transformer), Use bias=False, Attention dropout=0.2
Dense	(None, 128)	ReLU	L2 (0.01)	Dropout=0.2
Dense	(None, 64)	ReLU	L2 (0.01)	Dropout=0.2
Output	(None, 1)	Sigmoid (Binary) Softmax (Octal)	None	None
Optimizer = SGD, Learning Rate =0.001				

decentralized FL. ANNs are a flexible and powerful deep-learning network to easily handle large datasets and achieve state-of-the-art performance [30, 150, 191]. Table 7.4 presents the parameters at each layer for the ANN model embedded with the attention mechanism layers (including generalized attention, self-attention, and Transformer).

7.2.3 Attention Mechanisms

The proposed emotion recognition architecture implements three different attention mechanisms, explained in detail as follows:

1. **Generalized Attention:** The generalized attention mechanism allows the model to selectively focus on different modalities within the input data, improving its ability to capture relevant information and make accurate predictions [217]. For the proposed AFLEMP architecture, the input data is a fused feature vector of the EEG, ECG and GSR features. The Equation 7.6 presents computation from a weighted sum of the values, where the similarity between the query and the keys determines the weights. The weighted output is fed into the subsequent layer of the ANN model.

$$\text{Attention}_{Gen}(Q,K,V) = \text{softmax}(Q \cdot K)V \tag{7.6}$$

Here, the Softmax function in the attention layer computes the attention weights, $Q \cdot K$ is the dot product between the query vector (Q) and the matrix of key vectors (K). V is the matrix of value vectors.

2. **Self-Attention:** The self-attention mechanism is a method that empowers the model to selectively concentrate on different portions of the input sequence, thereby enhancing the precision of predictions [41, 68]. It has shown excellent results on the emotion recognition system based on physiological data [141, 266]. The single-head self-attention mechanism would compute a weighted sum of the input fused features based on their relevance to emotions. The weights are learned during training and depend on the input data. The generated weighted input sequence is then passed to the next layer of the ANN model, where the features are transformed and used for classification. The self-attention function is expressed mathematically as in Equation 7.7:

$$\text{Attention}_{Self}(Q,K,V) = \text{softmax}(\frac{QK^t}{\sqrt{d_k}})V \tag{7.7}$$

Here, QK^t is the dot product between the query vector Q and the transposed matrix of key vectors K^t. $\sqrt{d_k}$ is a scaling factor to prevent the dot product from growing too large. d_k is the dimensionality of the key vectors, which is typically smaller than the dimensionality of the query and value vectors.

3. **Transformer:** The transformer attention is a robust mechanism to learn the strong relationship between different features in a fused feature vector (EEG+ECG+GSR). It benefits emotion recognition, allowing the model to learn which features are most important for classifying emotions [271]. Its application in a federated environment reduces the data transfer requirements between the client and server. By enabling clients to focus on essential input data segments selectively, it minimizes the amount of data exchange. It reduces the communication overhead and makes the training process more efficient.

   The proposed AFLEMP architecture implements a transformer by incorporating multi-head self-attention (heads = 6), followed by a normalization layer. The output of this mechanism is fed into the ANN model for further classification. The mathematical computations is defined below in Equations 7.8, 7.9 and 7.10:

$$\text{Attention}_{Trans} = \text{MultiHead}(Q,K,V) \tag{7.8}$$

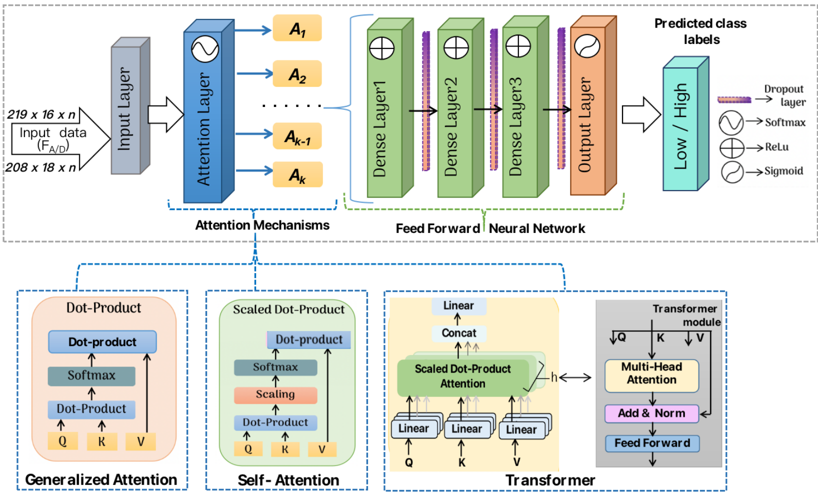


Figure 7.1: Architecture of the ANN with attention layers for AFLEMP.

$$\text{MultiHead}(Q,K,V) = \text{Concat}(head_1,...,head_6)X^o \tag{7.9}$$

$$\text{where, head}_w = \text{Attention}_{Self}(Q,K,V) \tag{7.10}$$

Equation 7.8 introduces multi-head attention, where six attention heads concurrently process the data using Equation 7.9 and 7.10, providing diverse perspectives. The resulting outputs are concatenated and transformed, capturing important patterns in the input. Equation 7.10 clarifies that each attention head operates using the same underlying attention mechanism, contributing distinct insights. Figure 7.1 illustrates the diagrammatic flow of the proposed AFLEMP architecture with attention layers (different mechanisms).

7.2.4 Architecture of AFLEMP

The proposed AFLEMP architecture aims to advance the field of emotion recognition by becoming a robust, privacy-preserving FL process that can effectively learn from diverse client data sets to create a more accurate and reliable global model for emotion recognition. The primary aim is to develop an FL process that effectively leverages the heterogeneous data distribution present in individual clients to build a more accurate model for emotion recognition using multi-modal physiological signal data. The proposed AFLEMP architecture adopts Horizontal Federated Learning (HFL), as client data distribution has similar features but differs in sample space. It is a promising approach to FL that leverages larger datasets and has the potential for increased model accuracy and generalization [270].

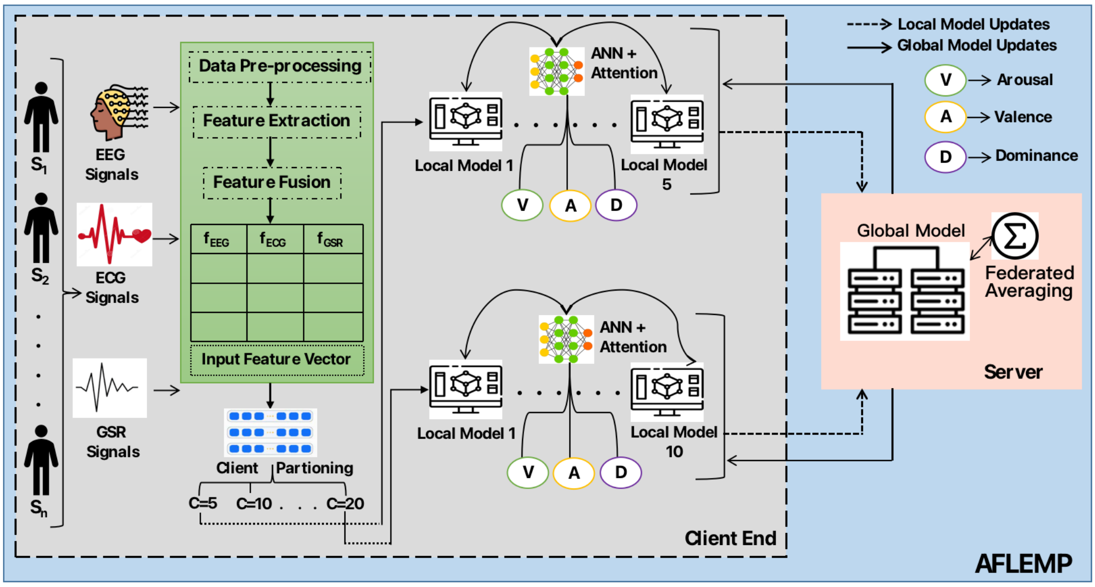


Figure 7.2: Complete workflow for AFLEMP architecture with AMIGOS [125].

1. **AFLEMP**

   The proposed Attention-based Federated Learning for Emotion recognition using the Multi-Modal Physiological data (AFLEMP) architecture is a coalesced federated learning approach. It integrates different attention mechanisms (Generalized attention, Self-attention, and Transformer) with an ANN to generate weights and perform classification at the client end. It performs federated averaging at the server to aggregate the local model's weights from clients. Figure 7.2 illustrates the complete architecture, showing every step, starting from the initial data collection, data pre-processing, feature extraction and fusion, federated environment creation and then the final classification at the client end.

2. **Federated Environment Setup**

   The proposed AFLEMP architecture uses TFF [208] as an FL tool to create a federated learning environment and uses Federated Averaging (FedAvg [136]) to collect and aggregate client model updates. Equation 7.11 gives the computation for the federated averaging at the server end. The algorithm for the FedAvg is described in detail earlier in Chapter 6.

$$w_t^g = \frac{1}{N_{total}} \sum_{i=1}^{N_{total}} w_{t,i}^l \tag{7.11}$$

Table 7.5: Federated data distribution for all clients (C) with different subjects (S) for AFLEMP.

AMIGOS (Subjects = 40)			
Clients = 5 **Subjects = 10**	**Clients = 10** **Subjects = 21**	**Clients = 15** **Subjects = 33**	**Clients = 20** **Subjects = 40**
C1 = S 1(1), C2 = S 2-3(2), C3 = S 4-5(2), C4 = S 6-7(2), C5 = S 8-10(3)	C1 = S 1(1), C2 = S 2-3(2), C3 = S 4-5(2), C4 = S 6-7(2), C5 = S 8-9(2), C6 = S 10-11(2), C7 = S 12-13(2), C8 = S 14-15(2), C9 = S 16-18(3), C10=S 19-21(3)	C1 = S 1(1), C2 = S 2(1), C3 = S 3-4(2), C4 = S 5-6(2), C5 = S 7-8(2), C6 = S 9-10(2), C7 = S 11-12(2), C8 = S 13-14(2), C9 = S 15-16(2), C10 = S 17-18(2), C11 = S 19-20(2), C12 = S 21-23(3), C13 = S 24-26(3), C14 = S 27-29(3), C15 = S 30-33(4)	C1 = S 1(1), C2 = S 2(1), C3 = S 3(1), C4 = S 4(1), C5 = S 5-6(2), C6 = S 7-8(2), C7 = S 9-10(2), C8 = S 11-12(2), C9 = S 13-14(2), C10 = S 15-16(2), C11 = S 17-18(2), C12 = S 19-20(2), C13 = S 21-22(2), C14 = S 23-24(2), C15 = S 25-26(2), C16 = S 27-28(2), C17 = S 29-30(2), C18 = S 31-33(3), C19 = S 34-36(3), C20 = S 37-40(4)
DREAMER (Subjects = 23)			
Clients = 3 **Subjects = 4**	**Clients = 5** **Subjects = 8**	**Clients = 7** **Subjects = 13**	**Clients = 10** **Subjects = 23**
C1 = S 1(1), C2 = S 2(1), C3 = S 3-4(2)	C1 = S 1(1), C2 = S 2(1), C3 = S 3-4(2), C4 = S 5-6(2), C5 = S 7-8(2)	C1 = S 1(1), C2 = S 2(1), C3 = S 3(1), C4 = S 4-5(2), C5 = S 6-7(2), C6 = S 8-10(3), C7 = S 11-13(3)	C1 = S 1(1), C2 = S 2(1), C3 = S 3(1), C4 = S 4-5(2), C5 = S 6-7(2), C6 = S 8-9(2), C7 = S 10-12(3), C8 = S 13-15(3), C9 = S 16-19(4), C10 = S 20-23(4)

Here w_t^g is the aggregated weight at the global server in time t, $w_{t,i}^l$ are the weights received from all local models in time t, and N_{total} is the total number of the local model participating for aggregation. The federated environment setup follows these steps: (1) *Creation of Local Model (client end)*; (2) *Creation of Global Model (server end)*; (3) *Local Model Update (client end).*

3. **Federated Data Partitioning**

   The proposed federated architecture adopts HFL, as all clients have the same features but differ in sample space (no. of subjects). Both the AMIGOS [125] and DREAMER [135] datasets are divided into four client divisions (C= 5, 10, 15, 20 /C=3, 5, 7, 10) as shown in Table 7.5, and the experiments ran for three rounds of aggregation (R= 50, 100, 200). One-way ANOVA testing is performed (with a significance value of 0.05 [199]) using IBM SPSS Statistics to test whether there are significant differences between the different client distributions. The p-value for AMIGOS [125] dataset is 0.045 and for DREAMER [135] is 0.037. Both are less than 0.05, implying significant differences between client distributions.

7.2.5 Experimental Setup

The tests and trials for the proposed AFLEMP architecture use Python 3.8 on a MacBook Air with a 1.6 GHz dual-core Intel Core i5 and Google Colab's Pro plus NVIDIA V100

Table 7.6: Train and test set of clients for proposed AFLEMP for both the datasets.

AMIGOS [125]				
No. of Clients	**C=5**	**C=10**	**C=15**	**C=20**
Training/Testing Clients	4/1	9/1	14/1	19/1
DREAMER [135]				
No. of Clients	**C=3**	**C=5**	**C=7**	**C=10**
Training/Testing Clients	2/1	4/1	6/1	9/1

GPU, CUDA version 11.2. Models are generated using the TensorFlow (TF version-2.6.0), TensorFlow Federated (TFF version-0.19.0) and Keras frameworks. During each round of aggregation, clients were required to train 200 epochs locally, with a batch size of 1024.

7.2.6 Evaluation Measures for AFLEMP

The following measures evaluate the proposed AFLEMP architecture (as detailed in Section 5.6). **Performance Measures:** Binary (low/high classes) and Octal (eight emotion classes) Classification Accuracy (Training Accuracy and Testing Accuracy), Precision, Recall, and F1-Score. **Scalability Measures:** Training Time and Model Accuracy with different client distributions and iteration rounds. **Communication Computation Measure:** Averaging time for different client distribution and iteration rounds, and Convergence Speed via Categorical Cross Entropy Loss.

7.2.7 Testing of the architecture

The proposed AFLEMP architecture utilizes a widely used Leave-one-out cross-validation (LOOCV) strategy to test the federated architecture. This approach treats each client as a test set once while using the rest of the clients for training [117, 154]. In each step of the LOOCV, from the C available clients, the samples of one client are used as the test set, and the samples of the remaining C-1 clients are used as the training set [125]. For instance, in the proposed federated learning architecture, when clients=5, the architecture allocates 4 clients for training, leaving the remaining 1 for testing, and this process iterates 5 times (Table 7.6 illustrates for all clients). The architecture reports results based on the average values across all iterations and evaluates model performance using metrics like accuracy, precision, recall, and the F1-score.

Table 7.7: Training and testing accuracies for **AFLEMP** for different attention mechanisms.

AMIGOS [125] (EEG+ECG+GSR)								
Rounds=200, Clients=20	**Binary classification (low/high classes)**						**Octal classification (8 emotion classes)**	
	Valence		**Arousal**		**Dominance**		**VAD**	
Attention Mechanisms	**Train accuracy**	**Test accuracy**	**Train accuracy**	**Test accuracy**	**Train accuracy**	**Test accuracy**	**Train accuracy**	**Test accuracy**
Without Attention	76.11%	75.34%	77.10%	78.21%	78.81%	77.45%	80.12%	79.84%
Generalized Attention	77.10%	76.11%	78.78%	77.43%	79.78%	78.67%	80.91%	80.21%
Self-Attention	78.95%	78.22%	79.44%	79.23%	79.88%	79.45%	81.56%	81.84%
Transformer	**82.77%**	**82.22%**	**80.58%**	**81.11%**	**81.50%**	**81.10%**	**83.04%**	**83.04%**
DREAMER [135] (EEG+ECG)								
Rounds=200, Clients=10	**Binary classification (low/high classes)**						**Octal classification (8 emotion classes)**	
	Valence		**Arousal**		**Dominance**		**VAD**	
Attention Mechanisms	**Train accuracy**	**Test accuracy**	**Train accuracy**	**Test accuracy**	**Train accuracy**	**Test accuracy**	**Train accuracy**	**Test accuracy**
Without Attention	70.10%	69.20%	69.70%	68.65%	71.65%	70.34%	73.88%	73.11%
Generalized Attention	72.11%	71.55%	71.20%	70.11%	72.11%	71.22%	74.10%	73.56%
Self-Attention	74.89%	73.55%	75.55%	74.45%	74.12%	73.78%	75.88%	75.65%
Transformer	**77.21%**	**76.56%**	**77.11%**	**76.14%**	**76.11%**	**75.55%**	**78.55%**	**77.34%**

7.3 Results and Discussion

In a federated environment, the local data stays at the client's end, and the global server does not have direct access. Consequently, emotion classification is carried out locally solely on each client. This section presents the results of the proposed AFLEMP architecture for classifying different emotion dimensions. The results are presented in the following categorical way:

7.3.1 Accuracy comparison for proposed multi-modal AFLEMP with different attention mechanisms

The proposed AFLEMP architecture experiments with three different attention mechanisms using two datasets, AMIGOS [125] and DREAMER [135], for which Table 7.7 presents the training and testing accuracies with 200 rounds of iterations. The following insights are derived from the tabular results.

1. ***Attention Mechanisms:*** Table 7.7 clearly shows that Transformer is performing best among the other attention mechanisms with testing accuracy of 83.04% on AMIGOS [125] and 77.34% on DREAMER [135] for octal classification (VAD), and 81.40% on AMIGOS [125] and 75.95% on DREAMER [135] for binary classification (average of valence, arousal, dominance).

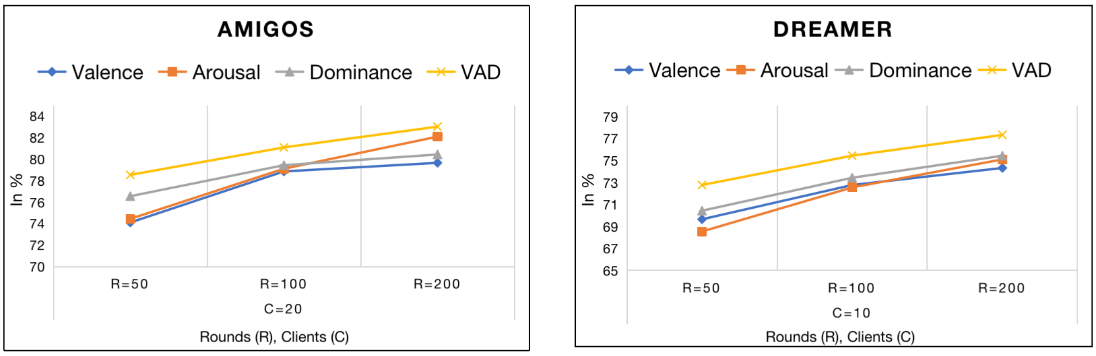


Figure 7.3: Testing accuracies for **AFLEMP** for Valence, Arousal, Dominance, and 3D-VAD.

2. ***Emotion dimensions:*** Table 7.7 gives the results for all the attention mechanisms with both binary (Valence, Arousal, Dominance individually) and octal (Valence-Arousal-Dominance together) classifications. It clearly shows that the octal classification with eight emotion classes (from VAD) is performing better than the binary classification with low and high classes (from valence, arousal, dominance individually) for each of the attention mechanisms. Figure 7.3 represents the testing accuracy for octal (VAD) and binary (valence, arousal and dominance) classifications for the Transformer with different rounds of iterations and maximum no. of clients (C=20 for AMIGOS [125] and C=10 for DREAMER [135]). It clearly illustrates that the proposed multi-modal AFLEMP architecture is performing better classification with VAD than binary classifications in all three rounds of iterations (R=50, 100 and 200).

3. ***Training and testing accuracies:*** The Training and Testing accuracies given in Table 7.7 show a difference of 1-2%, reflecting that there is no overfitting in the proposed AFLEMP architecture. Both the reported training and testing accuracies show the transformer's best performance with octal classification.

Figure 7.4 gives the testing accuracy curves with multiple iterations for multi-modal physiological data. The following points are inferred from the graph results:

1. ***Iteration Rounds:*** The best results are obtained with **200 rounds** of aggregation for all the client distribution, and this result is consistent across all three attention mechanisms and both datasets. It is the convergence point where the proposed AFLEMP architecture performs best.

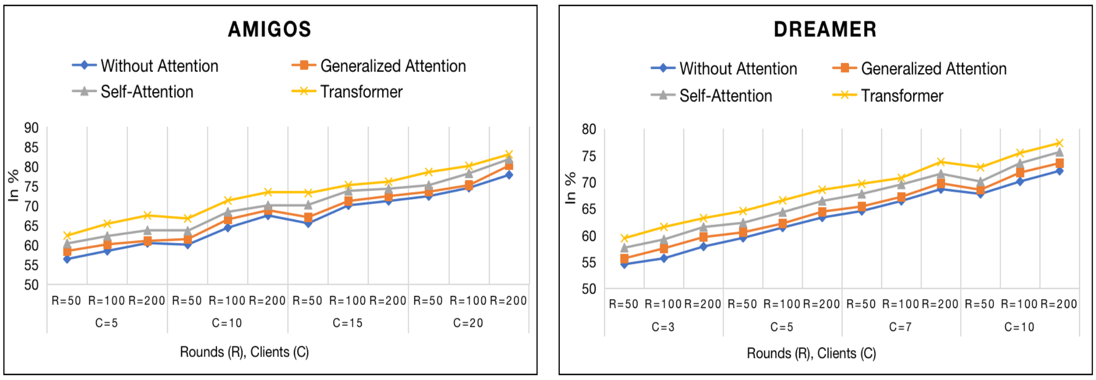


Figure 7.4: Testing accuracies for the **AFLEMP** for 3D-VAD in different iterations with different attention mechanisms.

Table 7.8: Evaluation measures for testing the AFLEMP (**with Transformer**) for 200 rounds.

Dataset ↓	Emotion Dimensions↓	Recall	Precision	F-1 Score	Accuracy
AMIGOS (EEG+ECG+GSR)	**Valence**	0.804	0.820	0.811	0.822
	Arousal	0.714	0.733	0.723	0.812
	Dominance	0.850	0.780	0.815	0.811
	VAD	0.860	0.830	0.845	0.830
DREAMER (EEG+ECG)	**Valence**	0.810	0.770	0.788	0.765
	Arousal	0.735	0.765	0.746	0.761
	Dominance	0.775	0.795	0.785	0.755
	VAD	0.787	0.872	0.795	0.773

2. ***Client Distributions:*** There is a significant increment in testing accuracy when **Clients=5 to Clients=20** for AMIGOS [125] (nearly 12-15%) and when **Clients=3 to Clients=10** for DREAMER [135] (nearly 13-15%).

3. ***Attention Mechanisms:*** The highest testing accuracy is achieved with **Transformer** (i.e. 83.04% for AMIGOS [125] and 77.34% for DREAMER [135]) outperforming the other two attention mechanisms.

7.3.2 Other performance measures for proposed multi-modal AFLEMP

7.3.2.1 Classification measures

The optimal values of evaluation measures, including Precision, Recall, F-1 score, and Testing Accuracy, are shown in Table 7.8. It shows that evaluation measures other than accuracy, i.e., precision, recall and f1-score is also giving optimal values for both octal and binary classifications for the proposed AFLEMP architecture using both datasets with the best-performing attention mechanism Transformer.

7.3.2.2 Scalability measures

1. **Accuracy:** For proving the proposed AFLEMP architecture to be scalable, its accuracy is computed for each round of iteration and with different client distributions as shown in Figure 7.4. The figure shows the testing accuracy for octal classification (VAD) with all three attention mechanisms for both datasets with Rounds=50, 100, and 200, and Client distributions, C=5, 10, 15, 20 for AMIGOS [125], and C=3, 5, 7, 10 for DREAMER [135].

2. **Training Time:** The training time contributes to the communication efficiency of the proposed AFLEMP architecture with different rounds of iterations and client distributions providing scalability. Table 7.9 presents the training time spent on the client end for running 200 local epochs for different client distributions (C=5, 10, 15 and 20 / C=3, 5, 7 and 10) and iterations of rounds (R=50, 100 and 200). The proposed AFLEMP architecture observes the best results when R=200, with a training time of 266.652 seconds for AMIGOS (with C=20) and 245.502 seconds for DREAMER (with C=10).

7.3.2.3 Communication Computation Measures

1. **Averaging Time:** Averaging time is majorly determining the impact on communication efficiency of the proposed AFLEMP architecture in terms of different rounds of iterations and client distributions. Table 7.9 presents the values in seconds, indicating the incremental change in averaging time resulting from variations in data size (C=5, 10, 15 and 20 / C=3, 5, 7 and 10) and the number of rounds (R=50, 100, and 200). The proposed AFLEMP architecture achieves the best results when R=200 with averaging time of 100.537 seconds for AMIGOS [125] (with Client =20) and 99.367 seconds for DREAMER [135] (with Client=10).

Table 7.9: Training and averaging time in seconds for the AFLEMP **(with Transformer).**

Epochs=200	AMIGOS [125]				DREAMER [135]			
Rounds=50	**Client=5**	**Client=10**	**Client=15**	**Client=20**	**Client=3**	**Client=5**	**Client=7**	**Client=10**
Training Time (in seconds) (Client End)	95.110	98.324	115.571	130.231	90.345	107.205	140.452	120.121
Averaging Time (in seconds) (Server End)	55.103	65.123	74.142	88.135	65.571	76.383	80.302	91.472
Rounds=100								
Training Time (in seconds) (Client End)	117.231	121.764	130.554	153.232	100.561	115.271	130.701	147.678
Averaging Time (in seconds) (Server End)	62.123	70.124	80.142	92.633	68.127	75.642	78.577	92.566
Rounds=200								
Training Time (in seconds) (Client End)	150.221	173.763	200.503	266.652	150.213	173.121	193.201	245.502
Averaging Time (in seconds) (Server End)	71.578	75.478	87.123	100.537	71.343	80.880	86.367	99.367

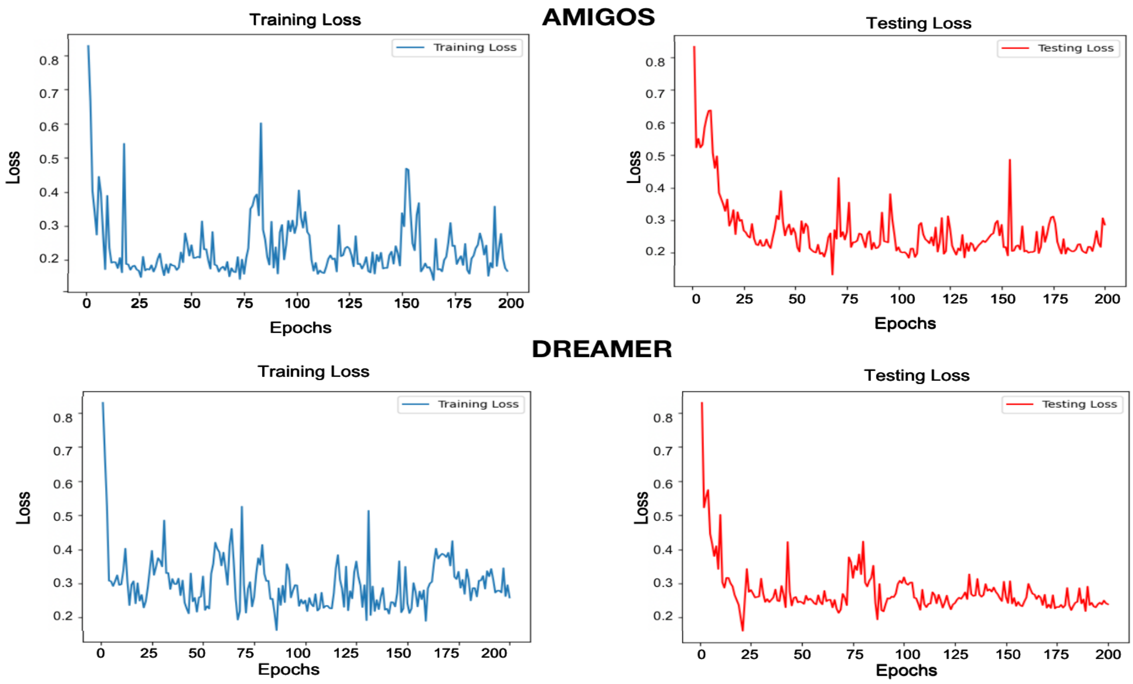


Figure 7.5: Loss curves for the AFLEMP for AMIGOS [125] and DREAMER [135] for 3D-VAD **(with Transformer)**.

2. **Convergence Speed:** To measure the convergence speed, categorical cross-entropy loss is computed, as show in Figure 7.5. It illustrates the categorical cross-entropy loss versus epoch curves for the AMIGOS [125] and DREAMER [135]

Table 7.10: Performance comparison with existing FL works for ERS with physiological signals.

FL for Emotion Recognition									
Reference	**Dataset**	**Physiological Signal**	**Classification Model**	**Modality**	**FL Methods**	**ED[1]**	**Testing Accuracy**	**S[3]**	**CC[2]**
Nandi et al. [36]	DEAP [233]	EDA, RESP	FFNN	Bi-Modal	FedAvg	2D	81.92% (VA)	✓	×
Tara Hassani [100]	CASE [134]	GSR	CNN	Single	FedAvg	2D	79% (VA)	×	×
Agrawal et al. [?]	DEAP [233]	EEG	2D-CNN	Single	FedAvg	3D	69.77% (avg. V,A,D)	×	×
Proposed (FL) AFLEMP	**AMIGOS [125]** **DREAMER [135]**	**EEG+ECG+GSR** **EEG+ECG**	**ANN+ Transformer**	**Multi-Modal**	**FedAvg**	**3D**	**83.04% (VAD)** **77.34% (VAD)**	✓	✓

[1]Emotion Dimension (ED), [2]Communication Computation (CC), [3]Scalability (S)

datasets for the proposed AFLEMP architecture. It reveals that the train and test sets follow a similar declining path for the eight emotion classes (VAD). The declining loss curves clearly indicate that the loss generated by the proposed AFLEMP architecture decreases progressively as the model improves its performance. This reduction in loss corresponds to the increase in accuracy over time, demonstrating that the model is learning and adapting effectively. By the time the training reaches 200 epochs, both the loss and accuracy stabilize, suggesting that the model has converged to an optimal solution.

7.3.3 Performance comparison of proposed multi-modal FL (AFLEMP) with existing FL works

Table 7.10 presents the comparative analysis between the proposed privacy-preserving federated AFLEMP architecture and other existing methods discussed in the literature utilizing FL for emotion recognition. It demonstrates that the proposed AFLEMP architecture outperforms the previous FL works for emotion recognition, achieving testing accuracy of 83.04% and 77.34% for AMIGOS [125] and DREAMER [135], respectively. Additionally, the proposed architecture also computes the scalability and communication computation, which is missing in the previous studies for FL.

7.4 Summary

The proposed novel *Attention-based Federated Learning for Emotion recognition using Multi-Modal Physiological data (AFLEMP)* architecture is able to accurately classify broader human emotions (using VAD) in a decentralized, federated learning environment. It integrates three attention mechanisms using an ANN model to train and classify different emotions at the client's end. The attention mechanism calculates the weights for essential features from multi-modal physiological data at each client end and sends the computed gradients to the global server. Embedding the attention mechanism in the proposed federated architecture reduces the amount of data transfer from client to server in a federated environment, improving communication efficiency and reducing the Variation Data Heterogeneity (VDH) in multi-modal physiological data. The result outcomes of the proposed AFLEMP architecture evidenced the reduction of data heterogeneity through improved communication, making the architecture robust.

The proposed AFLEMP architecture validates the three attention mechanisms, and their results prove that the Transformer performs best. It evaluates two emotion datasets, AMIGOS [125] and DREAMER [135], for multi-modal physiological signals - EEG, ECG, and GSR. Varying iterations evaluate the proposed AFLEMP architecture, including aggregation rounds (R=50, 100, and 200) and client distributions (C=5, 10, 15, 20/ C=3, 5, 7, 10). The consistency of the results for all clients suggests that the architecture is adept at handling the challenge of data heterogeneity. The proposed AFLEMP architecture is efficient (with manageable training and aggregation times) and scalable, maintains the accuracy and robustness of the emotion recognition model while preserving the privacy of the client's sensitive data and improves the previous federated learning approaches in emotion recognition.

<u>This chapter is based on the following work:</u>

- **Neha Gahlan**, and Divyashikha Sethia. ”AFLEMP: Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data.” **Biomedical Signal Processing and Control** 94 (2024): 106353. (**SCIE, Impact Factor: 5.1, Publisher: Elsevier**). Doi: `https://doi.org/10.1016/j.bspc.2024.106353.`

Chapter 8

Scaled Weighted Federated Averaging (SWFA) for Imbalanced Data Heterogeneity in Multi-modal ERS

In a real-world federated environment, each client may consist of different amounts or distributions of data, leading to an Imbalanced Data Heterogeneity (IDH), significantly impacting model aggregation at the server end. This imbalance can arise due to variations in the number of participants/clients, causing a disproportionate influence on the global model server. Such imbalances hinder the performance and generalizability of the ERS. However, existing FL works for ERS ignored this critical issue of IDH, resulting in models that may not perform well across real-world populations. The following chapter discusses this issue in greater detail, proposing a novel aggregation algorithm to address and reduce IDH, thereby improving the robustness and fairness of FL models for emotion recognition across distributed datasets.

This chapter introduces a novel Scaled Weighted Federated Averaging (SWFA) algorithm for aggregation at the server end. It is embedded (at the server end) in the proposed attention-based federated learning emotion recognition **AFLEMP** architecture using multi-modal physiological signal data (at the client end). The previous literature in FL introduced the FedAvg [51], FedPer [170], FedBoost [114] algorithms to average the weights at the server end received from each client [51, 283]. However, these algorithms cannot address the imbalances in the data distributed at the client end. Recent research proposed a Dynamic Weighted Federated Averaging

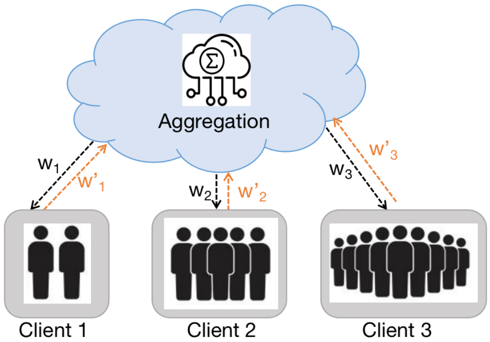


Figure 8.1: Imbalanced Data Heterogeneity (IDH) in a federated environment.

(DWFA) algorithm to address the imbalanced training and testing data at the client end [110]. However, it ignores the computational overheads, is susceptible to fluctuations in client performance, and misses client scalability, opening the way for better averaging.

Following are the main contributions of this chapter:

- ***Proposal of SWFA at server end in AFLEMP architecture:*** It implements the SWFA algorithm in AFLEMP at the server end for aggregation with a transformer at the client end for emotion recognition. SWFA aggregates the gradients computed from the multi-modal EEG, ECG, and GSR physiological signal data (*Multi-Modality*).

- ***Performance evaluation of proposed SWFA with AFLEMP architecture:*** It validates the two benchmark datasets: AMIGOS [125] and DREAMER [135] using Leave-one-out cross validation (LOOCV) technique (*Validation*). It provides scalability (in terms of rounds and iterations) and reports the communication computation (in terms of training and aggregation times) for the proposed SWFA with AFLEMP architecture. It provides a performance comparison of the proposed SWFA with other aggregation algorithms of FL (*with AFLEMP for emotion recognition*).

8.1 Imbalanced Data Heterogeneity (IDH)

Existing aggregation algorithms of federated learning fail to address the issue of IDH occurring at the client's end. Clients have different amounts of data available for training in a real-world scenario, which leads to IDH. For example, some clients may have a small amount of data, while others may have a large amount of data. The server aggregates these heterogeneous updates from the clients, but the imbalance primarily arises from the diversity in each client's data. Figure 8.1 gives an illustration of IDH in a federated environment.

Let there be an FL environment with clients denoted as $C = C^1, C^2, ...C^i$. Each client contains data from a different number of subjects, denoted as $S = S_1, S_2, S_3 ...., S_k$. The fused feature vector input consists of multi-modal physiological data (EEG, ECG, GSR) denoted as X ($X = [EEG + ECG + GSR]$). Equation 8.1 presents mathematically the IDH in different clients. For example, Client 1 (C^1) consists of data from subjects S_1 and S_2, whereas some other random client (C^4), consists of data from subjects S_{11} to S_{17} and so on.

$$||C^1_{S_1...S_2}|| \neq ||C^2_{S_3...S_5}|| \neq ||C^3_{S_6...S_{10}}|| \neq ||C^4_{S_{11}...S_{17}}|| \;.... \neq ||C^i_{S_{18}...S_k}|| \tag{8.1}$$

8.2 Experimental Methodology

8.2.1 Scaled Weighted Federated Averaging (SWFA)

In the standard FedAvg algorithm, the server aggregates local model updates from clients by computing a weighted average. The weight for each client is typically proportional to the size of its local dataset. It works well when the data is homogeneous and balanced across clients, but several issues arise when clients have imbalanced data distributions. In such a scenario of imbalanced data distributions, clients with larger datasets may dominate the model update, leading to a global model biased toward those clients.

The proposed Scaled Weighted Federated Averaging (SWFA) extends the standard FedAvg algorithm to handle the challenge of imbalanced data across clients in a federated environment. It focuses on improving the aggregation process at the server end by scaling and weighting client updates. SWFA introduces an additional scaling mechanism to account for these imbalances. The idea is to ensure that the aggregation process gives appropriate importance to each client's model update, regardless of data distribution. It involves scaling the client updates, in which each client's contribution is scaled based on the diversity/amount of the data it holds. It helps mitigate the impact of clients with large datasets. Equation 8.8 gives the computation for the scaled weighted federated averaging.

$$w_t^g = \frac{\sum_{i=1}^{N} w_i \cdot w_{t,i}^l}{\sum_{i=1}^{N} w_i} \tag{8.2}$$

Here, w_t^g is the aggregated weight at the global server in time t, and w_i is the weight assigned to client i, serving as the scaling factor. $w_{t,i}^l$ are the weights received from all client's local models in time t, and N is the total number of the local models participating in aggregation. Algorithm 2 provides the algorithm for the proposed SWFA as below:

Algorithm 2 Scaled Weighted Federated Averaging (SWFA)

Input: Client weights w_i for $i = 1, 2, \ldots, N$
Output: Scaled weighted average $\mathbf{w}_i^g$
Initialize global model weights $\mathbf{w}^g = 0$
Initialize total weight sum $S = 0$
for each client i from 1 to N **do**
    Update global model weights:

$$\mathbf{w}_i^g \leftarrow \mathbf{w}_i^g + w_i \cdot \mathbf{w}_{t,i}^l \tag{8.3}$$

    Update total weight sum:

$$S \leftarrow S + w_i \tag{8.4}$$

end for
Compute scaled weighted average for each parameter i:

$$\mathbf{w}_i^g = \frac{\sum_{i=1}^{N} w_i \cdot \mathbf{w}_{t,i}^l}{\sum_{i=1}^{N} w_i} \tag{8.5}$$

Return: Scaled weighted average $\mathbf{w}_i^g$

Key elements of the algorithm are:

- ***Scaling weights*** w_i**:** This is the scaling factor assigned to each client based on the data amounts. These weights control the contribution of each client's local model update at the global server.

- ***Local model weights*** $w_{t,i}^l$**:** These are the model weights that each client i calculates after training on its local data.

- ***Global model weights*** w_i^g***:*** This is the globally averaged model that is updated at the server after aggregating the client models.

Explanation of the steps:

- ***Step 1:*** The algorithm takes as input the client weights w_i for each client i (where $i = 1, 2, \ldots, N$). These weights determine the importance of each client in the aggregation process.

- ***Step 2:*** The algorithm outputs the **scaled weighted average** of model weights, denoted as w_i^g. This represents the new global model weights after federated averaging.

- ***Step 3:*** The global model weights are initialized to zero:

$$\mathbf{w}^g = 0$$

  This ensures that the aggregation starts from a neutral state.

- ***Step 4:*** A variable S is initialized to zero:

$$S = 0$$

  This variable will store the sum of client weights for normalization.

- ***Step 5:*** The algorithm iterates over all clients (1 to N) to aggregate their contributions.

- ***Step 6:*** For each client i, the local model weight $\mathbf{w}_{t,i}^l$ (trained on local data) is added to the global weight, scaled by the client's weight w_i:

$$\mathbf{w}_i^g \leftarrow \mathbf{w}_i^g + w_i \cdot \mathbf{w}_{t,i}^l$$

  This ensures that each client contributes proportionally based on their assigned weight w_i.

- ***Step 7:*** The total sum of client weights is updated as follows:

$$S \leftarrow S + w_i$$

- ***Step 8:*** The loop completes once all client updates have been aggregated.

- ***Step 9:*** The final global model weights are computed by normalizing the weighted sum:

$$\mathbf{w}_i^g = \frac{\sum_{i=1}^{N} w_i \cdot \mathbf{w}_{t,i}^l}{\sum_{i=1}^{N} w_i}$$

  This ensures that the global model update reflects a **true weighted average** rather than a simple sum.

- ***Step 10:*** The algorithm returns the scaled weighted average $\mathbf{w}_i^g$, which serves as the new global model update for the next training round.

SWFA helps the global model aggregator converge faster at the server end, as it balances client contributions more effectively, avoiding instability caused by data imbalances. Clients with disproportionately large or small datasets are appropriately scaled so that the global aggregator is not biased toward any single client.

8.2.2 AFLEMP with SWFA

The proposed multi-modal AFLEMP with SWFA architecture uses the Transformer with ANN at the client end (to reduce the VDH) and SWFA at the server end (to reduce the IDH). It classifies emotions using the three emotion dimensions: Valence, Arousal and Dominance from Mehrabian and Russell's 3D-VAD model of emotions [33, 127]. It performs binary classification for low/high Arousal, Valence and Dominance, and octal classification for eight different emotion classes using VAD. Figure 8.2 gives the architecture of AFLEMP with SWFA, and it is detailed in the following subsections.

8.2.2.1 Data Description

To evaluate the proposed AFLEMP with SWFA architecture, we perform comprehensive experiments on the publicly available multi-modal physiological datasets AMIGOS [125] and DREAMER [135]. Brief details of the two datasets are summarized in Table 8.1 and detailed information about the subjects, collection scenario, and physiological sensor devices and their pre-processing is given in section 3.9.1.

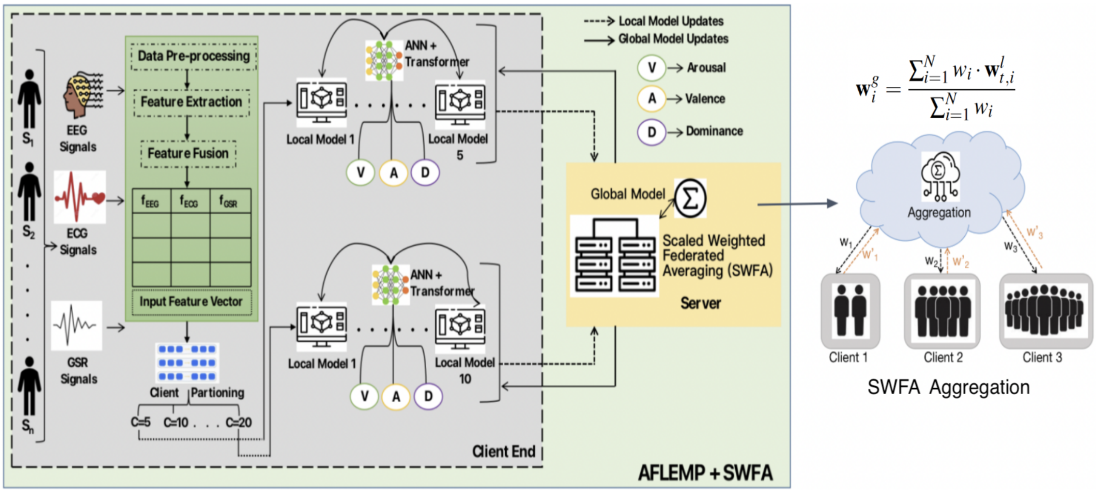


Figure 8.2: Architecture of proposed AFLEMP with proposed aggregation algorithm SWFA.

Table 8.1: Brief description of datasets for AFLEMP with SWFA.

Description/Dataset	**AMIGOS [125]**	**DREAMER [135]**
No. of subjects	40	23 (14 male, 9 female)
Physiological Signals	EEG, ECG, GSR	EEG, ECG
Video Content	16 short videos, 4 long videos	18 videos
Video Duration	57-155 seconds, 14-15 minutes	65 - 393 seconds
Label Matrix	16 x 3	18 x 3
Emotion Dimensions	Arousal, Valence, Dominance	
Emotion Assessment	SAM (Self Assessment Manikins)	
EEG electrodes	14	14
ECG electrodes	2	2
GSR electrodes	1	-
RESP electrodes	-	-

- The proposed AFLEMP with SWFA architecture considers the experiment involving 16 short videos from the AMIGOS [125] dataset, and 18 video stimuli from the DREAMER [135] dataset.

- For emotion evaluation, a threshold of 4.5 (lower than 4.5 as low and higher than 4.5 as high) was applied for the AMIGOS [125] dataset and a threshold of 3 (lower than 3 as low and higher than 3 as high) for the DREAMER [135] dataset.

Based on these dimensions, the proposed AFLEMP with SWFA architecture uses two types of classification labels as outlined below (detailed in earlier section 2.2.1):

1. **Binary Classification:** In this case, the classification labels obtained here are Low Valence (LV) / High Valence (HV), Low Arousal (LA) / High Arousal (HA),

Table 8.2: Extracted features from ECG, EEG and GSR signals for AFLEMP with SWFA.

Physiological Signal	Domain	Features
EEG (196 features)	**Time (6 x 14)**	Mean, Variance, Standard Deviation, Hjorth Parameters (Activity, Mobility, Complexity)
	Frequency (2 x 4 x 14)	Power Spectral Density (PSD), Band Power
ECG (12 features)	**Time (5 x 2)**	SDNN[1], RMSSD[2], HRV[3], SDSD[4], TINN[5]
	Frequency (1 x 2)	Power Spectral Density (PSD)
GSR (11 features)	**Time (2 x 5)**	SCL (Skin Conductance Level) SCR (Skin Conductance Resistance)
	Frequency (1 x 1)	Power Spectral Density (PSD)

and Low Dominance (LD) / High Dominance (HD) [172].

2. **Octal Classification:** In this case, the classification labels obtained here for Valence-Arousal-Dominance (VAD) are: HVHAHD / HVHALD / HVLAHD / HVLALD / LVHAHD / LVHALD / LVLAHD / LVLALD [172].

8.2.2.2 Feature Extraction

For the proposed AFLEMP with SWFA architecture, we extract 219 features from the AMIGOS [125] dataset and 208 from the DREAMER [135] dataset using an overlapping sliding window of 2 seconds (with 50% overlap), as listed in Table 8.2, (detailed in earlier section 2.3). The extracted features from the physiological signals data (EEG, ECG, GSR) are then combined using the Feature-Level Fusion (FLF) method as a concatenated feature vector. The concatenating is expressed mathematically in Equations 8.6 and 8.7.

$$F_A = f_{eeg}^{(196)} + f_{ecg}^{(12)} + f_{gsr}^{(11)} = f_{eeg+ecg+gsr}^{(219)} \tag{8.6}$$

$$F_D = f_{eeg}^{(196)} + f_{ecg}^{(12)} = f_{eeg+ecg}^{(208)} \tag{8.7}$$

Where, f_{eeg}, f_{ecg} and f_{gsr} present the number of features of EEG, ECG and GSR signals, respectively. F_A stands for concatenated feature vector for the AMIGOS [125] dataset and F_D for the DREAMER [135] dataset.

Table 8.3: Model parameters of the local model run at the client end in AFLEMP with SWFA.

Layers	Output Shape	Activation Function	Regularization	Hyperparameters
Input	(None, 219/208)	None	None	None
Attention Mechanism (Transformer)	(None, 219/208)	Softmax	None	Attention Heads=6, Use bias=False, Attention dropout=0.2
Dense	(None, 128)	ReLU	L2 (0.01)	Dropout=0.2
Dense	(None, 64)	ReLU	L2 (0.01)	Dropout=0.2
Output	(None, 1)	Sigmoid (Binary) Softmax (Octal)	None	None
Optimizer = SGD, Learning Rate =0.001				

8.2.2.3 Attention Mechanism - Transformer

The proposed AFLMEP with SWFA architecture uses the Transformer attention mechanism by incorporating multi-head self-attention (heads = 6), followed by a normalization layer. The output of this mechanism is fed into the ANN model for further classification.

8.2.2.4 Base Classifier - ANN

The proposed AFLEMP with SWFA architecture uses a Transformer with ANN at the client end. Table 8.3 presents the parameters at each layer for the ANN model with the Transformer. This model incorporates a Transformer-based attention mechanism followed by two dense layers for classification tasks. The input layer feeds into the attention mechanism, which has 6 attention heads, no bias, and a dropout, which further gets into the dense layers.

8.2.2.5 Federated data Partitioning

The proposed AFLEMP with SWFA architecture adopts HFL, as all clients have the same features but differ in sample space (no. of subjects). Both the AMIGOS [125] and DREAMER [135] dataset is divided into four client divisions (C= 5, 10, 15, 20 /C=3, 5, 7, 10) as shown in Table 8.4, and the experiments ran for three rounds of aggregation (R= 50, 100, 200). One-way ANOVA testing is performed (with a significance value of 0.05 [199]) using IBM SPSS Statistics to test whether there are significant differences between the different client distributions. The p-value for AMIGOS [125] dataset is 0.045 and for DREAMER [135] is 0.037. Both are less than 0.05, implying significant

Table 8.4: Federated data distribution for all clients (C) with different subjects (S) for AFLEMP with SWFA.

AMIGOS [125] (Subjects = 40)			
Clients = 5 Subjects = 10	**Clients = 10 Subjects = 21**	**Clients = 15 Subjects = 33**	**Clients = 20 Subjects = 40**
C1 = S 1(1), C2 = S 2-3(2), C3 = S 4-5(2), C4 = S 6-7(2), C5 = S 8-10(3)	C1 = S 1(1), C2 = S 2-3(2), C3 = S 4-5(2), C4 = S 6-7(2), C5 = S 8-9(2), C6 = S 10-11(2), C7 = S 12-13(2), C8 = S 14-15(2), C9 = S 16-18(3), C10=S 19-21(3)	C1 = S 1(1), C2 = S 2(1), C3 = S 3-4(2), C4 = S 5-6(2), C5 = S 7-8(2), C6 = S 9-10(2), C7 = S 11-12(2), C8 = S 13-14(2), C9 = S 15-16(2), C10 = S 17-18(2), C11 = S 19-20(2), C12 = S 21-23(3), C13 = S 24-26(3), C14 = S 27-29(3), C15 = S 30-33(4)	C1 = S 1(1), C2 = S 2(1), C3 = S 3(1), C4 = S 4(1), C5 = S 5-6(2), C6 = S 7-8(2), C7 = S 9-10(2), C8 = S 11-12(2), C9 = S 13-14(2), C10 = S 15-16(2), C11 = S 17-18(2), C12 = S 19-20(2), C13 = S 21-22(2), C14 = S 23-24(2), C15 = S 25-26(2), C16 = S 27-28(2), C17 = S 29-30(2), C18 = S 31-33(3), C19 = S 34-36(3), C20 = S 37-40(4)
DREAMER [135] (Subjects = 23)			
Clients = 3 Subjects = 4	**Clients = 5 Subjects = 8**	**Clients = 7 Subjects = 13**	**Clients = 10 Subjects = 23**
C1 = S 1(1), C2 = S 2(1), C3 = S 3-4(2)	C1 = S 1(1), C2 = S 2(1), C3 = S 3-4(2), C4 = S 5-6(2), C5 = S 7-8(2)	C1 = S 1(1), C2 = S 2(1), C3 = S 3(1), C4 = S 4-5(2), C5 = S 6-7(2), C6 = S 8-10(3), C7 = S 11-13(3)	C1 = S 1(1), C2 = S 2(1), C3 = S 3(1), C4 = S 4-5(2), C5 = S 6-7(2), C6 = S 8-9(2), C7 = S 10-12(3), C8 = S 13-15(3), C9 = S 16-19(4), C10 = S 20-23(4)

differences between client distributions.

8.2.2.6 Federated Environment via SWFA

The proposed algorithm uses TFF [208] as an FL tool to create a federated learning environment and uses SWFA to collect and aggregate client model updates. Equation 8.8 gives the computation for weighted federated averaging.

$$w_t^g = \frac{\sum_{i=1}^{N} w_i \cdot w_{t,i}^l}{\sum_{i=1}^{N} w_i} \tag{8.8}$$

Where w_t^g is the aggregated weight at the global server in time t, w_i is the weight assigned to client i, $w_{t,i}^l$ are the weights received from all client's local model in time t, and N is the total number of the local models participating for aggregation. Algorithm 2 provides the algorithm for the proposed SWFA (discussed above in Section 8.2.1). The federated environment setup follows these steps: (1) *Creation of Local Model (client end)*; (2) *Creation of Global Model (server end)*; (3) *Local Model Update (client end)*.

8.2.3 Experimental Setup

The tests and trials for the proposed AFLEMP with SWFA architecture use Python 3.8 on a MacBook Air with a 1.6 GHz dual-core Intel Core i5 and Google Colab's Pro plus NVIDIA V100 GPU, CUDA version 11.2. Models are generated using the

Table 8.5: Train and test set of clients for AFLEMP with SWFA.

AMIGOS [125]				
No. of Clients	**C=5**	**C=10**	**C=15**	**C=20**
Training/Testing Clients	4/1	9/1	14/1	19/1
DREAMER [135]				
No. of Clients	**C=3**	**C=5**	**C=7**	**C=10**
Training/Testing Clients	2/1	4/1	6/1	9/1

TensorFlow (TF version-2.6.0), TensorFlow Federated (TFF version-0.19.0) and Keras frameworks. During each round of aggregation, clients were required to train 200 epochs locally, with a batch size of 1024.

8.2.4 Evaluation Measures for AFLEMP with SWFA

The following measures evaluate the proposed AFLEMP with SWFA architecture (as detailed in Section 5.6). **Performance Measures:** Binary (low/high classes) and Octal (eight emotion classes) Classification Accuracy (Training Accuracy and Testing Accuracy), Precision, Recall, and F1-Score. **Scalability Measures:** Training Time and Model Accuracy with different client distributions and iteration rounds. **Communication Computation Measure:** Averaging time for different client distribution and iteration rounds, and Convergence Speed via Categorical Cross Entropy Loss.

8.2.5 Testing of the architecture

The Leave-one-out Cross Validation (LOOCV) strategy is used to test the AFLMEP with SWFA architecture. This approach treats each client as a test set once while using the rest of the clients for training [117, 125, 154]. Table 8.5 illustrates for all clients. The architecture reports results based on the average values across all iterations and evaluates model performance using metrics like accuracy, precision, recall, and the F1 measure.

8.3 Results and Discussion

In a federated environment, the local data stays at the client's end, and the global server does not have direct access. Consequently, emotion classification is carried out locally

Table 8.6: Training and testing accuracies for proposed AFLEMP with SWFA.

AMIGOS [125]								
Clients=20	**Valence**		**Arousal**		**Dominance**		**VAD**	
	Training Accuracy	**Testing Accuracy**	**Training Accuracy**	**Testing Accuracy**	**Training Accuracy**	**Testing Accuracy**	**Training Accuracy**	**Testing Accuracy**
Rounds=50	82.23%	81.11%	84.44%	83.21%	83.11%	82.12%	84.55%	83.11%
Rounds=100	84.34%	83.45%	86.21%	85.45%	85.55%	84.13%	87.11%	86.44%
Rounds=200	**88.98%**	**88.15%**	**89.78%**	**89.10%**	**88.11%**	**87.11%**	**90.96%**	**90.11**
DREAMER [135]								
Clients=10	**Valence**		**Arousal**		**Dominance**		**VAD**	
	Training Accuracy	**Testing Accuracy**	**Training Accuracy**	**Testing Accuracy**	**Training Accuracy**	**Testing Accuracy**	**Training Accuracy**	**Testing Accuracy**
Rounds=50	75.87%	74.32%	79.55%	78.21%	77.12%	76.45%	79.78%	78.34%
Rounds=100	79.94%	78.54%	82.55%	81.45%	80.15%	79.65%	82.11%	81.11%
Rounds=200	**84.88%**	**83.81%**	**85.66%**	**84.61%**	**84.15%**	**83.93%**	**86.34%**	**85.12%**

solely at each client's end. This section presents the results of the proposed AFLEMP with SWFA for classifying different emotion dimensions. The results are presented in the following subsections.

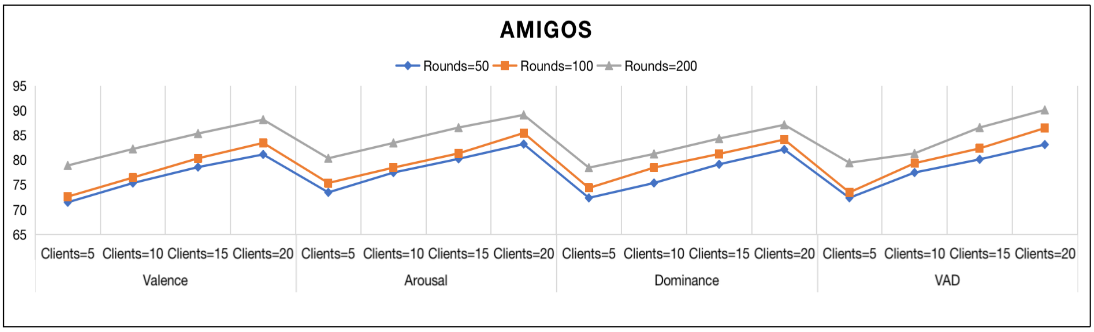
8.3.1 Performance measures for proposed multi-modal AFLEMP with SWFA

The proposed multi-modal AFLEMP architecture employs SWFA to create a global aggregator at the server end. It classifies emotions using two scenarios: one is binary classification, and the other is octal classification. The octal classification covers a broader range of emotions than the binary classification. The evaluation measures for the proposed AFLEMP with SWFA architecture are given in the following section below.

8.3.1.1 Classification Accuracies

Table 8.6 gives the Training and Testing accuracies For evaluating the performance of the proposed multi-modal AFLEMP with SWFA architecture for different iteration rounds. It gives the accuracy results for both the binary and octal classifications. The following points can be inferred from the Table 8.6:

1. It clearly states that the best results are obtained with 200 rounds of iterations with a maximum no. of clients. It gives results with only the maximum no of clients, but for the rest of the client distributions, Figure 8.3 graphically presents the testing accuracies with each client distribution and different rounds of iterations.



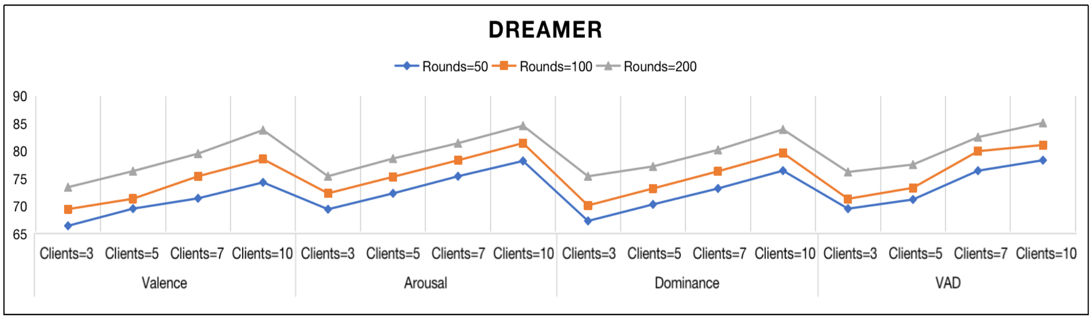


Figure 8.3: Testing accuracies for the proposed AFLEMP with SWFA for Valence, Arousal, Dominance, and 3D-VAD for all rounds of iterations.

2. The proposed AFLEMP with SWFA architecture achieves a testing accuracy of 88.15% (Valence), 89.10% (Arousal), 87.11% (Dominance), and 90.11% (VAD) for AMIGOS [125], and 83.81% (Valence), 84.61% (Arousal), 83.93% (Dominance), and 85.12% (VAD) for DREAMER [135]. It shows that octal classification (for eight emotion classes) gives better results than binary classification (only low/high classes).

3. It presents the training and testing accuracies, showing very little difference in their values, stating that there is no overfitting in the proposed AFLEMP with SWFA architecture.

8.3.1.2 Other Predictive Measures

To evaluate the predictive performance of the proposed multi-modal AFLEMP with SWFA architecture for different iteration rounds, we compute the Accuracy, Precision, Recall, and F1-score. Table 8.7 presents the values of recall, precision, f-1 score, and testing accuracy for the proposed multi-modal AFLEMP with SWFA architecture for 200 rounds of iterations for both datasets AMIGOS [125] and DREAMER [135] with

Table 8.7: Evaluation measures for testing the proposed AFLEMP with SWFA for 200 rounds.

Dataset ↓	Emotion Dimensions↓	Recall	Precision	F-1 Score	Testing Accuracy
AMIGOS [125] (EEG+ECG+GSR)	Arousal	0.847	0.892	0.861	0.898
	Valence	0.815	0.888	0.856	0.871
	Dominance	0.787	0.872	0.845	0.869
	VAD	0.871	0.955	0.913	0.901
DREAMER [135] (EEG+ECG)	Arousal	0.780	0.850	0.813	0.846
	Valence	0.791	0.889	0.837	0.835
	Dominance	0.792	0.835	0.812	0.838
	VAD	0.795	0.893	0.845	0.851

both binary and octal classifications. It shows that evaluation measures other than accuracy, i.e., precision, recall and f1-score is also giving optimal values for both octal and binary classifications for the proposed AFLEMP with SWFA architecture.

8.3.1.3 Communication Computation Measures

Training time (from the client end), averaging time (from the server end), and the model convergence speed via loss curves measures the communication computations of the proposed AFLEMP with SWFA architecture.

1. Table 8.8 presents the training and averaging times (in seconds) for all the iteration rounds and client distributions with both datasets for octal classifications (VAD). It shows that the maximum training time is 266.652 seconds for 200 rounds of iterations for AMIGOS [125] and 245.502 seconds for DREAMER [135]. Similarly, the averaging time presented by Table 8.8 shows that the proposed multi-modal AFLEMP with SWFA takes a maximum averaging time of 107.684 seconds by AMIGOS [125], and 101.467 seconds by DREAMER [135] with 200 rounds of iterations.

2. Table 8.8 presenting the values in seconds, indicating the incremental change in training, and averaging time resulting from variations in data size (C=5, 10, 15 and 20 / C=3, 5, 7 and 10) with the number of rounds (R=50, 100, and 200) proving the proposed multi-modal AFLEMP with SWFA architecture as a scalable architecture.

3. Figure 8.4 illustrates the categorical cross-entropy (for octal classification) loss versus epoch curves for the AMIGOS [125] and DREAMER [135] datasets for

Table 8.8: Training and averaging time in seconds for the proposed AFLMEP with SWFA.

Epochs=200	AMIGOS [125]				DREAMER [233]			
Rounds=50	**Client=5**	**Client=10**	**Client=15**	**Client=20**	**Client=3**	**Client=5**	**Client=7**	**Client=10**
Training Time (in seconds) (Client End)	95.110	98.324	115.571	130.231	90.345	107.205	140.452	120.121
Averaging Time (in seconds) (Server End)	65.212	72.321	80.342	96.212	68.671	76.932	82.302	90.102
Rounds=100								
Training Time (in seconds) (Client End)	117.231	121.764	130.554	153.232	100.561	115.271	130.701	147.678
Averaging Time (in seconds) (Server End)	70.671	75.354	84.342	100.563	70.367	79.172	81.367	96.326
Rounds=200								
Training Time (in seconds) (Client End)	150.221	173.763	200.503	266.652	150.213	173.121	193.201	245.502
Averaging Time (in seconds) (Server End)	74.211	80.431	89.621	107.684	70.123	81.110	89.221	101.467

the proposed AFLEMP with SWFA architecture. It reveals that the train and test sets follow a similar declining path for the eight emotion classes (VAD). The declining loss curves clearly indicate that the loss generated by the proposed AFLEMP with SWFA architecture decreases progressively as the model improves its performance. This reduction in loss corresponds to the increase in accuracy over time, demonstrating that the model is learning and adapting effectively. By the time the training reaches 200 epochs, both the loss and accuracy stabilize, suggesting that the model has converged to an optimal solution.

8.3.2 Performance comparison of proposed multi-modal AFLEMP with SWFA and other FL aggregation algorithms

The proposed multi-modal AFLEMP architecture leverages SWFA at the server side to serve as a global aggregator. Its performance is compared against other federated learning aggregation methods. Table 8.9 presents the testing classification accuracy and the time taken by each FL algorithm for server-side aggregation. The proposed multi-modal AFLEMP with SWFA outperforms the other FL algorithms, achieving testing accuracy of 90.11% and 85.12% with AMIGOS [125] and DREAMER [135],

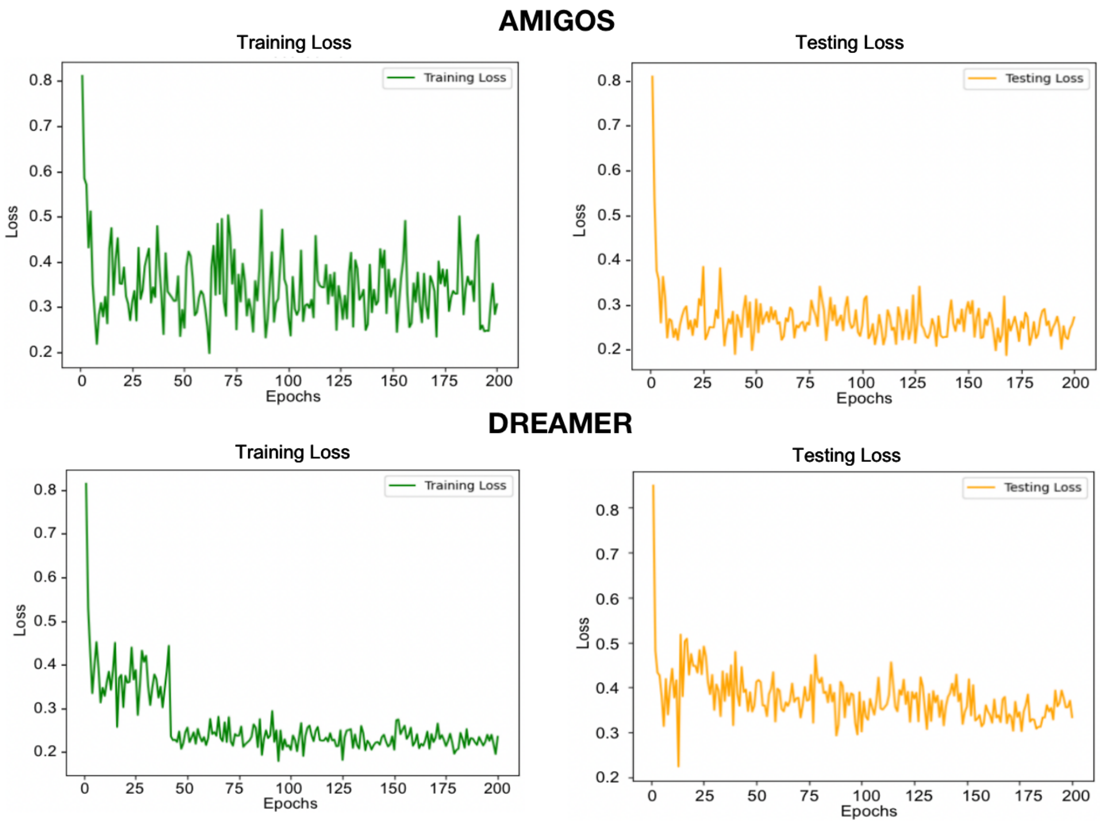


Figure 8.4: Loss curves for the proposed AFLEMP with SWFA for 3D-VAD.

Table 8.9: Testing accuracies and averaging time for AFLEMP with SWFA and other FL algorithms.

Clients=20, Rounds=200	AMIGOS [125]		DREAMER [135]	
	Testing Accuracy	**Averaging Time (in seconds)**	**Testing Accuracy**	**Averaging Time (in seconds)**
Fedavg	83.10%	100.537	77.31%	99.367
Fedper	84.12%	180.551	78.98%	190.545
FedBoost	86.67%	195.343	80.45%	201.432
DWFA	81.52%	201.713	74.92%	196.125
SWFA	**90.11%**	**107.684**	**85.12%**	**101.467**

respectively for octal classification. The SWFA takes 107.684 seconds with the AMIGOS dataset and 101.467 seconds with the DREAMER dataset for performing averaging at the server end with the maximum number of iteration rounds and clients. The time taken by SWFA is shorter than all other FL algorithms except FedAvg while achieving improved classification accuracy.

Table 8.10: Performance comparison with existing FL works for emotion recognition with physiological signals.

FL for Emotion Recognition									
Reference	**Dataset**	**Physiological Signal**	**Classification Model**	**Modality**	**FL Methods**	**ED[1]**	**Testing Accuracy**	**S[3]**	**CC[2]**
Nandi et al. [36]	DEAP [233]	EDA, RESP	FFNN	Bi-Modal	FedAvg	2D	81.92% (VA)	✓	×
Tara Hassani [100]	CASE [134]	GSR	CNN	Single	FedAvg	2D	79% (VA)	×	×
Agrawal et al. [149]	DEAP [233]	EEG	2D-CNN	Single	FedAvg	3D	69.77% (Avg. V,A,D)	×	×
Proposed F-MERS	AMIGOS [125] DREAMER [135]	EEG+ECG+GSR EEG+ECG	MLP	Multi-Modal	FedAvg	3D	83.02% (Avg. V,A,D) 75.19% (Avg. V,A,D)	✓	✓
Proposed AFLEMP	AMIGOS [125] DREAMER [135]	EEG+ECG+GSR EEG+ECG	ANN+ Transformer	Multi-Modal	FedAvg	3D	83.04% (Avg. V,A,D) 77.34% (Avg. V,A,D)	✓	✓
Proposed AFLEMP	**AMIGOS [125] DREAMER [135]**	**EEG+ECG+GSR EEG+ECG**	**ANN+ Transformer**	**Multi-Modal**	**SWFA**	**3D**	**90.11% (VAD) 85.12% (VAD)**	✓	✓

[1]Emotion Dimension (ED), [2]Communication Computation (CC), [3]Scalability (S)

8.3.3 Performance comparison of proposed multi-modal AFLEMP with SWFA with existing FL works

Table 8.10 presents the comparative analysis between the proposed privacy-preserving federated architecture and other existing methods discussed in the literature utilizing FL for emotion recognition. It demonstrates that the proposed AFLEMP with SWFA architecture outperforms the previous FL works as well as the proposed FL architectures (F-MERS and AFLEMP) for emotion recognition, achieving an accuracy of 90.11% and 85.12% for AMIGOS [125] and DREAMER [135], respectively. Additionally, the proposed AFLEMP with SWFA architecture also computes the scalability and communication computation, which is missing in the previous studies of FL for emotion recognition.

8.4 Summary

The proposed multi-modal AFLEMP architecture with the novel Scaled Weighted Federated Averaging (SWFA) algorithm can accurately and effectively classify broader human emotions (using VAD) in a decentralized, federated learning environment. SWFA is embedded (at the server end) in the proposed multi-modal AFLEMP architecture, generating a global aggregator at the server. It averages the updates from the local model running at the client end. It enhances the aggregation process by scaling the weights received from clients with large data, thereby reducing the Imbalanced Data Heterogeneity (IDH) occurring at the client end. The result outcomes of the proposed SWFA algorithm evidenced the reduction of data heterogeneity by improved classification accuracy (for both binary and octal) and reduced communication computations, making the AFLEMP architecture robust. It improves the fairness and effectiveness of the aggregation process, leading to a more robust global aggregator in federated learning environments.

The proposed AFLMEP with SWFA architecture integrates a Transformer with an ANN model to train at the client's end and SWFA at the server end. The Transformer calculates the weights for essential features from multi-modal physiological data at each client end and sends the computed weight gradients to the global server. Embedding the attention mechanism in the proposed federated architecture reduces the amount of data transfer from the client to the server in a federated environment, reducing the amount of data transfer, thereby reducing the Variation Data Heterogeneity (VDH) in multi-modal physiological data (at the client end). Embedding SWFA at the server end in the proposed federated architecture helps reduce the IDH by performing scaling and weighted averaging at the global server. The result outcomes of the proposed AFLEMP with SWFA architecture evidenced the reduction of data heterogeneities from the client and server end through improved communication, making the architecture robust.

This chapter is based on the following work:

- **J2: Neha Gahlan**, and Divyashikha Sethia. "AFLEMP: Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data." **Biomedical Signal Processing and Control** 94 (2024): 106353. (**SCIE, Impact Factor: 5.1, Publisher: Elsevier**). Doi: https://doi.org/10.1016/j.bspc.2024.106353.

Chapter 9

Conclusion, Future Work & Social Impact

This chapter summarizes the key findings and contributions of the research, reflecting on the overall objectives and how they have been addressed throughout the work. It delves into the implications of these findings, not only in terms of academic progress but also in terms of their practical applications in real-world settings. The chapter then outlines potential future work, focusing on areas where further exploration could enhance current methodologies or open up new research avenues. Finally, the societal impact of this research is discussed, emphasizing how advances in emotion recognition through physiological signals can contribute to various sectors such as healthcare, education, and human-computer interaction. The broader ethical considerations regarding physiological signals are also highlighted to ensure that future developments align with societal needs and values.

9.1 Conclusion & Future Work

This thesis establishes the theoretical basis and fundamental concepts needed to grasp the idea of emotion recognition and its relationship with physiological data. It evaluates emotions, various physiological signals used in emotion recognition, benchmark datasets, and methodologies for classifying emotional states. Majorly, it addresses data privacy concerns by introducing an innovative Federated Learning (FL) paradigm for emotion recognition. FL offers a promising approach for utilizing large and diverse datasets to train machine learning models while preserving data privacy, a crucial factor in safeguarding sensitive, emotional, mental healthcare, and physiological information.

In conclusion, this thesis is based on four key objectives: (1). Preserves **privacy** of ERS based on physiological signals, (2). Provides **multi-modality** with physiological signals in ERS, (3). Addresses **data heterogeneity (variation and imbalanced)** in federated ERS, (4). **Expands emotions** of ERS via 3D-VAD emotion dimensions.

Firstly, the thesis conducted a comprehensive literature review of the automated ERS using physiological signals to find the research gaps. It provides a literature review of the emotion models, physiological signals, the relation between emotions and physiological signals, technical background including data processing for physiological signals, ML and DL models, and the related works of FL for ERS. It is covered in Chapters 2, 3, 4.

Secondly, the thesis proposes a privacy-preserved FL-based Emotion recognition architecture using multi-modal physiological data (**F-MERS**) for classifying emotion states using a three-dimensional model of emotions covering a wide range of similar and complex emotions. The Proposed F-MERS validates three emotion benchmark datasets. The experimental results show that muli-modal performs better than single modalities, proving the proposed F-MERS to be robust, scalable (in terms of client distribution and number of iterations and rounds) and efficient in communication (in terms of time). The proposed F-MERS could not address the data heterogeneities of the multi-modal physiological data and the federated environment. It is covered in Chapter5.

Lastly, the thesis addresses two different types of data heterogeneities. One is the Variation Data Heterogeneity (VDH) existing in the multi-modal physiological data, and the other is Imbalanced Data Heterogeneity (IDH) occurring at the server end within a federated environment. To reduce these, the thesis proposed an Attention-based (**AFLEMP**) architecture to remove the VDH in the multi-modal physiological data. It incorporates a Scaled-Weighted Federated Averaging (SWFA) algorithm to reduce the IDH occurring at the server end within a federated environment. The Proposed AFLEMP validates two emotion benchmark datasets. The experimental results show improvement in classification accuracies and evaluation metrics, indicating the removal of VDH. Also, the communication efficiency in terms of time is compared for the proposed SWFA with other existing FL methods, resulting in better performance than other FL methods. It is covered in Chapters 5, 6, 7.

9.1.1 Lessons Learnt

Based on the literature review, experiments performed, and findings, the following are the learning points:

1. **Why is privacy essential for physiological data?** Physiological data is inherently personal and highly sensitive, encompassing signals such as EEG, ECG, RESP, GSR, Heart Rate, and others. Data from these signals are unique to each individual, revealing a lot about their health status and potentially providing insights into their emotional state, behaviour, and habits. Hence, they are considered sensitive. Compromising the privacy of physiological data can have profound repercussions on an individual's life. It opens the door for data attackers and exposes them to the risk of data breaches, which, in turn, can lead to various threats [21]. These threats include the potential exposure of an individual's health status, emotional stability, and even their biometric identity based on physiological signals. Therefore, safeguarding privacy is crucial when it comes to handling physiological data. For this, the proposed study ensures that physiological sensor data for emotion recognition is used to protect people's privacy and autonomy.

2. **How can complex emotions be mapped into different dimensions?** Complex emotions like fear, wrath, guilt, resentment, anxiety, and others have different levels of arousal, valence and dominance, are difficult to distinguish and cannot be mapped on the 2-dimensional emotion model by Russel [204]. Hence, the proposed study adopted Mehrabian's 3-dimensional model of emotions to map these complex emotions [33, 128]. The 3-dimensional model of emotion maps complex emotions into three dimensions: Arousal, Valence and Dominance. It understands better how emotions are experienced and influence one's behaviour. The 3-dimensional model of emotions maps emotions as arousal refers to the physiological activation or intensity level of emotions, ranging from low arousal (calm, relaxed) to high arousal (excited, anxious). Valence represents the pleasantness or unpleasantness of emotions, ranging from positive (happiness, joy) to negative (anger, sadness). Dominance reflects an emotion's sense of control or power, ranging from feeling dominant (empowered, in control) to

feeling submissive (helpless, powerless). This information develops better ways to cope with complex emotions and build stronger relationships.

3. **How do physiological signals contribute to emotion recognition?** Physiological signals provide valuable information about the body's physiological responses and indicate different emotional states.

    - EEG signals record the electrical hum of the brain, revealing its activity, as the brain acts as the maestro of emotions, conducting everything from physical actions and sensory experiences to language, memories, and feelings [175, 278].

    - GSR measures the skin's electrical conductivity and is also known as Electrodermal Activity (EDA) [156, 181, 205]. Skin conductivity varies with skin moisture level (sweating), showing variations in the Autonomous Nerve System associated with arousal, reflecting emotions such as stress, anxiety, and surprise. It is, in particular, a measure of arousal.

    - ECG signals are electric signals acquired to trace the action of the human heart and the potential fluctuations transmitted to the skin surface due to the heart's electrical activity (the contraction and relaxation of heart muscles) [113, 245]. Electrodes linked to the skin surface detect it.

    - The respiratory rate (RESP) physiological signal represents the frequency of breaths a person takes per minute, mirroring the inhaling and exhaling air rate. These breathing patterns closely connect to the emotional states [233]. When experiencing emotions like stress, fear, anxiety, or excitement, breathing quickens and becomes shallower. In contrast, during moments of relaxation and calmness, our breathing becomes slower and deeper. This interplay between respiratory rates highlights the importance of conscious breathing techniques in managing emotions [106]

4. **Why is multi-modality required in emotion recognition?** Here, multi-modality refers to the fusion or combination of different physiological signals required for emotion recognition. These different physiological signals provide

complementary information about emotions. Multiple physiological signals capture a more comprehensive picture of emotional states and increase emotion recognition accuracy [53]. Emotion recognition systems that rely on a single physiological signal are vulnerable to noise, artefacts, or biases inherent in that particular modality (physiological signal). Combining multiple modalities (physiological signals) creates more robust and adaptable systems capable of recognizing different emotions more accurately and precisely. In response, the proposed study combines the participating subjects' EEG, ECG, GSR and RESP signals.

5. **How can machine learning be used for automated emotion recognition systems?** Physiological signals data (EEG, ECG, GSR, RESP) obtained from different wearable sensors for recognizing emotions is preprocessed, from which the relevant features are extracted. The relevant features extracted are the inputs to train the Machine Learning and Deep Learning algorithms like SVM [252], DT [111], and KNN [85, 298], RNN [213], CNN [157, 173], and LSTM [98] for classifying the different emotion states. These algorithms are successful in attaining higher accuracies with physiological sensor data. It is worth noting that the success of these automated machine learning-based emotion recognition systems depends on the quality and diversity of the training data, the choice of relevant features, and the selection and optimization of the machine learning algorithm. The proposed study accommodates this by implementing an MLP classifier for emotion state classification.

6. **How federated learning paradigm is preserving data privacy in emotion recognition?** The traditional ML and DL architectures require complete access to the physiological data for training the model in an automated emotion recognition system. It compromises the privacy of the data as it requires complete access to physiological data for training purposes, giving easy access to data attackers. A new paradigm called Federated Learning (FL) is introduced by McMahan et al. [51] to resolve the issue of data privacy. FL is a promising approach that creates a decentralized environment with a local and global model at the client and server end, respectively [69]. It allows the local model updates to be sent

to a central server, combining them to create a global model [39, 99, 124]. This approach does not allow the global model to access the raw data used for training and hence preserves the privacy of the sensitive physiological data. The proposed study accommodates this approach by proposing the F-MERS architecture for emotion recognition, preserving the privacy of the sensitive physiological data while achieving good accuracy results.

9.1.2 Ethical Implications and Responsible Use

The ethical implications surrounding the automated ERS for emotion classification framework are significant and require thorough consideration. Automated ERS offers intriguing possibilities but raises several ethical concerns, mainly when applied to sensitive physiological data that require scholarly discussion and cautious implementation. Here is a breakdown of such critical issues:

9.1.2.1 Privacy Concerns

When utilizing technology for automated emotion recognition and classification using sensitive physiological data, a critical ethical consideration is required to safeguard individuals' privacy rights. These technologies involve collecting sensitive physiological data and then using it for automation via machine learning models. This process unintentionally violates an individual's right to privacy. Intrusive collection and analysis of emotions raise concerns about individual privacy and autonomy. Adhering to ethical standards and principles, such as obtaining informed consent and ensuring participant privacy, is crucial, as mentioned below:

1. *While collecting data, Informed Consent (Transparency):* Individuals are provided clear and concise information on how their data is collected, used, and stored.

2. *Granularity:* Consents are specific to the intended use of the data. Collecting data for research doesn't automatically allow its use for commercial purposes.

3. *Opt-out vs Opt-in:* Opt-in consent ensures individuals actively choose to participate, minimizing coercion.

4. *Voluntary Participation:* Participants must willingly agree to contribute their physiological data without coercion. They should have the option to withdraw at any point without facing negative consequences.

9.1.2.2 Potential Biases

The following biases need to be taken care of when using automated Emotion Recognition Systems (ERS) and collecting sensitive physiological data:

1. *Demographic and Cultural Biases:* These biases raise ethical concerns if the ERS favours certain demographic factors such as age, gender, and ethnicity over others, potentially leading to unequal representation. If the training data predominantly represents a specific group or culture, the system might struggle to accurately recognize and interpret emotions in individuals from other demographics.

2. *Algorithmic Biases:* The algorithms powering ERS may inadvertently perpetuate existing societal biases in the training data. For instance, if historical data contains imbalances or stereotypes, the system might unintentionally reinforce these biases, leading to unfair or inaccurate results.

3. *User Interface Bias:* The potential biases introduced through the user interface of the ERS might be biased toward users with different abilities. They should avoid discrimination based on interface preferences or limitations and be easily accessible.

The ethical implications of using automated ERS and collecting sensitive physiological data involve addressing potential biases, obtaining informed consent through transparency and voluntary participation, and ensuring responsible use by prioritizing data security, privacy, and fair deployment practices. Ethical considerations are paramount in shaping the development and deployment of emotion recognition technology to foster trust, fairness, and positive societal impact.

9.1.2.3 Responsible Uses

The proposed automated Emotion Recognition System (ERS) can be utilized in the following responsible ways:

1. *Data Security and Privacy:* The automated ERS must ensure robust measures to protect the collected physiological data. They secure storage and protect against unauthorized access. Emphasize the importance of respecting individuals' privacy rights throughout the data lifecycle.

2. *Mitigating Harm:* The automated ERS must implement safeguards to prevent potential harm from misusing emotion recognition technology. It includes addressing issues like emotional profiling, stigmatization, or the unintended consequences of relying on ERS in sensitive data.

3. *Fair and Equitable Deployment:* The automated ERS must strive for fair and equitable deployment of ERS across diverse populations. Avoid scenarios where certain groups may face disproportionate consequences or disadvantages due to the technology's application.

9.1.2.4 Limitations

One of the key challenges in applying the proposed F-MERS and AFLEMP to real-world scenarios is handling noisy and incomplete physiological signal datasets. Unlike controlled laboratory settings, real-world physiological signals often suffer from motion artefacts, environmental interference, and sensor inconsistencies. EEG signals may be affected by muscle movements and electrode displacement, ECG signals may exhibit baseline drift and powerline interference, while temperature fluctuations and skin conditions can impact GSR readings. These artefacts introduce variability in the extracted features, leading to performance degradation. To address this issue, different filtering methods can be applied to the signal data as a preprocessing step, like low and high-frequency filters, Butterworth filters, and bandpass filters. Additionally, Empirical Mode Decomposition (EMD) and Independent Component Analysis (ICA), as discussed above in section 2.5.1.2, can be applied to the EEG signal for artefact removal and denoising.

Another significant issue in real-world applications is incomplete datasets. Physiological signals are prone to missing values due to sensor failures, poor signal quality, or user non-compliance. In federated architectures, missing data can vary across clients, leading to biased updates and inefficient model training. To address this, future research should investigate federated data imputation techniques like FedMIWAE [105], FedTMI [297], Fedimpute [144] to handle missing data dynamically while preserving user privacy.

9.1.3 Future Work

There are certain restrictions on the proposed study, which opens the door for more investigation and testing. These are: (1). Since the current work involves feature-level fusion, one should experiment with decision-level fusion of several modalities. 2) Developing a more comprehensive method for emotion identification by combining physiological markers with additional bodily expressions of emotion, such as eye tracking, speech, and gesture. Since the current study only considers physiological signs, the proposed FL framework can be expanded in the future to integrate physical and physiological indicators.

9.2 Social Impact & Applications

This section outlines different applications for emotion recognition with federated learning that can improve real-world situations. The FL paradigm makes a wide range of applications possible by eliminating the need for data sharing while creating a machine-learning model. These are described as below:

1. **Smart Emotional Healthcare:** The automated ERSs are pivotal in discerning an individual's emotional states, which indicate their overall health. These systems leverage physiological signals such as heart rate, blood pressure, and skin temperature collected via smart wearables like smart bands and smartphones, forming an integral part of innovative healthcare systems. The proposed FL-based ERS mitigate the risk of information leakage and privacy attacks, ensuring highly protected data privacy in the healthcare industry by embedding them in the edge devices at the user's end (as shown in Figure 9.1). This approach significantly reduces vulnerabilities towards sensitive physiological

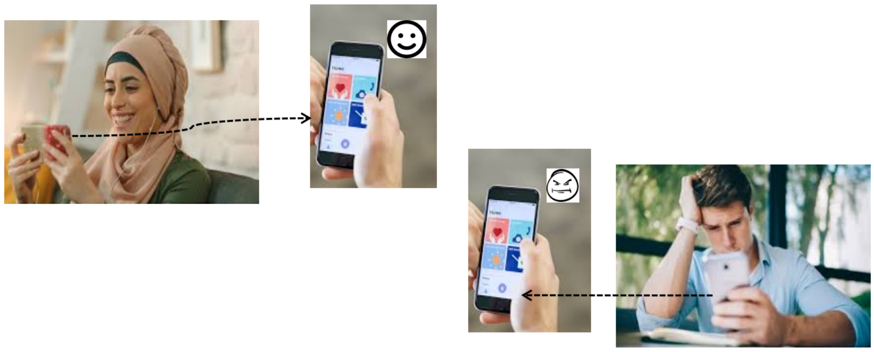


Figure 9.1: Health monitoring using smart wearable sensors

data associated with traditional centralized ERS, offering a safer alternative for recognizing emotions and health status via smart wearables.

Figure 9.2: In-Home health monitoring using smart wearable sensors

2. **In-Home Mental well-being Monitoring:**In-home health monitoring has garnered widespread interest among the global ageing population. In these scenarios, wearable IoT edge devices like smartphones and smartwatches play a crucial role in data acquisition. These devices, worn at home, track daily activities such as emotions, reactions, heart rate, and pulse rate (as shown in Figure 9.2). IoT devices efficiently collect vast amounts of data, which can be processed further. However, traditional healthcare centres utilize this data for diagnosis, applying ML and DL methods to classify emotional states while allowing third-party attackers to access sensitive data. To overcome these issues, the proposed robust FL-based ERS can be integrated with the IoT wearable devices to help report complex emotions recognized, enhance data privacy and improve communication, accuracy and training performance.

3. **E-Learning:** In E-Learning, FL enabled emotion recognition systems leverage wearable IoT devices to monitor physiological signals like heart rate and skin temperature, providing real-time insights into students' emotional states (as

Figure 9.3: Emotional health monitoring while E-learning.

shown in Figure 9.3). Unlike traditional methods, FL trains models directly on devices, ensuring data privacy by keeping sensitive information local. This approach personalizes learning experiences, offers real-time feedback to educators, enhances student support, and improves learning outcomes by adapting to students' emotional needs. FL's decentralized nature ensures robust data privacy and security, making it a transformative and trustworthy solution for modern E-Learning environments. For example, Gu et al. [248] proposed a privacy-preserved teacher-student network. The author used EEG signals in this network to recognize student emotions in an academic environment.

4. **Corporate Workers:** Employees in the corporate sector generally have a demanding workload every day, which causes a lot of stress. This ongoing stress can lead to an unhealthy lifestyle and an unwillingness to work, forcing individuals to quit their occupations or negatively affecting their health. Employees under such pressure could lose awareness of their emotional states. Helping people identify their current emotions and mental health is essential, enabling them to take breaks when necessary. Their complex and multiple emotions may be detected in real-time with data privacy using the proposed robust FL architecture, which can help them become more conscious and create a less stressful work atmosphere (as shown in Figure 9.4). Employees can work with greater confidence and integrity and less stress, which enhances their general well-being and productivity.

5. **Pet Robots:** These days, robots are being used increasingly in various settings, including retail, restaurants, hotels, airports, hospitals, and more [241]. These machines work using human-machine communication. A new study [186] claims

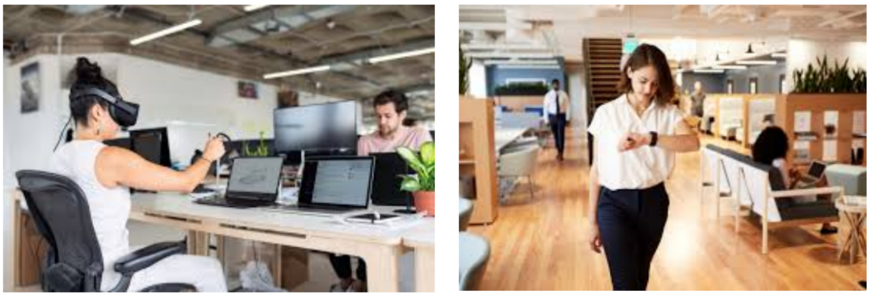


Figure 9.4: Emotional health monitoring in office corporate environments.

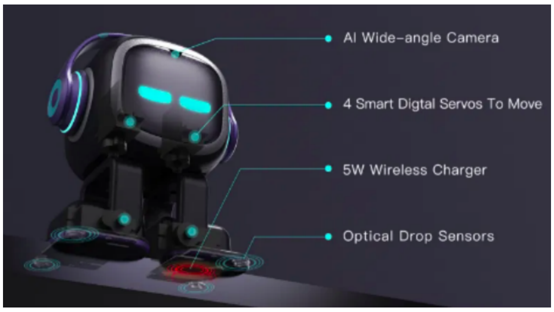


Figure 9.5: Pet Emo robot with interactive sensors

that the most acceptable robots are humanoid robots, which resemble humans in appearance. When these robots are in varied surroundings, they use smart devices and facial expressions to track human emotions. Once tracked, they can help people keep their emotions in check. For instance, the autonomous and intelligent EMO robot[1] (shown in Figure 9.5) developed by Emospark has an integrated neural network processor and optical drop sensors. It functions as a devoted companion and pet robot. Applications for iOS and Android can be connected to it with ease. Embedding the proposed FL architecture in such robots can enhance the data integrity, privacy while recognizing emotions and requests user experiences using slider buttons. The arousal, dominance, and valence emotions are the main focuses of the robotic paradigm.

6. **Designing Emotional UI:** These days, individuals can recognize emotions and discern moods thanks to user interfaces on smart TVs, smartphones, and wearables.

   "Technology is best when people are happy."[2].

   For example, Spikey Sanju's Employee Engagement app uses emojis on

[1] https://living.ai/emo/

[2] https://uxplanet.org/designing-emotional-ui-b11fa0fda5c

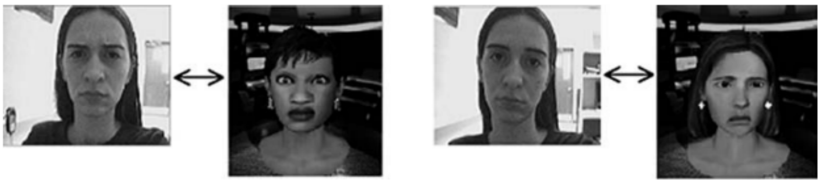


Figure 9.6: Animoji: Augmented reality masks to express emotions.

employees' displays to measure their emotions, such as joyful, furious, or sad [2]. Many UI designers use cameras to create cutting-edge means for users to express their feelings, such as Animoji[3] or augmented reality masks (an example is shown in Figure 9.6). These features allow people to convey emotions, significantly improving the user experience. By incorporating data privacy into these UI features through the proposed FL-based ERS, these devices can be safeguarded, enhanced, and less susceptible to data breaches while recognizing emotions.

[3] https://dribbble.com/shots/9524341-Employee-Engagement-App-Mood-Tracker-UI

References

[1] Flower A Friendly Federated Learning Framework. https://flower.dev/, Last accessed on 11 August, 2023.

[2] Mood Tracking Application. https://shorturl.at/5brHd, Last accessed on August 15, 2022.

[3] Biosemi ActiveTwo System. https://www.biosemi.com/, Last accessed on August 7, 2022.

[4] Emotiv Epoc Neuroheadset. https://www.emotiv.com/, Last accessed on August 7, 2022.

[5] EMOTIVE EPOC+. https://www.emotiv.com/epoc/, Last accessed on August 7, 2022.

[6] EMPATIC E4. https://www.empatica.com/en-gb/, Last accessed on August 7, 2022.

[7] ESI NeuroScan System. https://compumedicsneuroscan.com/applications/eeg/, Last accessed on August 7, 2022.

[8] Garmin HRM-DUAL. https://www.garmin.com/en-US/p/649059, Last accessed on August 7, 2022.

[9] Neurosky Mindwave. https://store.neurosky.com/pages/mindwave, Last accessed on August 7, 2022.

[10] Nexus-32 Amplifier. https://www.mindmedia.com/en/products/nexus-32/, Last accessed on August 7, 2022.

[11] Polar H10. https://shorturl.at/uBicK, Last accessed on August 7, 2022.

[12] RespiBAN. https://www.pluxbiosignals.com/products/respiban-ble, Last accessed on August 7, 2022.

[13] RING. https://shorturl.at/m6sb5, Last accessed on August 7, 2022.

[14] SHIMMER 2R. https://shimmersensing.com/, Last accessed on August 7, 2022.

[15] SHIMMER3. https://imotions.com/hardware/shimmer3-gsr/, Last accessed on August 7, 2022.

[16] VITALCONNECT. https://vitalconnect.com/, Last accessed on August 7, 2022.

[17] CGX Quick Systems. https://www.brainproducts.com/solutions/cgx-quick/, Last accessed on August 7, 2024.

[18] DSI-24. https://wearablesensing.com/dsi-24/, Last accessed on August 7, 2024.

[19] Neuphony Headband. https://neuphony.com/product/brain-wearables/neuphony-eeg-headband/, Last accessed on August 7, 2024.

[20] Samsung Gear live. https://www.samsung.com/us/support/mobile/wearables/smartwatches/gear-live, Last accessed on August 8, 2022.

[21] Data Breach. https://tinyurl.com/2p8b57ax, Last accessed on February 3, 2023.

[22] Federated AI Technology Enabler (FATE). https://fate.fedai.org/, Last accessed on February 3, 2023.

[23] Model Inversion Attacks and Data Protection Law. https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0083, Last accessed on June 10, 2023.

[24] AI Security Model Hacking with Model Inversion Attacks. https://tinyurl.com/cspy6wj6, Last accessed on June 10, 2024.

[25] Data Breach Exposes. https://tinyurl.com/mr2r262n, Last accessed on May 5, 2023.

[26] A. Anuragi, D.S. Sisodia and R.B. Pachori. "EEG-Based Cross-Subject Emotion Recognition using Fourier-Bessel Series Expansion based Empirical Wavelet Transform and NCA Feature Selection Method". *Information Sciences*, 610:508–524, 2022.

[27] A. Bhattacharyya, et al. "A Novel Multivariate-Multiscale Approach for Computing EEG Spectral and Temporal Complexity for Human Emotion Recognition". *IEEE Sensors Journal*, 21(3):3579–3591, 2020.

[28] A. Bhatti, et al. "AttX: Attentive Cross-Connections for Fusion of Wearable Signals in Emotion Recognition". *arXiv preprint arXiv:2206.04625*, 2022.

[29] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas. "Human Emotion Recognition: Review of Sensors and Methods". *Sensors*, 20(3):592, 2020.

[30] A. Fallatah and H.R. Dajani. "Accurate Detection of Speech Auditory Brainstem Responses Using a Spectral Feature-based ANN Method". *Biomedical Signal Processing and Control*, 44:307–313, 2018.

[31] A. Iyer, et al. "CNN and LSTM Based Ensemble Learning for Human Emotion Recognition using EEG Recordings". *Multimedia Tools and Applications*, 82(4):4883–4896, 2023.

[32] A. Lalitha, et al. "Peer-to-peer Federated Learning on Graphs". *arXiv preprint arXiv:1901.11173*, 2019.

[33] A. Mehrabian and J.A. Russell. *"An Approach to Environmental Psychology."* the MIT Press, 1974.

[34] A. Mert and A. Akan. "Emotion Recognition from EEG Signals by Using Multivariate Empirical Mode Decomposition". *Pattern Analysis and Applications*, 21:81–89, 2018.

[35] A. Metallinou and S. Narayanan. "Annotation and Processing of Continuous Emotional Attributes: Challenges and Opportunities". In *Proc. IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2013.

[36] A. Nandi and F. Xhafa. "A Federated Learning Method for Real-Time Emotion State Classification from Multi-Modal Streaming". *Methods*, 2022.

[37] A. Ortony, G.L. Clore and A. Collins. "The Cognitive Structure of Emotions Cambridge". *UK: Cambridge University Press9*, 1988.

[38] A. Quattrone. *"Inferring Sensitive Information from Seemingly Innocuous Smartphone Data."* PhD thesis, University of Melbourne, Australia, 2016.

[39] A. Rahman, et al. “Federated Learning-based AI Approaches in Smart Healthcare: Concepts, Taxonomies, Challenges and Open Issues”. *Cluster Computing*, pages 1–41, 2022.

[40] A. Topic and M. Russo. “Emotion Recognition Based on EEG Feature Maps Through Deep Learning Network”. *Engineering Science and Technology, an International Journal*, 24(6):1442–1454, 2021.

[41] A. Vaswani, et al. “Attention is All You Need”. *Advances in Neural Information Processing Systems*, 30, 2017.

[42] A.D. Bigirimana and N. Siddique and D. Coyle. “A Hybrid ICA-Wavelet Transform for Automated Artefact Removal in EEG-Based Emotion Recognition”. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 004429–004434, 2016.

[43] A.G. Roy, at. al. “Braintorrent: A Peer-to-peer Environment for Decentralized Federated learning”. *arXiv preprint arXiv:1905.06731*, 2019.

[44] A.K. Eesee. “The Suitability of the Galvanic Skin Response (GSR) as a Measure of Emotions and the Possibility of Using the Scapula as an Alternative Recording Site of GSR”. In *Proc. 2nd IEEE International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE)*, pages 80–84, 2019.

[45] A.Nandi, F. Xhafa and R. Kumar. “A Docker-based federated learning framework design and deployment for multi-modal data stream classification”. *Computing*, pages 1–35, 2023.

[46] A.S. Rajpoot, et al. “Subject Independent Emotion Recognition using EEG Signals Smploying Attention Driven Neural Networks”. *Biomedical Signal Processing and Control*, 75:103547, 2022.

[47] A.Zadoo, et al. “A Review on Churn Prediction and Customer Segmentation Using Machine Learning”. In *Proc. IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, volume 1, pages 174–178, 2022.

[48] B. Chakravarthi, et al. “EEG-based Emotion Recognition Using Hybrid CNN and LSTM Classification”. *Frontiers in Computational Neuroscience*, 16:1019776, 2022.

[49] B. Gu, et al. "Privacy-Preserving Asynchronous Vertical Federated Learning Algorithms for Multiparty Collaborative Learning". *IEEE Transactions on Neural Networks and Learning Systems*, 33(11):6103–6115, 2021.

[50] B. Khaleghi, et al. "Multisensor Data Fusion: A Review of the State-of-the-art". *Information fusion*, 14(1):28–44, 2013.

[51] B. McMahan, et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data". In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[52] B. Zali-Vargahan, et al. "Deep Time-Frequency Features and Semi-Supervised Dimension Reduction for Subject-Independent Emotion Recognition from Multi-Channel EEG Signals". *Biomedical Signal Processing and Control*, 85:104806, 2023.

[53] C. Busso, et al. "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information". In *Proc. ACM 6th International Conference on Multimodal Interfaces*, pages 205–211, 2004.

[54] C. Kirschbaum, K. Pirke and .H. Hellhammer. "The 'Trier Social Stress Test'–A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting". *Neuropsychobiology*, 28(1-2):76–81, 1993.

[55] C. Li, et al. "Emotion Recognition of Human Physiological Signals Based on Recursive Quantitative Analysis". In *Proc. IEEE 10th International Conference on Advanced Computational Intelligence (ICACI)*, pages 217–223, 2018.

[56] C. Lutz and G.M. White. "The Anthropology of Emotions". *Annual Review of Anthropology*, pages 405–436, 1986.

[57] C. Peter, E. Ebert and H. Beikirch. "A Wearable Multi-Sensor System for Mobile Acquisition of Emotion-Related Physiological Data". In *Proc. International Conference on Affective Computing and Intelligent Interaction*, pages 691–698. Springer, 2005.

[58] C. Schiweck, et al. "Heart Rate and High Frequency Heart Rate Variability During Stress as Biomarker for Clinical Depression. A Systematic Review". *Psychological Medicine*, 49(2):200–211, 2019.

[59] C. Tan, M. Šarlija and N. Kasabov . "NeuroSense: Short-Term Emotion Recognition and Understanding Based on Spiking Neural Network Modelling of Spatio-Temporal EEG Patterns". *Neurocomputing*, 434:137–148, 2021.

[60] C. Vidaurre, T.H. Sander and A. Schlögl. “BioSig: The Free and Open Source Software Library for Biomedical Signal Processing”. *Computational Intelligence and Neuroscience*, 2011(1):935364, 2011.

[61] C. Xiefeng, et al. “Heart Sound Signals can be used for Emotion Recognition”. *Scientific Reports*, 9(1):6486, 2019.

[62] C. Ye and B.V. Kumar and M.T. Coimbra. “Heartbeat Classification using Morphological and Dynamic Features of ECG Signals”. *IEEE Transactions on Biomedical Engineering*, 59(10):2930–2941, 2012.

[63] C.D. Spielberger, et al. “The State-Trait Anxiety Inventory”. *Revista Interamericana de Psicologia/Interamerican Journal of Psychology*, 5(3 & 4), 1971.

[64] C.E Izard. “Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues”. *Annual Review of Psychology*, 60:1–25, 2009.

[65] M.T. Cicero. *“Cicero on The Emotions: Tusculan disputations 3 and 4”*. University of Chicago Press, 2002.

[66] C.Y. Park, et al. “K-EmoCon, A Multimodal Sensor Dataset for Continuous Emotion Recognition in Naturalistic Conversations”. *Scientific Data*, 7(1):1–16, 2020.

[67] D. Gao, et al. “HHHFL: Hierarchical Heterogeneous Horizontal Federated Learning for Electroencephalography”. *arXiv preprint arXiv:1909.05784*, 2019.

[68] D. Li, et al. “Spatial-Frequency Convolutional Self-Attention Network for EEG Emotion Recognition”. *Applied Soft Computing*, 122:108740, 2022.

[69] D. Li, Z. Luo and B. Cao. “Blockchain-based Federated Learning Methodologies in Smart Environments”. *Cluster Computing*, 25(4):2585–2599, 2022.

[70] D. Nath, et al. “An Efficient Approach to EEG-Based Emotion Recognition using LSTM Network”. In *2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA)*, pages 88–92. IEEE, 2020.

[71] D. Watson and L.A. Clark and A. Tellegen. “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales.” *Journal of Personality and Social Psychology*, 54(6):1063, 1988.

[72] Charles Darwin. “In F. Darwin,(Ed.), The Life and Letters of Charles Darwin: Including an Autobiographical Chapter (Cambridge Library Collection-Darwin, Evolution and Genetics)”, 2009.

[73] D.O. Bos, et al. “EEG-Based Emotion Recognition”. *The Influence of Visual and Auditory Stimuli*, 56(3):1–17, 2006.

[74] D.P. Tobón Vallejo and A. El Saddik. “Emotional States Detection Approaches Based on Physiological Signals for Healthcare Applications: A Review”. *Connected Health in Smart Cities*, pages 47–74, 2020.

[75] E. Alickovic and Z. Babic. “The Effect of Denoising on Classification of ECG Signals”. In *Proc. IEEE 25th International Conference on Information, Communication and Automation Technologies (ICAT)*, pages 1–6, 2015.

[76] E. Horvitz and D. Mulligan. “Data, Privacy, and the Greater Good”. *Science*, 349(6245):253–255, 2015.

[77] E. Younis, et al. “Evaluating Ensemble Learning Methods for Multi-modal Emotion Recognition Using Sensor Data Fusion”. *Sensors*, 22(15):5611, 2022.

[78] E.A. Veltmeijer, C. Gerritsen and K.V. Hindriks. “Automatic Emotion Recognition for Groups: A Review”. *IEEE Transactions on Affective Computing*, 14(1):89–107, 2021.

[79] E.K Krumova, et al. “Long-Term Skin Temperature Measurements–A Practical Diagnostic Tool in Complex Regional Pain Syndrome”. *Pain*, 140(1):8–22, 2008.

[80] Electrophysiology, Task Force of the European Society of Cardiology the North American Society of Pacing. “Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use”. *Circulation*, 93(5):1043–1065, 1996.

[81] F. Agrafioti, D. Hatzinakos and A.K. Anderson. “ECG Pattern Analysis for Emotion Detection”. *IEEE Transactions on Affective Computing*, 3(1):102–115, 2011.

[82] F. Galvão, S.M. Alarcão and M.J. Fonseca. “Predicting Exact Valence and Arousal Values from EEG”. *Sensors*, 21(10):3414, 2021.

[83] F. Li, et al. “A Classification Framework for Depressive Episode using RR Intervals from Smartwatch”. *IEEE Transactions on Affective Computing*, 2023.

[84] F. Nasoz, et al. “Emotion Recognition from Physiological Signals Using Wireless Sensors for Presence Technologies”. *Cognition, Technology & Work*, 6(1):4–14, 2004.

[85] F. Nasoz et al. "Emotion Recognition from Physiological Signals using Wireless Sensors for Presence Technologies". *Cognition, Technology & Work*, 6(1):4–14, 2004.

[86] F. Ringeval, et al. "Prediction of Ssynchronous Dimensional Emotion Ratings from Audiovisual and Physiological Data". *Pattern Recognition Letters*, 66:22–30, 2015.

[87] G. Rigas, et al. "A user Independent, Biosignal Based, Emotion Recognition Method". In *Proc. International Conference on User Modeling*, pages 314–318. Springer, 2007.

[88] G. Valenza, A. Lanata and E.P. Scilingo. "The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition". *IEEE transactions on Affective Computing*, 3(2):237–249, 2011.

[89] G. Valenza, et al. "Revealing Real-Time Emotional Responses: A Personalized Assessment Based on Heartbeat Dynamics". *Scientific Reports*, 4(1):1–13, 2014.

[90] H. Candra, et al. "Investigation of Window Size in Classification of EEG-Emotion Signal with Wavelet Entropy and Support Vector Machine". In *Proc. IEEE 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7250–7253, 2015.

[91] H. Chen, et al. "3D Model-Based Continuous Emotion Recognition". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1836–1845, 2015.

[92] H. Durrant-Whyte, and T.C. Henderson. "Multisensor Data Fusion". *Springer Handbook of Robotics*, pages 867–896, 2016.

[93] H. Ferdinando, T. Seppänen and E. Alasaarela. "Comparing Features From ECG Pattern and HRV Analysis for Emotion Recognition System". In *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–6. IEEE, 2016.

[94] H. Guo, et al. "Heart Rate Variability Signal Features for Emotion Recognition by using Principal Component Analysis and Support Vectors Machine". In *Proc. IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 274–277, 2016.

[95] H. Lee, et al. "Risk of Data Leakage in Estimating the Diagnostic Performance of a Deep-Learning-Based Computer-Aided System for Psychiatric Disorders". *Scientific Reports*, 13(1):16633, 2023.

[96] H. Palo, M.N. Mohanty and M. Chandra. "Use of Different Features for Emotion Recognition Using MLP Network". In *Computational Vision and Robotics*, pages 7–15. Springer, 2015.

[97] H. Sandler, et al. "Positive Emotional Experience: Induced by Vibroacoustic Stimulation using A Body Monochord in Patients With Psychosomatic Disorders: Is Associated With An Increase in EEG-Theta and A Decrease in EEG-Alpha Power". *Brain Topography*, 29(4):524–538, 2016.

[98] H. Tang, et al. "Multimodal Emotion Recognition using Deep Neural Networks". In *Proc. International Conference on Neural Information Processing*, pages 811–819. Springer, 2017.

[99] H. Zhu, et al. "Federated Learning on Non-IID data: A Survey". *Neurocomputing*, 465:371–390, 2021.

[100] T. Hassani. Federated Emotion Recognition with Physiological Signals-GSR. Master's thesis, 2021.

[101] H.P. Martinez, Y. Bengio and G.N. Yannakakis. "Learning Deep Physiological Models of Affect". *IEEE Computational Intelligence Magazine*, 8(2):20–33, 2013.

[102] H.U.R Siddiqui, et al. "Respiration Based Non-Invasive Approach for Emotion Recognition using Impulse Radio Ultra Wide Band Radar and Machine Learning". *Sensors*, 21(24):8336, 2021.

[103] H.U.R Siddiqui, et al. "Emotion Classification using Temporal and Spectral Features from IR-UWB-Based Respiration Data". *Multimedia Tools and Applications*, 82(12):18565–18583, 2023.

[104] I. Alsafy and M. Diykh. "Developing a Robust Model to Predict Depth of Anesthesia from Single Channel EEG Signal". *Physical and Engineering Sciences in Medicine*, 45(3):793–808, 2022.

[105] I Balelli, et al. Fed-MIWAE: Federated Imputation of Incomplete Data via Deep Generative Models. *arXiv preprint arXiv:2304.08054*, 2023.

[106] I. Homma and Y. Masaoka. "Breathing Rhythms and Emotions". *Experimental Physiology*, 93(9):1011–1021, 2008.

[107] I. Lin, et al. "Cardiac Autonomic and Cardiac Vagal Control uring and After Depressive and Happiness Autobiographical Memories in Patients with Major Depressive Disorder". *Frontiers in Psychiatry*, 13:878285, 2022.

[108] I. Siegert, R. Böck and A. Wendemuth. "Inter-Rater Reliability for Emotion Annotation in Human-Computer Interaction: Comparison and Methodological Improvements". *Journal on Multimodal User Interfaces*, 8:17–28, 2014.

[109] J. Chen, et al. "Feature-Level Fusion of Multimodal Physiological Signals for Emotion Recognition". In *Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 395–399, 2015.

[110] J. Chen, et al. "Federated Learning for Bearing Fault Diagnosis with Dynamic Weighted Averaging". In *Proc. IEEE International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*, pages 1–6, 2021.

[111] J. Cheng, et al. "Emotion Recognition From Multi-Channel EEG Via Deep Forest". *IEEE Journal of Biomedical and Health Informatics*, 25(2):453–464, 2020.

[112] J. Costa, et al. "EmotionCheck: A Wearable Device to Regulate Anxiety Through False Heart Rate Feedback". *GetMobile: Mobile Computing and Communications*, 21(2):22–25, 2017.

[113] J. Delaney, and D. Brodie. "Effects of Short-Term Psychological Stress on the Time and Frequency Domains of Heart-Rate Variability". *Perceptual and Motor Skills*, 91(2):515–524, 2000.

[114] J. Hamer, M. Mohri and A.T. Suresh. "Fedboost: A Communication-Efficient Algorithm for Federated Learning". In *International Conference on Machine Learning*, pages 3973–3983. PMLR, 2020.

[115] J. Kröger. "Unexpected Inferences from Sensor Data: A Hidden Privacy Threat in the Internet of Things". In *Internet of Things. Information Processing in an Increasingly Connected World: First IFIP International Cross-Domain Conference, IFIPIoT 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-19, 2018, Revised Selected Papers 1*, pages 147–159. Springer, 2019.

[116] J. LaRocco, M.D. Le and D.G.Paeng. "A Systemic Review of Available Low-Cost EEG Headsets used for Drowsiness Detection". *Frontiers in Neuroinformatics*, page 42, 2020.

[117] J. Lee and S. Yoo. "Design of User-customized Negative Emotion Classifier Based on Feature Selection Using Physiological Signal Sensors". *Sensors*, 18(12):4253, 2018.

[118] J. Liu, Y. Su and Y. Liu. "Multi-modal Emotion Recognition with Temporal-band Attention Based on LSTM-RNN". In *Pacific Rim Conference on Multimedia*, pages 194–204. Springer, 2017.

[119] J. Morris. "Observations: SAM: the Self-Assessment Manikin; An Efficient Cross-cultural Measurement of Emotional Response". *Journal of Advertising Research*, 35(6):63–68, 1995.

[120] J. S. Heinisch, I. Hübener and K. David. "The Impact of Physical Activities on the Physiological Response to Emotions". In *Proc. IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 824–829, 2018.

[121] J. Shukla, et al. "Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity". *IEEE Transactions on Affective Computing*, 12(4):857–869, 2019.

[122] J. Wagner, J. Kim and E. André. "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification". In *Proc. IEEE International Conference on Multimedia and Expo*, pages 940–943, 2005.

[123] J. Wang, et al. "Functional Brain Network Based Multi-Domain Feature Fusion of Hearing-Impaired EEG Emotion Identification". *Biomedical Signal Processing and Control*, 85:105013, 2023.

[124] J. Xu, et al. "Privacy Preserving Personalized Blockchain Reliability Prediction via Federated Learning in IoT Environments". *Cluster Computing*, 25(4):2515–2526, 2022.

[125] J.A. Miranda-Correa, et al. "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups". *IEEE Transactions on Affective Computing*, 12(2):479–493, 2018.

[126] J.A. Russell. "Affective space is bipolar." *Journal of personality and Social Psychology*, 37(3):345, 1979.

[127] J.A. Russell. "A circumplex model of affect." *Journal of Personality and Social Psychology*, 39(6):1161, 1980.

[128] J.A. Russell and A. Mehrabian. “Evidence for a Three-Factor Theory of Emotions”. *Journal of Research in Personality*, 11(3):273–294, 1977.

[129] J.E. Arco et al. “Identifying HRV patterns in ECG Signals as Early Markers of Dementia”. *Expert Systems with Applications*, 243:122934, 2024.

[130] J.K. Vuoskoski and T. Eerola. “The Role of Mood and Personality in the Perception of Emotions Represented by Music”. *Cortex*, 47(9):1099–1106, 2011.

[131] J.S. Kumar and P. Bhuvaneswari. “Analysis of Electroencephalography (EEG) Signals and Its Categorization- A Study”. *Procedia Engineering*, 38:2525–2536, 2012.

[132] K. Dehdar, et al. “Allergen-Induced Anxiety-like Behavior is Associated with Disruption of Medial Prefrontal Cortex-Amygdala Circuit”. *Scientific Reports*, 9(1):19586, 2019.

[133] K. Kipli, et al. “GSR Signals Features Extraction For Emotion Recognition”. In *Proceedings of Trends in Electronics and Health Informatics: TEHI 2021*, pages 329–338. Springer, 2022.

[134] K. Sharma, et al. “A Dataset of Continuous Affect Annotations and Physiological Signals for Emotion Analysis”. *Scientific Data*, 6(1):1–13, 2019.

[135] Katsigiannis, et al. “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(1):98–107, 2017.

[136] K.J. Rahman et al. “Challenges, Applications and Design Aspects of Federated Learning: A Survey”. *IEEE Access*, 9:124682–124700, 2021.

[137] K.J. Rahman, et al. “Challenges, Applications and Design Aspects of Federated Learning: A Survey”. *IEEE Access*, 9:124682–124700, 2021.

[138] K.P. Rawn and P.S. Keller. “Child Emotion Lability is Associated with Within-task Changes of Autonomic Activity During a Mirror-Tracing Task”. *Psychophysiology*, page e14354, 2023.

[139] L. Canales, et al. “EmoLabel: Semi-Automatic Methodology for Emotion Annotation of Social Media Text”. *IEEE Transactions on Affective Computing*, 13(2):579–591, 2019.

[140] L. Constantine and H. Hajj. “A Survey of Ground-Truth in Emotion Data Annotation”. In *Proc. IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 697–702, 2012.

[141] L. Gong, et al. "EEG emotion recognition using attention-based convolutional transformer neural network". *Biomedical Signal Processing and Control*, 84:104835, 2023.

[142] L. Petrescu, et al. "Integrating Biosignals Measurement in Virtual Reality Environments for Anxiety Detection". *Sensors*, 20(24):7088, 2020.

[143] L. Santamaria-Granados, et al. "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)". *IEEE Access*, 7:57–67, 2018.

[144] L. Siqi, et al. "Fedimpute: Privacy-Preserving Missing Value Imputation for Multi-Site Heterogeneous Electronic Health Records". *Available at SSRN 4930174*.

[145] L.D. Sharma and A. Bhattacharyya. "A Computerized Approach for Automatic Human Emotion Recognition using Sliding Mode Singular Spectrum Analysis". *IEEE Sensors Journal*, 21(23):26931–26940, 2021.

[146] Liakopoulos, et al. "CNN-Based Stress and Emotion Recognition in Ambulatory Settings". In *Proc. IEEE 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–8, 2021.

[147] Liu, Zixuan et al. "Progress in Data Acquisition of Wearable Sensors". *Biosensors*, 12(10):889, 2022.

[148] L.Zhang, et al. "BioVid Emo DB: A Multimodal Database for Emotion Analyses Validated by Subjective Ratings". In *Proc. IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6, 2016.

[149] M. Agrawal and M.A. Anwar and R. Jindal. "FedCER-Emotion Recognition Using 2D-CNN in Decentralized Federated Learning Environment". In *Proc. IEEE 6th International Conference on Information Systems and Computer Networks (ISCON)*, pages 1–5, 2023.

[150] M. Alex et al. "Discrimination of Genuine and Acted Emotional Expressions using EEG Signal and Machine Learning". *IEEE Access*, 8:191080–191089, 2020.

[151] M. Ali et al. "A Globally Generalized Emotion Recognition System Involving Different Physiological Signals". *Sensors*, 18(6):1905, 2018.

[152] M. Benedek, Mathias and C. Kaernbach. "A Continuous Measure of Phasic Electrodermal Activity". *Journal of Neuroscience Methods*, 190(1):80–91, 2010.

[153] M. Egger, M. Ley and S. Hanke. “Emotion Recognition from Physiological Signal Analysis: A Review”. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.

[154] M. Khateeb, S.M. Anwar and M. Alnowami. “Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset”. *IEEE Access*, 9:12134–12142, 2021.

[155] M. Kumar, D. Pandey and P. Shrivastva. “Effect of GSR Biofeedback Relaxation Training on Blood Glucose and Anxiety Level of Type 2 Diabetic Patients”. *International Journal of Indian Psychology*, 4(1):82, 2016.

[156] M. Liu, et al. “Human Emotion Recognition Based on Galvanic Skin Response Signal Feature Selection and SVM”. In *Proc. IEEE International Conference on Smart City and Systems Engineering (ICSCSE)*, pages 157–160, 2016.

[157] M. Meng, et al. “EEG-based emotion recognition with cascaded convolutional recurrent neural networks”. *Pattern Analysis and Applications*, pages 1–13, 2023.

[158] M. Moshawrab, et al. “Reviewing Federated Learning Aggregation Algorithms; Strategies, Contributions, Limitations and Future Perspectives”. *Electronics*, 12(10):2287, 2023.

[159] M. Ragot, et al. “Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors”. In *Proc. International Conference on Applied Human Factors and Ergonomics*, pages 15–22. Springer, 2017.

[160] M. Shukla and O. Seneviratne. “MentalHealthAI: Utilizing Personal Health Device Data to Optimize Psychiatry Treatment”. In *AMIA Annual Symposium Proceedings*, volume 2023, page 641. American Medical Informatics Association, 2023.

[161] M. Soleymani, et al. “A Multimodal Database for Affect Recognition and Implicit Tagging”. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2011.

[162] M. Soukhtanlou, et al. “Electrophysiological Processing of Happiness During Conscious and Sub-conscious Awareness in Depression”. *Neurology, Psychiatry and Brain Research*, 33:32–38, 2019.

[163] M. Tamir. “What Do People Want to Feel and Why? Pleasure and Utility in Emotion Regulation”. *Current Directions in Psychological Science*, 18(2):101–105, 2009.

[164] M. Wang, et al. "Identifying Personal Physiological Data Risks to the Internet of Everything: The Case of Facial Data Breach Risks". *Humanities and Social Sciences Communications*, 10(1):1–15, 2023.

[165] M.A. Anwar et al. "FedEmo: A Privacy-Preserving Framework for Emotion Recognition using EEG Physiological Data". In *Proc. IEEE 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 119–124, 2023.

[166] M.A. Hasnul, et al. "Electrocardiogram-based Emotion Recognition Systems and Their Applications in Healthcare—A Review". *Sensors*, 21(15):5015, 2021.

[167] M.A. Hasnul, et al. "Augmenting ECG Data With Multiple Filters For a Better Emotion Recognition System". *Arabian Journal for Science and Engineering*, pages 1–22, 2023.

[168] M.B.H Wiem and Z. Lachiri. "Emotion Classification in Arousal Valence Model using MAHNOB-HCI Database". *International Journal of Advanced Computer Science and Applications*, 8(3), 2017.

[169] M.D. Hssayeni and B. Ghoraani. "Multi-Modal Physiological Data Fusion for Affect Estimation Using Deep Learning". *IEEE Access*, 9:21642–21652, 2021.

[170] M.G. Arivazhagan et al. "Federated Learning with Personalization Layers". *arXiv preprint arXiv:1912.00818*, 2019.

[171] M.H Purnomo. "Analysis the Dominant Location of Brain Activity in Frontal Lobe Using K-means Method". In *Proc. IEEE International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering 2009*, pages 1–3, 2009.

[172] M.K. Singh, et al. "A Deep Learning Approach for Subject-Dependent & Subject-Independent Emotion Recognition Using Brain Signals with Dimensional Emotion Model". *Biomedical Signal Processing and Control*, 84:104928, 2023.

[173] M.N. Dar, et al. "CNN and LSTM-Based Emotion Charting using Physiological Signals". *Sensors*, 20(16):4551, 2020.

[174] M.P. Tarvainen, et al. "Kubios HRV–Heart Rate Variability Analysis Software". *Computer Methods and Programs in Biomedicine*, 113(1):210–220, 2014.

[175] M.R. Kumar and R.Y. Srinivasa. "Epileptic Seizures Classification in EEG Signal Based on Semantic Features and Variational Mode Decomposition". *Cluster Computing*, 22:13521–13531, 2019.

[176] N. Kumari, S. Anwar and V. Bhattacharjee. "Time Series-Dependent Feature of EEG Signals for Improved Visually Evoked Emotion Classification Using EmotionCapsNet". *Neural Computing and Applications*, 34(16):13291–13303, 2022.

[177] N. Rehman and D.P. Mandic. "Multivariate Empirical Mode Decomposition". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2117):1291–1302, 2010.

[178] N. Saffaryazdi, et al. "Emotion Recognition in Conversations using Brain and Physiological Signals". In *Proc. of the 27th International Conference on Intelligent User Interfaces*, pages 229–242, 2022.

[179] N. Thammasan, K. Fukui and M. Numao. "Application of Deep Belief Networks in EEG-based Dynamic Music-Emotion Recognition". In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 881–888. IEEE, 2016.

[180] N.E. Huang, et al. "The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis". *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995, 1998.

[181] N Nourbakhsh, Y Wang, and F Chen. Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pages 420–423, 2012.

[182] N.T. Pham, et al. "Speech Emotion Recognition using Overlapping Sliding Window and Shapley Additive Explainable Deep Neural Network". *Journal of Information and Telecommunication*, pages 1–19, 2023.

[183] O. AlShorman et al. "The Effects of Emotional Stress on Learning and Memory Cognitive Functions: An EEG Review Study in Education". In *Proc. IEEE 6th International Conference on E-Learning (econf)*, pages 177–182, 2020.

[184] O. Bălan, et al. "Emotion Classification Based on Biophysical Signals and Machine Learning Techniques". *Symmetry*, 12(1):21, 2019.

[185] P. Bota, et al. "Emotion Assessment Using Feature Fusion and Decision Fusion Classification Based on Physiological Data: Are We There Yet?" *Sensors*, 20(17):4723, 2020.

[186] P. Christou, A. Simillidou and M.C. Stylianou. "Tourists' Perceptions Regarding the Use of Anthropomorphic Robots in Tourism and Hospitality". *International Journal of Contemporary Hospitality Management*, 2020.

[187] P. Das, A. Khasnobish and Tibarewala. "Emotion Recognition Employing ECG and GSR Signals as Markers of ANS". In *Proc. IEEE Conference on Advances in Signal Processing (CASP)*, pages 37–42. IEEE, 2016.

[188] P. Ekman. "An Argument for Basic Emotions". *Cognition & Emotion*, 6(3-4):169–200, 1992.

[189] P. Gong, H.T Ma and Y. Wang. "Emotion Recognition Based on the Multiple Physiological Signals". In *Proc.IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 140–143, 2016.

[190] P. Jahankhani, V. Kodogiannis and K. Revett. "EEG Signal Classification using Wavelet Feature Extraction and Neural Networks". In *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*, pages 120–124. IEEE, 2006.

[191] P. Mini, T. Thomas and R. Gopikakumari. "Wavelet Feature Selection of Audio and Imagined/Vocalized EEG Signals for ANN Based Multimodal ASR System". *Biomedical Signal Processing and Control*, 63:102218, 2021.

[192] P. Qi, et al. "A blockchain-based Secure Internet of Medical Things Framework for Stress Detection". *Information Sciences*, 628:377–390, 2023.

[193] P. Sarkar and A. Etemad. "Self-Supervised ECG Representation Learning for Emotion Recognition". *IEEE Transactions on Affective Computing*, 2020.

[194] P. Sarma and S. Barma. "Emotion Recognition by Distinguishing Appropriate EEG Segments Based on Random Matrix Theory". *Biomedical Signal Processing and Control*, 70:102991, 2021.

[195] P. Schmidt, et al. "Introducing WESAD, A Multimodal Dataset for Wearable Stress and Affect Detection". In *Proc. ACM 20th International Conference on Multimodal Interaction*, pages 400–408, 2018.

[196] P. Schmidt, et al. "Wearable Affect and Stress Recognition: A Review". *arXiv preprint arXiv:1811.08854*, 2018.

[197] P. Vanhaesebrouck, A. Bellet and M. Tommasi. “Decentralized Collaborative Learning of Personalized Models Over Networks”. In *Artificial Intelligence and Statistics*, pages 509–517. PMLR, 2017.

[198] P. Vos, et al. “The Tell-tale: What do Heart Rate; Skin Temperature and Skin Conductance Reveal about Emotions of People with Severe and Profound Intellectual Disabilities?” *Research in Developmental Disabilities*, 33(4):1117–1127, 2012.

[199] P. Xiao, et al. “Inter-Subject Prediction of Pediatric Emergence Delirium using Feature Selection and Classification from Spontaneous EEG Signals”. *Biomedical Signal Processing and Control*, 80:104359, 2023.

[200] P.C. Schmid and M.M. Schmid. “Mood Effects on Emotion Recognition”. *Motivation and Emotion*, 34:288–292, 2010.

[201] P.G. Malghan and M.K. Hota. “A Review on ECG Filtering Techniques for Rhythm Analysis”. *Research on Biomedical Engineering*, 36:171–186, 2020.

[202] P.Huang. “Decoding Emotions: Intelligent Visual Perception for Movie Image Classification Using Sustainable AI in Entertainment Computing”. *Entertainment Computing*, 50:100696, 2024.

[203] P.J Bota, et al. “A Review, Current Challenges, and Future Possibilities on Emotion Recognition using Machine Learning and Physiological Signals”. *IEEE Access*, 7:140990–141020, 2019.

[204] P.J. Lang. “The Emotion Probe: Studies of Motivation and Attention.” *American psychologist*, 50(5):372, 1995.

[205] P.J. Lang, et al. *“International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual”*. NIMH, Center for the Study of Emotion & Attention Gainesville, FL, 2005.

[206] Q. Gao et al. “EEG Based Emotion Recognition Using Fusion Feature Extraction Method”. *Multimedia Tools and Applications*, 79:27057–27074, 2020.

[207] Q. Jing, et al. “Quantifying The pPerformance of Federated Transfer Learning”. *arXiv preprint arXiv:1912.12795*, 2019.

[208] Q. Li, et al. “A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection”. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3347–3366, 2021.

[209] Q. Mao, et al. “Learning Salient Features for Speech Emotion Recognition Using Convolutional Neural Networks”. *IEEE Transactions on Multimedia*, 16(8):2203–2213, 2014.

[210] Q. Xiong , et al. “A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI”. *Computational and Mathematical Methods in Medicine*, 2020, 2020.

[211] Q. Yang, et al. “Vertical Federated Learning”. In *Federated Learning*, pages 69–81. Springer, 2020.

[212] R. Agarwal, M. Andujar and S. Canavan. “Classification of Emotions using EEG Activity Associated with Different Areas of The Brain”. *Pattern Recognition Letters*, 162:71–80, 2022.

[213] R. Duan, J. Zhu and B. Lu. “Differential Entropy Feature for EEG-based Emotion Classification”. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84. IEEE, 2013.

[214] R. El Mokadem and Y. Ben Maissa and Z. El Akkaoui. “Federated Learning for Energy Constrained Devices: A Systematic Mapping Study”. *Cluster Computing*, 26(2):1685–1708, 2023.

[215] R. Hartmann, et al. “Heart Rate Variability as Indicator of Clinical State in Depression”. *Frontiers in Psychiatry*, 9:735, 2019.

[216] R. Jenke, A. Peer and M. Buss. “Feature Extraction and Selection for Emotion Recognition from EEG”. *IEEE Transactions on Affective computing*, 5(3):327–339, 2014.

[217] R. Pandya. “Generalized Attention Mechanism and Relative Position for Transformer”. *arXiv preprint arXiv:2208.10247*, 2022.

[218] R. Patel, et al. “Suppression of Eye-Blink Associated Artifact Using Single Channel EEG Data by Combining Cross-Correlation with Empirical Mode Decomposition”. *IEEE Sensors Journal*, 16(18):6947–6954, 2016.

[219] R. Plutchik. “A Psycho Evolutionary Theory of Emotions”. *Social Science Information*, 1982.

[220] R. Renger. “A Review of the Profile of Mood States (POMS) in the Prediction of Athletic Success”. *Journal of Applied Sport Psychology*, 5(1):78–84, 1993.

[221] R. Subramanian, et al. "ASCERTAIN: Emotion and Personality Recognition using Commercial Sensors". *IEEE Transactions on Affective Computing*, 9(2):147–160, 2016.

[222] R.A. Prodhan, et al. "Emotion Recognition from Brain Wave Using Multitask Machine Learning Leveraging Residual Connections". In *Proc. International Conference on Machine Intelligence and Emerging Technologies*, pages 121–136. Springer, 2022.

[223] R.M. Mehmood and H.J. Lee. "Emotion Classification of EEG Brain Signal Using SVM and KNN". In *Proc. IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–5, 2015.

[224] J. Roessler and P.A. Gloor. "Measuring Happiness Increases happiness". *Journal of Computational Social Science*, 4(1):123–146, 2021.

[225] R.S. Antunes et al. "Federated Learning for Healthcare: Systematic Review and Architecture Proposal". *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2022.

[226] S. Agarwal and B. Santra and D.P. Mukherjee. "Anubhav: Recognizing Emotions Through Facial Expression". *The Visual Computer*, 34(2):177–191, 2018.

[227] S. Bagherzadeh et al. "A Subject-iIndependent Portable Emotion Recognition System using Synchrosqueezing Wavelet Transform Maps of EEG Signals and ResNet-18". *Biomedical Signal Processing and Control*, 90:105875, 2024.

[228] S. Basu, et al. "Effects of Emotion on Physiological Signals". In *2016 IEEE Annual India Conference (INDICON)*, pages 1–6. IEEE, 2016.

[229] S. Dinde, and A. Paithane. "Human Emotion Recognition using Electrocardiogram Signals". *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(2):194–197, 2004.

[230] S. Dutta, et al. "An Analysis of Emotion Recognition Based on GSR Signal". *ECS Transactions*, 107(1):12535, 2022.

[231] S. Folschweiller and J.F. Sauer. "Respiration-driven brain oscillations in emotional cognition". *Frontiers in Neural Circuits*, 15:761812, 2021.

[232] S. Gannouni, et al. "Emotion Detection using Electroencephalography Signals and A Zero-time Windowing-Based Epoch Estimation and Relevant Electrode Identification". *Scientific Reports*, 11(1):7071, 2021.

[233] S. Koelstra, et al. "DEAP: A Database for Emotion Analysis; using Physiological Signals". *IEEE transactions on Affective Computing*, 3(1):18–31, 2011.

[234] S. Liu, et al. "Subject-Independent Emotion Recognition of EEG Signals Based on Dynamic Empirical Convolutional Neural Network". *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(5):1710–1721, 2020.

[235] S. Nita, et al. "A New Data Sugmentation Convolutional Neural Network for Human Emotion Recognition based on ECG Signals". *Biomedical Signal Processing and Control*, 75:103580, 2022.

[236] S. Sharma, et al. "Secure and Efficient Federated Transfer Learning". In *Proc. IEEE International Conference on Big Data (Big Data)*, pages 2569–2576, 2019.

[237] S. Wu, et al. "Estimation of Valence of Emotion using Two Frontal EEG Channels". In *Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1127–1130, 2017.

[238] Saganowski, et al. Consumer Wearables and Affective Computing for Wellbeing Support. In *MobiQuitous 2020-17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 482–487, 2020.

[239] S.D. Kreibig. "Autonomic Nervous System Activity in Emotion: A Review". *Biological Psychology*, 84(3):394–421, 2010.

[240] S.G. Simon, et al. "Taking Context To Heart: Momentary Emotions, Menstrual Cycle Phase, and Cardiac Autonomic Regulation". *Psychophysiology*, 58(4):e13765, 2021.

[241] S.H. Chuah, and J. Yu. "The Future of Service: The Power of Emotion in Human-Robot Interaction". *Journal of Retailing and Consumer Services*, 61:102551, 2021.

[242] Shikha, et al. "Stacked Sparse Autoencoder and Machine Learning Based Anxiety Classification Using EEG Signals". In *Proc. ACM First International Conference on AI-ML-Systems*, pages 1–7, 2021.

[243] Shikha, et al. "HRV and GSR as Viable Physiological Markers for Mental Health Recognition". In *Proc. IEEE 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 37–42, 2022.

[244] S.K. Khare, et al. "Emotion Recognition and Artificial Intelligence: A Systematic Review (2014–2023) and Research Recommendations". *Information Fusion*, 102:102019, 2024.

[245] S.Z. Bong and M. Murugappan and S. Yaacob. “Analysis of Electrocardiogram (ECG) Signals for Human Emotional Stress Classification”. In *Proc. International Conference on Intelligent Robotics, Automation, and Manufacturing*, pages 198–205. Springer, 2012.

[246] T. Chen, et al. “Emotion Recognition using Empirical Mode Decomposition and Approximation Entropy”. *Computers & Electrical Engineering*, 72:383–392, 2018.

[247] T. Chilimbi, et al. “Project Adam: Building an Efficient and Scalable Deep Learning Training System”. In *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)*, pages 571–582, 2014.

[248] T. Gu, et al. “Frame-Level Teacher-Student Learning With Data Privacy for EEG Emotion Recognition”. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[249] T. Li, et al. “Federated Learning: Challenges, Methods, and Future Directions”. *IEEE Signal Processing Magazine*, 37(3):50–60, 2020.

[250] T. Ryffel, et al. “A Generic Framework for Privacy Preserving Deep Learning”. *arXiv preprint arXiv:1811.04017*, 2018.

[251] T. Song , et al. “EEG Emotion Recognition using Dynamical Graph Convolutional Neural Networks”. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2018.

[252] T. Tuncer and S. Dogan and A. Subasi. “LEDPatNet19: Automated Emotion Recognition Model Based on Nonlinear LED Pattern Feature Extraction Function using EEG Signals”. *Cognitive Neurodynamics*, pages 1–12, 2021.

[253] T. Wink and Z. Nochta. “An Approach for Peer-to-peer Federated Learning”. In *Proc. 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*, pages 150–157, 2021.

[254] T. Xue, et al. “Annotation Tool for Precise Emotion Ground Truth Label Acquisition while Watching 360 VR Videos”. In *Proc. IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 371–372, 2020.

[255] T. Zhang, et al. “Cornet: Fine-Grained Emotion Recognition for Video Watching using Wearable Physiological Sensors”. *Sensors*, 21(1):52, 2020.

[256] Torres. Cristian, et al. “SVM-based Feature Selection Methods for Emotion Recognition from Multimodal Data”. *Journal on Multimodal User Interfaces*, 11(1):9–23, 2017.

[257] T.W. Bae and K.K. Kwon. “Efficient Real-Time R and QRS Detection Method using a Pair of Derivative Filters and Max Filter for Portable ECG Device”. *Applied Sciences*, 9(19):4128, 2019.

[258] V. Bajaj and R.B. Pachori. “Classification of Seizure and Non-seizure EEG Signals using Empirical Mode Decomposition”. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):1135–1142, 2011.

[259] V. Bhavya, and T. Ahamed. “Mood Identification in People Using ECG Signals”. *International Research Journal of Engineering and Technology (IRJET)*, 2018.

[260] D. Vandervoort. “The Importance of Emotional Intelligence in Higher Education”. *Current Psychology*, 25(1):4–7, 2006.

[261] W. Guo, G. Xu and Y. Wang. “Horizontal and Vertical Features Fusion Network Based on Different Brain Regions for Emotion Recognition”. *Knowledge-Based Systems*, 247:108819, 2022.

[262] W. Guo, G. Xu and Y. Wang. “Multi-Source Domain Adaptation With Spatio-Temporal Feature Extractor for EEG Emotion Recognition”. *Biomedical Signal Processing and Control*, 84:104998, 2023.

[263] W. Li, et al. “A Bi-Stream Hybrid Model with MLPBlocks and Self-attention Mechanism for EEG-Based Emotion Recognition”. *Biomedical Signal Processing and Control*, 86:105223, 2023.

[264] W. Liu, et al. “Multimodal Emotion Recognition using Deep Canonical Correlation Analysis”. *arXiv preprint arXiv:1908.05349*, 2019.

[265] W. Luping, W. Wei and L. Bo. “CMFL: Mitigating Communication Overhead for Federated Learning”. In *Proc. IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 954–964, 2019.

[266] W. Tao, et al. “EEG-Based Emotion Recognition Via Channel-Wise Attention and Self Attention”. *IEEE Transactions on Affective Computing*, 2020.

[267] W.A. Awan et al. “An Ensemble Learning Method for Emotion Charting Using Multimodal Physiological Signals”. *Sensors*, 22(23):9480, 2022.

[268] X. Deng, and J. Zhu, and S. Yang. “SFE-Net: EEG-Based Emotion Recognition with Symmetrical Spatial Feature Extraction”. In *Proc 29th ACM International Conference on Multimedia*, pages 2391–2400, 2021.

[269] X. Li, et al. "Exploring EEG Features in Cross-Subject Emotion Recognition". *Frontiers in Neuroscience*, 12:162, 2018.

[270] X. Yin and Y. Zhu and J. Hu. "A Comprehensive Survey of Privacy-Preserving Federated Learning: A Taxonomy, Review, and Future Directions". *ACM Computing Surveys (CSUR)*, 54(6):1–36, 2021.

[271] X. Zhong, et al. "Bi-hemisphere Asymmetric Attention Network: Recognizing Emotion from EEG Signals Based on the Transformer". *Applied Intelligence*, 53(12):15278–15294, 2023.

[272] Y. Chen, and X. Sun, and Y. Jin. "Communication-Efficient Federated Deep Learning with Layerwise Asynchronous Model Update and Temporally Weighted Aggregation". *IEEE Transactions on Neural Networks and Learning Systems*, 31(10):4229–4238, 2019.

[273] Y. Chen, et al. "Asynchronous Online Federated Learning for Edge Devices with Non-IID Data". In *Proc. IEEE International Conference on Big Data (Big Data)*, pages 15–24, 2020.

[274] Y. Cheng, G. Liu and H. Zhang. "The Research of EMG Signal in Emotion Recognition based on TS and SBS Algorithm". In *Proc. 3rd IEEE International Conference on Information Sciences and Interaction Sciences*, pages 363–366, 2010.

[275] Y. Gao, et al. "EEG-GCN: Spatio-temporal and Self-Adaptive Graph Convolutional Networks for Single and Multi-view EEG-based Emotion Recognition". *IEEE Signal Processing Letters*, 29:1574–1578, 2022.

[276] Y, Hsu, et al. "Automatic ECG-based Emotion Recognition in Music Listening". *IEEE Transactions on Affective Computing*, 11(1):85–99, 2017.

[277] Y. Kim and J. Kim. "Human-like Emotion Recognition: Multi-label Learning from Noisy Labeled Audio-Visual Expressive Speech". In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5104–5108, 2018.

[278] Y. Lin, et al. "Multilayer Perceptron for EEG Signal Classification during Listening to Emotional Music". In *TENCON 2007-2007 IEEE Region 10 Conference*, pages 1–3. IEEE, 2007.

[279] Y. Liu, et al. "A Secure Federated Transfer Learning Framework". *IEEE Intelligent Systems*, 35(4):70–82, 2020.

[280] Y. Liu, et al. "Fedcoin: A Peer-to-peer Payment system for Federated Learning". In *Federated Learning: Privacy and Incentive*, pages 125–138. Springer, 2020.

[281] Y. Liu, et al. "Vertical Federated Learning: Concepts, Advances, and Challenges". *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[282] Y. Nagai, C.I. Jones and A. Sen. "Galvanic Skin Response (GSR)/Electrodermal/Skin Conductance Biofeedback on Epilepsy: A Systematic Review and Meta-Analysis". *Frontiers in Neurology*, 10:377, 2019.

[283] Y. Su, et al. "Multi-party Diabetes Mellitus risk prediction based on secure federated learning". *Biomedical Signal Processing and Control*, 85:104881, 2023.

[284] Y. Wu, et al. "Privacy Preserving Vertical Federated Learning for Tree-based Models". *arXiv preprint arXiv:2008.06170*, 2020.

[285] Y. Xu, et al. "Frontal Alpha EEG Asymmetry Before and After Positive Psychological Interventions for Medical Students". *Frontiers in Psychiatry*, 9:432, 2018.

[286] Y. Yang and H.H. Chen. "Machine Recognition of Music Emotion: A Review". *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):1–30, 2012.

[287] Y. Yang, et al. "Emotion Recognition from Multi-Channel EEG Through Parallel Convolutional Recurrent Neural Network". In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2018.

[288] Y. Zhang, C. Cheng and Y. Zhang. "Multimodal Emotion Recognition Using A Hierarchical Fusion Convolutional Neural Network". *IEEE Access*, 9:7943–7951, 2021.

[289] Y. Zhang, et al. "Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and stratified cross validation". *IEEE Access*, 4:8375–8385, 2016.

[290] Y. Zhang, S. Zhang and X. Ji. "EEG-based Classification of Emotions using Empirical Mode Decomposition and Autoregressive Model". *Multimedia Tools and Applications*, 77:26697–26710, 2018.

[291] Yang, Cheng-Jie, et al. "A Convolution Neural Network Based Emotion Recognition System Using Multimodal Physiological Signals". In *Proc. IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, pages 1–2, 2020.

[292] Z. Cheng, et al. “A Novel ECG-Based Real-Time Detection Method of Negative Emotions in Wearable Applications”. In *Proc. IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 296–301, 2017.

[293] Z. Lan, et al. “Real-time EEG-Based Emotion Monitoring Using Stable Features”. *The Visual Computer*, 32:347–358, 2016.

[294] Z. Li, et al. “Emotion Recognition from EEG using RASM and LSTM”. In *Proc. International Conference on Internet Multimedia Computing and Service*, pages 310–318. Springer, 2017.

[295] Z. Li, et al. “Towards Effective Clustered Federated Learning: A Peer-to-peer Framework with Adaptive Neighbor Matching”. *IEEE Transactions on Big Data*, 2022.

[296] Z. Wang, S.B. Ho and E. Cambria. “A Review of Emotion Sensing: Categorization Models and Algorithms”. *Multimedia Tools and Applications*, 79(47):35553–35582, 2020.

[297] Z. Yao and C. Zhao. “FedTMI: Knowledge Aided Federated Transfer Learning for Industrial Missing Data Imputation”. *Journal of Process Control*, 117:206–215, 2022.

[298] Z. Zhou, et al. “An AI-Empowered Affect Recognition Model for Healthcare and Emotional Well-Being using Physiological Signals”. *Cluster Computing*, 26(2):1253–1266, 2023.

[299] W. Zheng and B. Lu. “Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks”. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015.

[300] W. Zheng, J. Zhu, and B. Lu. “Identifying Stable Patterns Over Time for Emotion Recognition From EEG”. *IEEE Transactions on Affective Computing*, 10(3):417–429, 2017.

[301] C. Zuo, Z. Lin, and Y. Zhang. “Why Does Your Data Leak? Uncovering the Data Leakage in Cloud from Mobile pps”. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1296–1310. IEEE, 2019.

Biography - Supervisor

Divyashikha Sethia is an Assistant professor at the Delhi Technological University, serving from 2010 till present. She has worked in the software industry in Telecom and Networking in software and test automation development in leading companies like Cisco System, 2Wire, Force10 Networks, and Future Software (acquired by Aricent) for more than 5 years in the US and India.

Biography - Student

Neha Gahlan is a Ph.D. student in the Software Engineering Department at the Delhi Technological University. She received her bachelor's and master's degrees in Computer Science Engineering from Guru Gobind Singh Indraprastha University, India, in 2019 and 2021, respectively. Her research interests include human cognition, human-machine interaction, edge computing, and signal processing.

Biomedical Signal Processing and Control 94 (2024) 106353

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

AFLEMP: Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data

Neha Gahlan *, Divyashikha Sethia

Delhi Technological University, New Delhi, Delhi, 110042, India

ARTICLE INFO

Keywords:
Federated Learning
Emotion recognition
Multi-modal Physiological data
Attention mechanisms
Data heterogeneity

ABSTRACT

Automated emotion recognition systems utilizing physiological signals are essential for affective computing and intelligent interaction. Combining the multiple physiological signals is more precise and effective in accurately assessing a person's emotional state. These automated emotion recognition systems using conventional machine learning techniques require complete access to the physiological data for emotion state classification, compromising sensitive data privacy. Federated Learning (FL) resolves this issue by preserving the user's privacy and sensitive physiological data while recognizing emotions. However, existing FL methods have limitations in handling data heterogeneity in the physiological data and do not measure communication efficiency and scalability. In response to these challenges, this paper proposes a unique novel framework called *AFLEMP (Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data)* integrating attention mechanism-based Transformer with an Artificial Neural Network (ANN) model. The framework reduces two types of data heterogeneity: (1) Variation Heterogeneity (VH) in multi-modal EEG, GSR, and ECG physiological signal data using attention mechanisms and (2) Imbalanced Data Heterogeneity (IDH) in the FL environment using scaled weighted federated averaging. This paper validates the proposed *AFLEMP* framework on two publicly available emotion datasets, AMIGOS and DREAMER, achieving an average accuracy of 88.30% and 84.10%, respectively. The proposed *AFLEMP* framework proves robust, scalable, and efficient in communication. *AFLEMP* is the first FL framework to propose for emotion recognition using multi-modal physiological signals while reducing data heterogeneity and outperforming existing FL methods.

1. Introduction

The field of affective computing, concentrating on detecting and quantifying human emotions, has emerged as a promising domain within Human-Computer Interaction (HCI) research. It aims to enable machines to comprehend and identify the emotional state of humans. Additionally, it focuses on providing appropriate feedback to regulate, respond, and evaluate emotions. Furthermore, emotion analysis is critical in healthcare, where it helps comprehend a human's cognitive and behavioural functioning.

Over the last decade, there has been extensive research on the connection between emotions and physiological signals [1–5]. Applying various physiological signals to discern and accurately delineate different emotional states has become a promising strategy. Physiological signals like Electroencephalogram (EEG), Electrocardiogram (ECG), Galvanic Skin Response (GSR), Heart Rate (HR), Electrodermal Activity (EDA), Respiration (RESP) and Blood Volume Pulse (BVP) are capable of accurately detecting and measuring a person's emotional state in real-time [6–9]. These are preferable over physical sensors like facial expressions, speech, gestures, and postures for recognizing emotions [10–12] because physical sensors capture external manifestations of emotions. Various factors influence the physical sensors, including cultural differences, individual variations, or conscious attempts to hide or display emotions. On the other hand, physiological responses are more relative to the autonomic nervous system, which operates involuntarily and is less susceptible to conscious control or external influence [13]. In addition, physiological sensors are less intrusive than physical sensors. These sensors provide valuable information about the intensity of the emotional response. For example, signals such as ECG and heart rate help detect low and high arousal levels. EEG activity changes in the brain's frontal and temporal regions are associated with emotional responses like happiness, sadness, and fear. GSR measures changes in the skin's electrical conductance and causes changes in emotional arousal.

Using a single physiological signal is not reliable for emotion recognition due to the presence of artefacts and noises that might distort the signal's features, impacting the accuracy of the emotion recognition

* Corresponding author.
E-mail addresses: nehagahlan_2k21phdcs02@dtu.ac.in (N. Gahlan), divyashikha@dtu.ac.in (D. Sethia).

https://doi.org/10.1016/j.bspc.2024.106353
Received 5 June 2023; Received in revised form 4 March 2024; Accepted 8 April 2024
Available online 13 April 2024
1746-8094/

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19467-3

Federated learning in Emotion Recognition Systems based on physiological signals for privacy preservation: a review

Neha Gahlan[1] · **Divyashikha Sethia**[1]

Received: 16 September 2022 / Revised: 16 February 2024 / Accepted: 21 May 2024

Abstract

Automated Emotion Recognition Systems (ERS) with physiological signals help improve health and decision-making in everyday life. It uses traditional Machine Learning (ML) methods, requiring high-quality learning models for physiological data (sensitive information). However, automated ERS enables data attacks and leaks, significantly losing user privacy and integrity. This privacy problem can be solved using a novel Federated Learning (FL) approach, which enables distributed machine learning model training. This review examines 192 papers focusing on emotion recognition via physiological signals and FL. It is the first review article concerning the privacy of sensitive physiological data for an ERS. The paper reviews the different emotions, benchmark datasets, machine learning, and federated learning approaches for classifying emotions. It proposes a novel multi-modal Federated Learning for Physiological signals based on Emotion Recognition Systems *(Fed-PhyERS)* architecture, experimenting with the AMIGOS dataset and its applications for a next-generation automated ERS. Based on critical analysis, this paper provides the key takeaways, identifies the limitations, and proposes future research directions to address gaps in previous studies. Moreover, it reviews ethical considerations related to implementing the proposed architecture. This review paper aims to provide readers with a comprehensive insight into the current trends, architectures, and techniques utilized within the field.

Keywords Emotion recognition · Federated learning · Physiological signals · Wearable sensors · Privacy

1 Introduction

Emotions play a pivotal role in shaping individuals' lives, exerting a profound influence on their feelings, current state of mind, and decision-making processes. Humans exhibit two

✉ Divyashikha Sethia
divyashikha@dtu.ac.in

Neha Gahlan
nehagahlan_2k21phdcs02@dtu.ac.in

[1] Department of Software Engineering, Delhi Technological University, Rohini, New Delhi 10042, Delhi, India

Cluster Computing
https://doi.org/10.1007/s10586-023-04133-4

Federated learning inspired privacy sensitive emotion recognition based on multi-modal physiological sensors

Neha Gahlan[1] · Divyashikha Sethia[1]

Received: 5 June 2023 / Revised: 21 August 2023 / Accepted: 27 August 2023

Abstract
Traditional machine learning classifiers can automatically evaluate human behaviour and emotion recognition tasks. However, prior research work does not secure users' privacy and personal information because they need complete access to sensitive physiological data. The recently introduced Federated Learning (FL) paradigm can address this problem. FL allows the local model updates to be sent to a central server, combining them to create a global model. It does not allow the global model to access the raw data used to train it. Motivated by the core concept of FL, this paper proposes a novel FL-based Multi-modal Emotion Recognition System (F-MERS) framework combining EEG, GSR, ECG, and RESP physiological sensors data. It uses Multi-layer Perceptron (MLP) as a base model for classifying complex emotions in three dimensions: Valence, Arousal, and Dominance (VAD). The work validates the F-MERS framework with three emotion benchmark datasets, DEAP, AMIGOS, and DREAMER, achieving accuracies of 87.90%, 89.02%, and 79.02%, respectively. It is the first FL-enabled framework for recognizing complex emotions in three dimensions (VAD) with multi-modal physiological sensors. The proposed study assesses the F-MERS framework in two scenarios: (1). Subject dependent and (2). Subject independent, making the framework more generalized and robust. The experimental outcomes indicate that the F-MERS framework is scalable, efficient in communication, and offers privacy preservation over the baseline Non-FL MLP model.

Keywords Emotion recognition system · Federated learning · Physiological sensors · Multi-modal · Privacy · MLP

1 Introduction

Emotions influence a person's physical health and decision-making abilities [1]. For instance, people are more prone to suffering from poor mental health during emotionally stressful times. Also, when a person is not feeling well, their emotional state is unbalanced. Hence, determining the emotional states is vital to ensure improved emotional wellness. There are two categories of indicators for recognizing emotions, as described below:

✉ Neha Gahlan
nehagahlan_2k21phdcs02@dtu.ac.in

Divyashikha Sethia
divyashikha@dtu.ac.in

[1] Department of Software Engineering, Delhi Technological University, Rohini, New Delhi 110042, India

- **Physical indicators:** One is human bodily indicators, such as facial expression [2, 3], speech [4], gesture [5], Eye-tracking [6, 7], posture, and others, which have the advantage of being easy to collect. Nevertheless, it is quite easy for people to alter their body signs, such as their voice or facial expression, to hide their genuine emotions while interacting with others. People might, for instance, smile during a formal social gathering even if they are experiencing bad emotions. Hence, there is no way to guarantee the correctness of these indicators.
- **Physiological indicators:** These indicators capture the electrical activities (physiological responses) of the human body using physiological sensors like - Electroencephalogram (EEG) [8–10], Electrocardiogram (ECG) [11], Electrodermal Activity (EDA) [12], Heart Rate (HR) [13], Blood Volume Pulse (BVP) [14] and Respiration Rate (RESP) [15]. These indicators can

Published online: 24 September 2023

2024 5th International Conference for Emerging Technology (INCET)
Karnataka, India. May 24-26, 2024

Multifaceted Discrete Emotion Recognition from EEG Physiological Signals via Machine Learning Techniques

Amriti Gupta
Dept. of Software Engineering
Delhi Technological University
amritigupta_mc21a11_19@dtu.ac.in

Kunal Nayyar
Dept. of Electronics & Communication Engineering
Delhi Technological University
kunalnayyar_ec21a17_19@dtu.ac.in

Harshit Chopra
Dept. of Computer Engineering
Delhi Technological University
harshitchopra_co21a4_13@dtu.ac.in

Neha Gahlan
Dept. of Software Engineering
Delhi Technological University
nehagahlan_2k21phdcs02@dtu.ac.in

Divyashikha Sethia
Dept. of Software Engineering
Delhi Technological University
divyashikha@dtu.ac.in

2024 5th International Conference for Emerging Technology (INCET) | 979-8-3503-6115-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCET61516.2024.10592936

***Abstract*—Emotions, intricate mental states shaped by neurophysiological alterations influenced by cognitive processes, sensory inputs, behaviours, and diverse experiences, remain a compelling domain for exploration. The Electroencephalogram (EEG) is a potent tool that directly quantifies these emotional variations by capturing brain signals. Recent strides in emotion recognition have harnessed traditional machine learning classifiers to automate the identification of human emotions with remarkable success. This paper delves into the relatively underexplored Discrete Emotion Model, unveiling its capacity to achieve outstanding accuracy despite historical reservations regarding its effectiveness, using the ECSMP (Emotion, Cognition, Sleep, and Multi-model Physiological signals) dataset. It has two distinct environments, Video watching, and CANTAB-based cognitive assessment phases, enabling seamless data collection and analysis. This research effectively quantifies emotion for binary classification (classifies the type of emotion felt) and multiclass classification (classifies the intensity of emotion felt), elevating emotion recognition capabilities through the synergy of EEG technology. The exceptional performance of XGBoost, with a 96.5% accuracy rate in binary classification and 95% in multiclass emotion recognition, highlights its prowess compared to the other models tested.**

***Index Terms*—Emotion Recognition, Physiological signals**

I. Introduction

Emotions are states experienced by individuals that influence their behaviours and thoughts. They arise from both physical and physiological reactions to internal or external stimuli. Differentiating between emotions can be done through facial expressions, behaviours, and physiological responses. Physiological signals are precious in accurately reflecting a person's emotions in real-time. Monitoring these signals has become more accessible through wearable devices such as smartphones and smartwatches. However, improved emotional well-being necessitates developing techniques to extract emotional states from physiological measurements acquired via mobile devices. Electroencephalogram (EEG) is a widely employed and effective physiological signal for discerning emotions [1], [2]. This efficacy arises from its direct capture of signals emanating from the brain.

A. Emotion Model

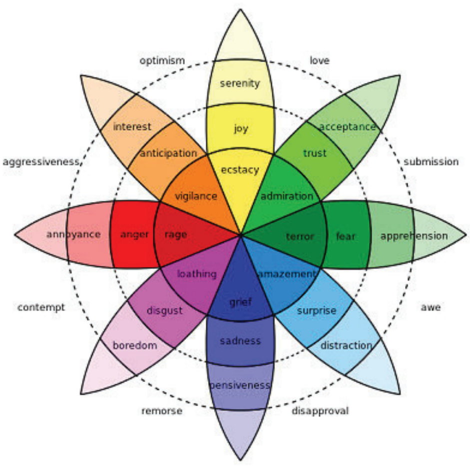


Fig. 1: Plutchik's Emotion Model [3]

Scholars and psychologists have proposed two prominent and widely accepted theories in studying emotions: the Discrete Emotion Model and the Dimensional Emotion Model [4], [5].

From a discrete perspective, influenced by Ekman's theory [15] and evolutionary insights from Darwin and Tomkins, six primary emotions emerge: happiness, sadness, anger, disgust, fear, and surprise. Additional feelings result from intricate combinations or variations of these core emotions. Plutchik [16] introduces a wheel model as shown in Fig. 1, classifying eight primary emotions by intensity. It is similar to mixing primary colors to create secondary hues. Izard [17]

FedEmo: A Privacy-Preserving Framework for Emotion Recognition using EEG Physiological Data

Mohd Ayaan Anwar
Manan Agrawal
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
mohdayaananwar_2k19co232@dtu.ac.in
mananagrawal_2k19co210@dtu.ac.in

Neha Gahlan
Dept. of Software Engineering
Delhi Technological University
Delhi, India
nehagahlan_2k21phdcs02@dtu.ac.in

Divyashikha Sethia
Dept. of Software Engineering
Delhi Technological University
Delhi, India
divyashikha@dtu.ac.in

Gaurav Kumar Singh
Dept. of Software Engineering
Delhi Technological University
Delhi, India
gauravkumarsingh_2k21dsc06@dtu.ac.in

Rishabh Chaurasia
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
rishabhchaurasia_2k18co291@dtu.ac.in

2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS) | 978-1-6654-7706-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/COMSNETS56262.2023.10041308

***Abstract*—Emotions are intricate mental states triggered by neurophysiological adjustments linked to ideas, sensations, behavioral reactions, and a level of pleasure or annoyance. These changes are best traced with the physiological signal Electroencephalogram (EEG), as it records the direct sensations sent by the brain. Recent research on emotion classification methods employs conventional machine learning classifiers to access human emotions and perform automatic emotion recognition tasks. However, they lack in securing users' privacy and sensitive information because they need access to all data. A newly introduced framework Federated Learning (FL), can resolve this problem. It is an approach that aims to create a global model classifier without requiring access to users' local data. This study proposes a novel FL framework, Federated learning for Emotion recognition *(FedEmo)*, for emotion state classification from physiological signal EEG while preserving users' data privacy. It uses Artificial Neural Network (ANN) as a baseline model for classifying emotional states: Arousal, Valence, and Dominance. Adding the concept of federated learning to build a framework *FedEmo* prevents loss of privacy as it enables the local training on the client's end with an updated model from the global server without compromising privacy. The proposed *FedEmo* framework approach achieves accuracies of 63.3%, 56.7%, and 52.2% for Valence, Arousal, and Dominance, respectively, using the well-known DREAMER dataset. These results are comparable to the basic centralized ANN model with the additional development of privacy preservation.**

***Index Terms*—Emotion Recognition, Physiological signals, Federated Learning, Data Privacy**

I. Introduction

Emotions are interpersonal states that have an impact on the behavior and thoughts of a person. They originate from physical and physiological reactions and internal or external stimuli. Face expressions, behavioral patterns, and physiological reactions can all be used to discriminate between various emotions [1]. Out of these, physiological signals are the most preferred parameters as they map the exact emotions of a person at any instance. Recording such signals is performed by the different smart wearables [2]. Recent technology developments have made it easy to access physiological signal monitoring via wearable devices, including smartphones, smartwatches (Empatica E4), and many others. Emotional well-being can be better cared for by creating new techniques to extract emotional states from physiological measurements utilizing mobile devices often, which will have a favorable impact on physical health. The Electroencephalogram (EEG) is a popular and effective physiological signal for identifying different emotions. These signals record the current emotions of a person by sending the signals directly to the brain [3].

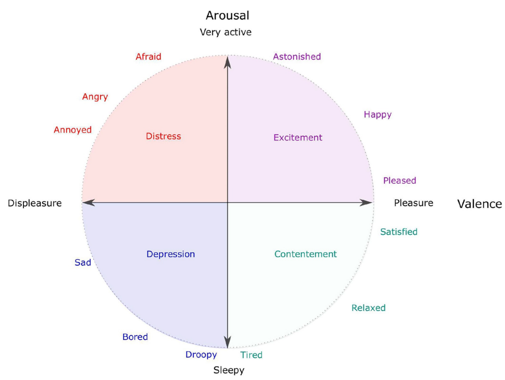


Fig. 1. Russell's circumplex model of affect [4]

Relative to this, Russell [4] proposed a two-dimensional model of emotions. This model categorizes emotions based on valence and arousal (as shown in Fig.1). Mehrabian and Russell [5], [6] extended the two-dimensional model into a three-

2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1) | 978-1-6654-5627-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIA57370.2023.10169633

Three Dimensional Emotion State Classification based on EEG via Empirical Mode Decomposition

Neha Gahlan
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
nehagahlan_2k21phdcs02@dtu.ac.in

Divyashikha Sethia
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
divyashikha@dtu.ac.in

***Abstract*—Electroencephalography (EEG) is useful for mapping emotions directly from the brain, but its heterogeneous signals make it challenging to extract features accurately. Prior works for emotion classification uses EEG data without removing data heterogeneity leading to misclassification or inaccurate classification. This paper proposes an EMD-based methodology for EEG data that segments signals into multiple IMFs to remove heterogeneity and extract significant features. The proposed approach uses a Feed-Forward Neural Network (FFNN) to classify emotions via the VAD model and shows a 5-6% increment in accuracy, precision, and recall scores for emotion classification. Experimental results demonstrate good evaluation performance scores for classifying emotional states on two publicly accessible emotional datasets, AMIGOS and DREAMER.**

***Index Terms*—Emotion Classification, Electroencephalogram (EEG), Empirical Mode Decomposition (EMD), Temporal and Spectral features**

I. Introduction

Electroencephalographic (EEG) signals are the electric impulses recorded to assess brain functions. The brain controls humans' emotional activities, including physical movement, sensory processing, language & communication, memory, and emotions. It investigates various techniques for recognizing emotional states and monitors the cognitive state of humans to assess their emotional health. EEG signals are easily analyzed, and their changes are easily observable, making them a popular and effective means of identifying distinct emotions. Therefore, it should be considered an important element while recognizing emotions. Researchers have become more interested in identifying emotions from EEG signals in recent years [1], [2], [3]. Hence, extracting EEG features and classifying them for emotion recognition have become intriguing research topics.

Wearable devices like smartphones and smartwatches (e.g. Empatica E4) have made it possible to monitor the physiological signals, thanks to recent technological developments. The physiological measurements collected by wearable devices serve as inputs for emotion classification systems driven by artificial intelligence. Automated systems for classifying emotional states are made possible by the use of Deep Learning (DL) and Machine Learning (ML) models. These models can automatically identify correlations between measurements taken under different conditions, analyze large volumes of data, and classify various emotional states. Among the algorithms utilized in these systems are Support Vector Machine (SVM) [2], Decision Trees (DT), K-Nearest Neighbor (KNN) [1], Artificial Neural Network (ANN) [1], and Long Short-Term Memory (LSTM).

Why is EMD required for EEG signals?

The EEG signals are often composed of multiple underlying modes, each with frequency, phase, and amplitude, which can be challenging to identify and separate using traditional linear techniques. The billions of interconnected neurons in EEG signals render them non-linear and non-stationary. These inconsistencies in the EEG signals may lead to misclassification or inaccurate emotional state classification. This issue can be resolved by segmenting the original EEG signals and identifying the segment that causes non-stationarity in the original signal. Huang et al. [4] introduced a signal processing technique based on the Hilbert transformation called Empirical Mode Decomposition (EMD) to address this issue. EMD decomposes the original EEG signals into a set of Intrinsic Mode Functions (IMFs), representing the underlying oscillatory modes present in the signal. These IMFs can then be analyzed individually to gain insight into the different components of EEG signals, such as theta waves and alpha waves. Which IMF component classifies best? By decomposing the EEG signal into IMFs, EMD provides a more detailed representation of the underlying patterns in the signal, which can help analyze brain responses to emotional activities. Additionally, EMD also denoises the signal by removing high-frequency noise that may be present in the signal. EMD has been applied to predicting and detecting seizures from EEG signals [5]. It successfully interprets the inter-modulation distortion and non-stationarity in EEG signals and is adopted for the proposed EEG-based emotion classification framework.

The recent studies that employed EMD on the EEG data are shown in Table I for emotion classification. Zhuang et al. [2] applied EMD to the EEG signals gathered from 32 channels placed on the scalp of the DEAP dataset. The authors used the machine learning-based algorithm SVM to classify only two emotional states: Valence and Arousal. Ahmet et al. [1] used Multivariate Empirical Mode Decomposition (MEMD) for the low & high arousal and valence emotional states.

Emotion Recognition from Facial Expressions using Deep Recurrent Attention Network

Neha Gahlan
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
nehagahlan_2k21phdcs02@dtu.ac.in

Divyashikha Sethia
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
divyashikha@dtu.ac.in

2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS) | 979-8-3503-8311-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/COMSNETS59351.2024.10427068

***Abstract*—Facial expressions serve as a powerful non-verbal communication tool. They convey emotions effectively without the need for verbal expressions. Extensive research endeavors have been dedicated to this field, primarily focusing on employing Neural Networks (NNs) for feature extraction and inference. However, these studies exhibit substantial variation concerning NN architectures and other contributing factors. Hence, this paper introduces a Deep Convolutional Recurrent Attention Network (DCRAN) model architecture with notable performance improvement for Facial Emotion Recognition (FER). The proposed DCRAN model's effectiveness is demonstrated through validation using the well-known FER2013 dataset, achieving a test accuracy of 81.1%.**

***Index Terms*—Facial Emotion Recognition (FER), Neural Networks (NNs), FER2013**

I. Introduction

IoT (Internet of Things) devices can gather facial image data from various sources, such as security cameras, smartphones, and wearable devices. This multi-source data collection enables a more comprehensive dataset for emotion recognition. IoT systems can stream facial image data in real-time to emotion recognition algorithms, allowing for immediate analysis and response [1]. Real-time processing is crucial for applications like mental health monitoring and human-computer interaction, especially for remote monitoring of children and the elderly.

Facial expression recognition serves to identify emotions depicted in facial images, revealing glimpses into the individual's character and behavior. In the 20th century, Ekman and Friesen [2], prominent American psychologists, established a universal set of six fundamental emotions, including sadness, surprise, anger, fear, disgust, and happiness, that transcend cultural boundaries.

In recent years, the field of neural networks has experienced significant advancements driven by the progress in deep learning techniques. Its objective is to enable machines to perceive and interpret visual information like human vision, allowing them to extract meaningful insights, recognize objects, infer context, and perform various tasks related to image analysis and understanding. Deep learning models play a crucial role in this process by extracting complex and meaningful features from images, allowing the model to identify intricate patterns and important details necessary for accurate image classification. The remarkable capability of deep learning models to learn and adapt from large datasets has greatly enhanced the precision of image classification tasks, as evidenced by prior research. Deep learning models can comprehend complex relationships among features and achieve remarkably accurate predictions by harnessing the capabilities of neural networks and the intricate interconnections of multiple layers.

Feelings and emotions are reflected via different facial expressions, exhibiting emotions like anger, disgust, fear, happiness, sadness, and surprise. These emotional cues are captured and documented in several prominent datasets commonly utilized for research in facial expression recognition. Notable datasets in this field encompass JAFFE [3], Cohn-Kanade Dataset (CK) [4], Extended Cohn-Kanade Dataset (CK+) [5], KDEF [6], AffectNet [7], Static Facial Expression in the Wild (SFEW) [8], and FER2013 [9]. Based on prior research in facial expression recognition, FER2013 [9] stands out as the most extensively employed dataset owing to its substantial data size.

This study aims to conduct emotion recognition via facial expressions. The proposed study utilizes the renowned Facial Emotion Recognition 2013 (FER2013) dataset [9]. A large number of CNN-based pre-trained networks are used with the FER2013 dataset for feature extraction and classification [10], [11], [12], [13]. These network models are less intriguing due to the overfitting and bias issues during the classification process. Hence, we propose a hybrid network to overcome this issue and perform accurate classifications.

Motivated by the remarkable ability of deep learning models to extract hierarchical and abstract features, this paper aims to harness their power in capturing intricate features and details that play a pivotal role in accurate image classification. In light of this, the proposed approach introduces a novel Deep Convolutional Recurrent Attention Network (DCRAN) based on deep learning principles. The proposed DCRAN model focuses explicitly on facial image classification to recognize and classify emotions. By leveraging convolutional and recurrent neural networks, the DCRAN model combines features, enabling it to capture and interpret facial expressions' subtle nuances effectively. The utilization of attention mechanisms further enhances the model to consider relevant features of facial regions, leading to improved emotion recognition performance. Through adopting deep learning techniques

GALGOTIAS UNIVERSITY | NAAC GRADE A+ Accredited University | ICCSAI | Technically & Financially sponsored by IEEE UP SECTION (INDIA)

IEEE CONFERENCE
RECORD NO. : #56668

International Conferenceon Communication, Security and Artificial Intelligence ICCSAI - 2022

23rd- 24th December 2022

Certificate of Presenter

This is to certify that Prof./Dr./Mr./Ms Neha Gahlan has presented paper ID 186 entitled Fed-MLP: Federated Framework for Recognizing Human Emotions via Physiological Signal EEG during the **International Conference on communication, security and artificial intelligence, ICCSAI-2022,** (Technically Co-Sponsored by **IEEE UP SECTION**) held on **23rd-24th December 2022** organised by Galgotias University, Greater Noida, Uttar Pradesh, India.

Dr. Avadhesh Kumar	Dr. Nitin Kumar Gaur	Dr. K. Mallikharjuna Babu
Conference Chair, ICCSAI-2022	Conference Chair, ICCSAI-2022	Patron, ICCSAI-2022
PVC, Galgotias University (U.P.)	Registrar, Galgotias University (U.P.)	Vice Chancellor, Galgotias University (U.P.)

NATIONAL BOARD of ACCREDITATION | NATIONAL INSTITUTIONAL RANKING FRAMEWORK | INSTITUTION'S INNOVATION COUNCIL GALGOTIAS UNIVERSITY | MoE's INNOVATION CELL (GOVERNMENT OF INDIA)

IEEE Record No: 57370

ALLIANCE UNIVERSITY

Private University established in Karnataka State by Act No.34 of year 2010
Recognized by the University Grants Commission (UGC), New Delhi

Celebrating 25 years SILVER JUBILEE of Alliance Education

IEEE

IEEE BANGALORE SECTION

ALLIANCE TECHNOLOGY CONFERENCE (ATCON-1)
(HYBRID MODE)

CERTIFICATE OF PRESENTATION

This certificate is presented to

Dr./Prof./Mr./Ms. Neha Gahlan

of Delhi Technological University, Delhi

for presenting the paper titled Three Dimensional Emotion State Classification based on EEG via Empirical Mode Decomposition

in the 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1) hosted by Alliance College of Engineering and Design, Alliance University, Bengaluru on April 21-22, 2023.

Dr. Reeba Korah
Interim Dean
Alliance College of Engineering and Design

N. Mish

Dr. Nivedita Mishra
Registrar
Alliance University

Jain College of Engineering
Belagavi, India

IEEE
BANGALORE SECTION

5th International Conference of Emerging Technology (INCET 2024)

24th – 26th May 2024

Certificate

*This is to certify that Dr./Prof./Mr./Ms. **Neha Gahlan** has presented paper entitled **Multifaceted Discrete Emotion Recognition from EEG Physiological Signals via Machine Learning Techniques** in 5th International Conference of Emerging Technology (INCET 2024) during 24th to 26th May 2024.*

Dr. Krupa Rasane
Convener - INCET 2024

Dr. J. Shivakumar
General Chair - INCET 2024

GALGOTIAS UNIVERSITY   NAAC GRADE A+ Accredited University   ICCSAI   Technically & Financially sponsored by IEEE UP SECTION (INDIA)   IEEE CONFERENCE RECORD NO. : #56668

International Conferenceon Communication, Security and Artificial Intelligence ICCSAI - 2022

23rd- 24th December 2022

Certificate of Presenter

This is to certify that Prof./Dr./Mr./Ms .... Neha Gahlan .... has presented paper ID .... 186 .... entitled Fed-MLP: Federated Framework for Recognizing Human Emotions via Physiological Signal EEG .... during the **International Conference on communication, security and artificial intelligence, ICCSAI-2022,** (Technically Co-Sponsored by **IEEE UP SECTION**) held on **23rd-24th December 2022** organised by Galgotias University, Greater Noida, Uttar Pradesh, India.

Dr. Avadhesh Kumar	**Dr. Nitin Kumar Gaur**	**Dr. K. Mallikharjuna Babu**
Conference Chair, ICCSAI-2022	Conference Chair, ICCSAI-2022	Patron, ICCSAI-2022
PVC, Galgotias University (U.P.)	Registrar, Galgotias University (U.P.)	Vice Chancellor, Galgotias University (U.P.)

NBA NATIONAL BOARD of ACCREDITATION | nirf NATIONAL INSTITUTIONAL RANKING FRAMEWORK | INSTITUTION'S INNOVATION COUNCIL GALGOTIAS UNIVERSITY | MoE's INNOVATION CELL (GOVERNMENT OF INDIA)