Democratic and Popular Algerian Republic
Ministry of Higher Education and Scientific Research



Yahia Fares University of Medea
Faculty of Sciences
Department of Mathematics and Computer Science

Master Thesis

Submitted for obtaining the Master's Degree in
Computer Science
Specialty: **Applied Artificial Intelligence**

Machine and Deep Learning Techniques for Emotion Recognition Using EEG Signals and Computer Games

---

Prepared and Presented by:

Benkhaled Amina Douniazed
Rezazi Aya

Supervisor:

Dr.Boudoukhani Nesrine

2025–2026



إهداء آية

بسم الله الرحمن الرحيم

﴿ رَبِّ أَوْزِعْنِي أَنْ أَشْكُرَ نِعْمَتَكَ الَّتِي أَنْعَمْتَ عَلَيَّ وَعَلَىٰ وَالِدَيَّ وَأَنْ أَعْمَلَ صَالِحًا تَرْضَاهُ ﴾

(سورة النمل، الآية 19)

الحمد لله أولاً وآخراً، الحمد لله الذي بنعمته تتمّ الصالحات، وبتوفيقه تُيسَّر الصعاب، وتُنجز الأعمال، وتتحقق الغايات.

أهدي ثمرة هذا الجهد إلى من كان لهم الفضل، بعد الله سبحانه وتعالى، في بلوغي هذه المرحلة من مسيرتي العلمية:

إلى نفسي، التي صبرت حين اشتدّ التعب، وآمنت أن بعد كل صبرٍ أملٌ يُكتب.

إلى والديّ الكريمين، سندي في الحياة، ونور دربي في كلّ خطوةٍ وعطاء، لم يبخلا يوماً بالدعاء أو الوفاء.

إلى إخوتي وعائلتي جميعاً، الذين كان لقربهم أثرٌ طيبٌ في نفسي، ودافعاً يرافقني في كل طريقٍ ومسير.

إلى صديقتي دنيا زاد، رفيقة هذا الإنجاز، التي شاركتني العزم والإصرار في كل لحظةٍ وقرار.

إلى جميع أصدقائي، الذين كانوا جزءاً من هذه الرحلة، وتركوا في طريقي أثراً لا يُمحى وذكريات لا تُنسى.

إلى كل من علّمني حرفاً، من أساتذتي الذين رافقوني طوال مسيرتي، وكان لهم الفضل في بناء معرفتي وصقل فكري.

وإلى الأستاذة بودكاني، المشرفة على هذا المشروع، تقديراً لما قدمته من توجيه وإرشاد ومتابعة، وما أبدته من تعاون وتشجيع، فجزاها الله خير الجزاء.

وفي الختام، أسأل الله أن يبارك هذا العمل، وأن يجعله نافعاً ومثمراً، وأن يوفقني لما فيه الخير والسداد.

وما توفيقي إلا بالله عليه توكلت وإليه أنيب



إهداء

أمينة دنيازاد

بسم الله الرحمن الرحيم

﴿وَلَقَدْ آتَيْنَا دَاوُودَ وَسُلَيْمَانَ عِلْمًا وَقَالَا الْحَمْدُ لِلَّهِ الَّذِي فَضَّلَنَا عَلَىٰ كَثِيرٍ مِنْ عِبَادِهِ الْمُؤْمِنِينَ﴾

(سورة النمل، الآية 15)

الحمد لله الذي فضّلني على كثيرٍ من عباده المؤمنين، ومهّد لي طريق العلم، وهداني لنور كتابه، فكان القرآن لي أنيسًا وثباتًا.

أهدي ثمرة هذا الجهد والعمل، إلى القلوب التي نبضت معي طوال المسير:

إلى من كان دعاؤهما نورًا يشقُّ دربي، أمي وأبي.

إلى من شدَّ الله بهما عضدي، إخوتي: عادل ومحمد أمين.

إلى جدتي الغالية، وإلى من تركوا في نفسي أثرًا طيبًا لا يُنسى، عائلتي .

إلى من تمنيت أن يشهد نجاحي، جدي رحمه الله.

إلى من كانوا خير الرفقة والأنس، وإلى من شاركتني مشقة الطريق، آية.

إلى كل من علمني حرفًا، وأخص بالذكر مشرفتي في هذا العمل، الدكتورة بودكاني نسرين، تقديرًا لتوجيهاتها ودعمها.

أسأل الله أن يجعل هذا الجهد خالصًا لوجهه الكريم، وأن يتقبله منا، وأن يجعله علمًا يُنتفع به.

Abstract

Emotion recognition using electroencephalography (EEG) signals is a rapidly growing research area in affective state analysis, owing to its ability to directly reflect the underlying neural activity associated with participants' internal emotional states. However, this task remains challenging due to significant inter-subject variability and the inherently non-stationary nature of EEG signals, which makes the extraction of stable and discriminative emotional patterns particularly difficult. To address these challenges, this study proposes two classification approaches. The first is a Deep Learning (DL) model that integrates Convolutional Neural Networks (CNN), Multi-Head Attention mechanisms (MHA), and Bidirectional Long Short-Term Memory (BiLSTM) networks to effectively learn spatial and temporal representations from EEG signals. The second approach is a Global Stacking Ensemble framework that leverages multiple base classifiers to enhance performance, stability, and generalization capability. The proposed models are evaluated on the GAMEEMO dataset, which contains four emotional classes, under two experimental settings: Subject-Dependent and Subject-Independent using a Leave-One-Subject-Out (LOSO) cross-validation strategy. Experimental results show that the DL model with Intra-Subject Fine-Tuning achieves an accuracy of 97.78% in the Subject-Dependent setting and 60.31% in the Subject-Independent setting. In contrast, the ensemble model achieves accuracies of 95.31% and 81.70%, respectively. The experimental results clearly demonstrate the effectiveness of Deep Learning (DL) models in capturing highly discriminative emotional representations from EEG signals under Subject-Dependent settings, while Ensemble Learning (EL) methods consistently achieve superior generalization ability and enhanced robustness against inter-subject variability. Quantitatively, the proposed approach outperforms existing state-of-the-art methods by 11.26% and 3.39% in the Subject-Dependent and Subject-Independent settings, respectively.

Keywords: *Emotion Recognition*; *Electroencephalography (EEG)*; *Deep Learning*; *Multi-Head Attention*; *Ensemble Learning*; *GAMEEMO Dataset*; *Inter-Subject Variability.*

الملخّص

يُعَدُّ التعرف على المشاعر باستخدام إشارات تخطيط الدماغ الكهربائي (EEG) من المجالات النشطة بحثيًا في تحليل الحالات العاطفية، نظرًا لقدرتها على عكس النشاط العصبي المرتبط مباشرة بالحالات العاطفية الداخلية للمستخدمين. ومع ذلك، لا تزال هذه المهمة تمثل تحديًا بحثيًا بسبب التباين بين الأفراد (inter-subject variability) والطبيعة غير المستقرة زمنيًا لإشارات EEG، مما يجعل استخراج أنماط عاطفية ثابتة ومميزة تحديًا كبيرًا. وللتعامل مع هذه التحديات، تقترح هذه الدراسة نهجين للتصنيف؛ يعتمد الأول على نموذج تعلم عميق يجمع بين الشبكات العصبية الالتفافية (CNN)، وآلية الانتباه متعدد الرؤوس (Multi-Head Attention)، وشبكات الذاكرة ثنائية الاتجاه (BiLSTM)، بهدف تعلم التمثيلات المكانية والزمنية من إشارات EEG بشكل فعال. أما النهج الثاني فيعتمد على نموذج تجميعي عالمي (Global Stacking Ensemble) يوظف مجموعة من المصنفات لتعزيز الأداء وتحسين الاستقرار وزيادة القدرة على التعميم. تم تقييم النماذج المقترحة على مجموعة بيانات GAMEEMO متعددة الفئات (أربع حالات عاطفية) ضمن إعداديين تجريبيين: الإعداد المعتمد على الفرد (Subject-Dependent) والإعداد المستقل عن الفرد (Subject-Independent) باستخدام استراتيجية Leave-One-Subject-Out (LOSO). تظهر النتائج التجريبية بوضوح فعالية نماذج التعلم العميق (Deep Learning (DL)) في التقاط تمثيلات عاطفية مميزة من إشارات EEG ضمن إعداد Subject-Dependent، في حين تُحقق طرق التعلم التجميعي (Ensemble Learning (EL)) باستمرار قدرة أعلى على التعميم ومرونة أكبر في التعامل مع التباين بين المشاركين (Inter-Subject Variability). وقد حقق النهج المقترح تحسنًا بنسبة 11.26% و3.39% مقارنة بأحدث الأساليب (State-of-the-Art) في إعدادَي Subject-Dependent وSubject-Independent على التوالي.

الكلمات المفتاحية: التعرف على المشاعر (Emotion Recognition)؛ تخطيط الدماغ الكهربائي (Electroencephalography, EEG)؛ التعلم العميق (Deep Learning)؛ الانتباه متعدد الرؤوس (Multi-Head Attention)؛ التعلم التجميعي (Ensemble Learning)؛ مجموعة بيانات GAMEEMO (GAMEEMO Dataset)؛ التباين بين الأفراد (Inter-Subject Variability).

Contents

Contents i

List of Tables v

List of Figures vii

List of Acronyms x

General Introduction 1

Problem Statement . . . 2

Research Objectives . . . 2

1 General Background on Electroencephalography and Emotion Recognition 4

1.1 Introduction . . . 5

1.2 Electroencephalography (EEG) . . . 5

1.2.1 What is Electroencephalography (EEG)? . . . 5

1.2.2 Components of an EEG System . . . 6

1.2.3 EEG Electrodes and Distribution Systems . . . 7

1.2.4 How EEG Signals are Recorded? . . . 9

1.2.5 Characteristics of Electroencephalogram (EEG) Signals . . . 9

1.2.6 Types of EEG Signals . . . 11

1.2.7 Advantages, Disadvantages, and Challenges of EEG . . . 12

1.2.8 General Applications of Electroencephalography (EEG) . . . 12

1.3 Emotion Recognition . . . 14

1.3.1 What is Emotion Recognition? . . . 14

1.3.2 Emotion Recognition using EEG . . . 15

1.3.3 Computer Games as Emotion Elicitation Environments . . . 15

1.4 Relationship between EEG Signals, Emotions, and Computer Games . . . 16

1.5 Conclusion . . . 17

2 Literature Review of EEG-Based Emotion Recognition 18

2.1 Introduction . . . 19

2.2 EEG Signal Acquisition Methods . . . 19

2.2.1 Electrode Placement and Configuration . . . 20

2.2.2 Experimental Protocol and Stimulus Environment . . . 20

2.2.3 Signal Recording and Data Quality Control . . . 20

2.3 Annotation for EEG-Based Emotion Recognition 20
2.4 Datasets for EEG-Based Emotion Recognition 23
2.5 Signal Preprocessing for EEG Signals 24
2.6 Feature Extraction for EEG Signals 26
2.6.1 Time-Domain Features 27
2.6.2 Frequency-Domain Features 27
2.6.3 Time-Frequency Domain Features 28
2.6.4 Non-Linear Domain Features 29
2.7 Data Augmentation for EEG Signals 31
2.8 Background of Machine Learning and Deep Learning for EEG Analysis 32
2.8.1 Traditional Machine Learning Models 32
2.8.2 Deep Learning Models 35
2.9 Related Work 41
2.9.1 Random-Split Evaluation 41
2.9.2 Subject-Dependent Evaluation 43
2.9.3 Subject-Independent Evaluation 45
2.10 Discussion on Emotion Recognition through EEG Analysis 47
2.11 Conclusion 47

3 Research Methodology and Technical Framework 48
3.1 Introduction 49
3.2 Problem Statement 49
3.2.1 Data Scarcity and Accessibility 49
3.2.2 EEG Noise Susceptibility 49
3.2.3 Signal Amplitude Variability 50
3.2.4 Inter-Subject Variability and Generalization Gap 50
3.3 Motivation 50
3.3.1 Addressing Data Scarcity 50
3.3.2 Handling EEG Noise Sensitivity 50
3.3.3 Addressing Signal Amplitude Variability 51
3.3.4 Managing Inter-Subject Variability and Generalization Gap 51
3.4 Dataset Overview 51
3.4.1 Description of the Dataset 51
3.4.2 Participants and Recording Procedures 51
3.4.3 Data Structure and Technical Details 52
3.5 Emotion Representation and Annotation 52
3.5.1 Emotion Representation using the Self-Assessment Manikin (SAM) 52
3.5.2 Annotation Strategy 53
3.6 Proposed Method 56
3.6.1 Preprocessing of EEG Signals 56
3.6.2 Feature Extraction for Machine Learning and Ensemble Learning Models 59
3.6.3 Feature Extraction for Deep Learning and Hybrid Models 60

3.6.4 Machine Learning-Based Models ... 61
3.6.5 Deep Learning-Based Models ... 61
3.6.6 Ensemble Learning-Based Models ... 63
3.6.7 Hybrid Deep Learning and Ensemble Learning Models ... 65
3.6.8 Self-Training Strategy with Hybrid Deep Learning ... 74
3.7 Conclusion ... 74

4 Experimental Setup 75
4.1 Introduction ... 76
4.2 Evaluation Strategy and Performance Metrics ... 76
4.2.1 Subject-Dependent Setting ... 76
4.2.2 Subject-Independent Setting ... 76
4.2.3 Evaluation Metrics ... 77
4.3 Tools and Work Environment ... 78
4.4 Conclusion ... 79

5 Results and Discussion for the Subject-Dependent Approach 80
5.1 Introduction ... 81
5.2 Experimental Setup ... 81
5.2.1 Loss Function and Output Activation ... 81
5.2.2 Hyperparameters for ML and EL Models ... 82
5.2.3 Hyperparameters for DL and Hybrid Models ... 83
5.3 Results and Discussion ... 84
5.3.1 Results of ML-Based Models ... 84
5.3.2 Results of DL-Based Models ... 88
5.3.3 Results of EL-Based Models ... 91
5.3.4 Results of Hybrid Models ... 93
5.3.5 Results of Self-Training Strategy with Hybrid Deep Learning ... 97
5.4 Comparison with Literature ... 102
5.5 Conclusion ... 103

6 Results and Discussion for the Subject-Independent Approach 104
6.1 Introduction ... 105
6.2 Experimental Setup ... 105
6.2.1 Loss Function and Output Activation ... 105
6.2.2 Hyperparameters for ML and EL Models ... 105
6.2.3 Hyperparameters for Hybrid Models ... 106
6.3 Results and Discussion ... 107
6.3.1 Results of ML-Based Models ... 107
6.3.2 Results of EL-Based Models ... 111
6.3.3 Results of Hybrid Models ... 114
6.3.4 Results of Self-Training Strategy with Hybrid Deep Learning ... 117
6.4 Comparison with Literature ... 122

6.5 Conclusion 123

GENERAL CONCLUSION **124**

A Subject-Dependent Classification Results **132**

B Subject-Independent Classification Results **143**

C Detailed Subject Annotations and Demographics **148**

List of Tables

2.1 Comparison of Common EEG Datasets for Emotion Recognition. 23
2.2 Signal Preprocessing for EEG Signals. 24
2.3 EEG Time-Domain Feature Extraction Methods and Objectives. 27
2.4 EEG Frequency-Domain Feature Extraction Methods and Objectives. 28
2.5 EEG Time-Frequency Domain Feature Extraction Methods and Objectives. 29
2.6 EEG Non-Linear Domain Feature Extraction Methods and Objectives. 29
2.7 EEG Data Augmentation Strategies. 31
2.8 Summary of Random-Split Evaluation for EEG-Based Emotion Recognition Studies. . 42
2.9 Summary of Subject-Dependent Evaluation for EEG-Based Emotion Recognition Studies. 44
2.10 Summary of Subject-Independent Evaluation for EEG-Based Emotion Recognition Studies. 46
3.1 Emotional Labels Distribution Across Game Sessions per Subject. 54
3.2 Frequency Distribution of Emotional Labels Across Game Sessions for All Participants. 55
3.3 Feature Grouping Strategy and Number of Features per Group. 60
5.1 Hyperparameter Settings for Machine Learning and Ensemble Learning Models 82
5.2 Hyperparameter Settings for Deep and Hybrid Learning Models. 83
5.3 Average Accuracy (%) for ML Models in Subject-Dependent Approach Across Different Feature Sets Using StandardScaler Normalization. 84
5.4 Comparative Average Classification Accuracy (%) of DL Models Across EEG Feature Extraction Methods with StandardScaler Normalization. 88
5.5 Average Accuracy (%) for Ensemble Learning Models in Subject-Dependent Approach Across Different Feature Sets Using StandardScaler Normalization. 91
5.6 Average Accuracy (%) for Hybrid Deep Learning Architectures Using StandardScaler Normalization and (75%) Overlapping Ratio. 93
5.7 Impact of Normalization Techniques on the Average Classification Accuracy of Self-Training Strategy. 97
5.8 Subject-Wise Detailed Classification Performance (Accuracy, Macro and Weighted Metrics) for RobustScaler Across 27 Subjects. 99
5.9 Comparison with Related Work for EEG-Based Emotion Recognition under Random and Subject-Dependent Split Settings. 102
6.1 Hyperparameter Settings for Deep and Hybrid Learning Models. 106
6.2 Average Accuracy (%) for Machine Learning Models in Subject-Independent Approach Across Different Feature Sets Using StandardScaler Normalization. 107

6.3 Average Accuracy (%) for Ensemble Learning Models in Subject-Independent Approach Across Different Feature Sets Using StandardScaler Normalization 112
6.4 Average Accuracy (%) for Hybrid Deep Learning Architectures with Corresponding Overlapping Ratios. 114
6.5 Subject-Wise Detailed Classification Performance (Accuracy, Macro and Weighted Metrics) for RobustScaler Across 20 Subjects. 118
6.6 Comparison with Related Work for Subject-Independent EEG-Based Emotion Recognition. 122

List of Figures

1.1 Basic EEG System Components. 6
1.2 Classification of EEG Electrode Types [**1**]. 8
1.3 The International 10-20 System for EEG and Its Mapping [**2**]. 8
1.4 Power Spectral Density (PSD) Analysis and EEG Frequency Band Distribution [**3**]. . . 10
1.5 Comparison of Standard EEG Frequency Bands [**4**]. 11
1.6 An Overview of BCI Systems: Components, Processing, and End-Uses [**1**]. 13
1.7 Neural Correlates of Different Emotional Responses.[1] 15
1.8 Experimental Setup for EEG-Based Emotion Recognition during Gaming.[2] 16
2.1 Illustration of the Baseline EEG Data Acquisition Framework. 19
2.2 The 2D Valence-Arousal Space for Dimensional Emotion Representation [**5**]. 21
2.3 Self-Assessment Manikin (SAM) Scales for Subjective Emotion Labeling [**6**]. 22
2.4 Overview of Feature Extraction for EEG Signals. 26
2.5 Illustration of How the Support Vector Machines (SVM) Algorithm Works [**7**]. 33
2.6 Illustration of How the K-Nearest Neighbors (K-NN) Algorithm Works [**7**]. 33
2.7 1D Deep Convolutional Neural Network (CNN) Structure [**8**]. 35
2.8 The Architecture of the EEGNet [**9**]. 36
2.9 Basic Architecture of a Recurrent Neural Network [**10**]. 36
2.10 Structure of an LSTM Cell [**11**]. 37
2.11 Bidirectional LSTM (BiLSTM) Architecture [**11**]. 37
2.12 Architecture of a Gated Recurrent Unit (GRU) [**11**]. 38
2.13 Architecture of a Deep Multilayer Perceptron (MLP) Neural Network [**12**]. 39
2.14 Hybrid Deep Learning Architecture Integrating CNN and LSTM Modules for Emotion Recognition [**13**]. 40
3.1 SAM Scale Used by Subject 16 to Label Valence and Arousal Levels after Game 4. . . . 53
3.2 Visualization of the frequency distribution of emotional labels across game sessions (G1–G4) for all participants. 55
3.3 Overview of The proposed Method for Emotion Recognition Using EEG Signals. 56
3.4 Preprocessing Pipeline for EEG Signals. 57
3.5 Proposed Machine Learning Pipeline Based on Feature Extraction. 61
3.6 Overview of the Proposed Deep Learning and Hybrid Models. 61
3.7 Illustrated Architecture of the Proposed MLPNN Model. 62
3.8 Illustrated Architecture of the Proposed BiLSTM Model. 62
3.9 Illustrated Architecture of the Proposed Deep BiLSTM Model. 63
3.10 Overview of Ensemble Learning Approaches: Local Stacking, Global Stacking (Fusion), and Majority Voting. 64

3.11 Illustrated Architecture of the Proposed CNN + LSTM Model. 65
3.12 Illustrated Architecture of the Proposed CNN + BiLSTM + MHA Model. 66
3.13 Illustrated Architecture of the Proposed CNN + BiGRU + MHA Model. 66
3.14 Illustrated Architecture of the Proposed Transformer + Deep BiLSTM Model. 67
3.15 Illustrated Architecture of the Proposed CNN + MHA + BiLSTM Model. 68
3.16 Illustrated Architecture of the Proposed CNN + MHA + BiLSTM + Ensemble Learning Model. 69
3.17 Illustrated Architecture of the Proposed EEGNet + Attention Model. 70
3.18 Illustrated Architecture of the Proposed DAN Model. 71
3.19 Illustrated Architecture of the Proposed DAMGCNN + Attention Model. 72
3.20 Architecture of the LaBraM Framework **[14]**. 73
3.21 Illustrated Architecture of the Proposed Self-Training Strategy. 74
4.1 Subject-Dependent vs. Subject-Independent Validation Strategies. 77
5.1 Performance Comparison of ML Models Across Distinct Feature Selection Groups (Dependent Approach). 85
5.2 Heatmap Analysis of Classification Accuracy Across ML Models and Feature Sets (Dependent Approach). 85
5.3 Tracking plot illustrating the performance variations of subjects S28, S14, and S09 across different feature sets using the XGBoost classifier. 86
5.4 Subject-Wise Accuracy Distribution and Outlier Analysis Across ML Classifiers and Feature sets (Dependent Approach). 87
5.5 Average Classification Accuracy Across 28 Subjects Using DL Models. 89
5.6 Subject-Wise Accuracy Distribution and Outlier Analysis Across DL Models. 90
5.7 Tracking Plot of Intra-Subject Performance Variations for Subject S14 Across Different Architectures and Feature Sets. 91
5.8 Global Stacking vs Majority Voting Accuracy (28 Subjects)(Dependent Approach). 92
5.9 Distribution of Classification Accuracy Across 28 Subjects Using Local Stacking & Global Stacking vs Majority Voting (Dependent Approach). 93
5.10 Average Classification Accuracy (%) of Hybrid Deep Learning Models Across 28 Subjects (Dependent Approach). 95
5.11 Distribution of Classification Accuracy Across 28 Subjects Using Hybrid Deep Learning Architectures 96
5.12 Tracking Plot of Intra-Subject Performance Variations Across Hybrid Models. 97
5.13 Performance Distribution of Intra-Subject Fine-Tuning Using StandardScaler and RobustScaler. 98
5.14 Performance Evaluation and Confusion Matrix for Subject 14 (S14). 100
5.15 Performance Evaluation and Confusion Matrix for Subject 20 (S20). 100
5.16 Performance Evaluation and Confusion Matrix for Subject 18 (S18). 101
6.1 Performance Comparison of Machine Learning Models Across Distinct Feature Selection Groups (Independent Approach). 108
6.2 Heatmap Analysis of Classification Accuracy Across Machine Learning Models and Feature Sets (Independent Approach). 108

6.3 Subject-Wise Accuracy Distribution and Outlier Analysis Across Machine Learning Classifiers and Feature Subsets (Independent Approach). 110
6.4 Tracking plots of intra-subject performance variations across feature sets for participants exhibiting high variability and relatively stable performance. 111
6.5 Global Stacking vs Majority Voting Accuracy (28 Subjects)(Independent Approach). . 112
6.6 Distribution of Classification Accuracy Across 28 Subjects Using Local Stacking & Global Stacking vs Majority Voting (Independent Approach). 113
6.7 Average Classification Accuracy (%) of Hybrid Deep Learning Models Across 28 Subjects (Independent Approach). 115
6.8 Subject-Wise Accuracy Distribution and Outlier Analysis Across Hybrid Deep Learning Models (Independent Approach). 116
6.9 Tracking Plot Illustrating Intra-Subject Performance Variations for Subjects S14 and S28 Across Different Hybrid Models. 117
6.10 Performance Distribution of Intra-Subject Fine-Tuning (Before and After) Using RobustScaler. 119
6.11 Global Classification Performance and Confusion Matrices. 120
6.12 Valence–Arousal Coordinate Frequency Distribution and Class Mapping Across the Four Experimental Games. 121

List of Acronyms

ADC	Analog-to-Digital Converter
ADHD	Attention Deficit Hyperactivity Disorder
AF3 / AF4	Anterior Frontal electrodes
AIO	All In One (Total AIO Feature Group)
BCI	Brain–Computer Interface
BiGRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short-Term Memory
CIT-EmotionNet	CNN Interactive Transformer Emotion Network
CNN	Convolutional Neural Network
CSV	Comma-Separated Values (file format for EEG data)
CV	Cross-Validation
DAMGCN	Dynamic Adaptive Multigraph Convolutional Network
DAN	Domain-Adversarial Neural Network
DASM	Differential Asymmetry (Spatial Feature)
DE	Differential Entropy
DL	Deep Learning
DSSANet	Directional Spatial and Spectral Attention Network
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram / Electroencephalography
EL	Ensemble Learning
ELU	Exponential Linear Unit (Activation Function)
EMD	Empirical Mode Decomposition
EMG	Electromyogram (Muscle Activity Artifacts)
EOG	Electrooculogram (Eye Movement Artifacts)
ERP	Event-Related Potential
F3 / F4 / F7 / F8	Frontal electrodes
FC5 / FC6	Fronto-Central electrodes

FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
FS	Sampling Frequency (e.g., 128 Hz)
GANs	Generative Adversarial Networks
GCN	Graph Convolutional Network
GPU	Graphics Processing Unit (used to accelerate computations, e.g., Tesla T4)
GRL	Gradient Reversal Layer
GRU	Gated Recurrent Unit
GSR	Galvanic Skin Response
HANV	High Arousal – Negative Valence
HAPV	High Arousal – Positive Valence
HV-Flow	Horizontal and Vertical Flow
IADS	International Affective Digitized Sound System
ICA	Independent Component Analysis
IMFs	Intrinsic Mode Functions
IQR	Interquartile Range
KNN	K-Nearest Neighbors
LaBraM	Large Brain Model
LANV	Low Arousal – Negative Valence
LAPV	Low Arousal – Positive Valence
LDA	Linear Discriminant Analysis
LDS	Linear Dynamical Systems
LightGBM	Light Gradient Boosting Machine
LOSO	Leave-One-Subject-Out
LOTO	Leave-One-Trial-Out
LR	Logistic Regression
LSTM	Long Short-Term Memory
MEG	Magnetoencephalography
MHA	Multi-Head Attention
ML	Machine Learning
MLP	Multilayer Perceptron

MLPNN	Multilayer Perceptron Neural Network
mRMR	Minimum Redundancy Maximum Relevance
O1 / O2	Occipital electrodes
P7 / P8	Parietal electrodes
PANAS	Positive Affect and Negative Affect Schedules
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PSD	Power Spectral Density
RAM	Random Access Memory
RASM	Rational Asymmetry
RCAR	Rule Classification based on Association Rules
RCE	Random Channels Exchange
ReLU	Rectified Linear Unit
RF	Random Forest
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
RNN	Recurrent Neural Network
SAM	Self-Assessment Manikin
SGD	Stochastic Gradient Descent
SMOTE	Synthetic Minority Over-sampling Technique
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
T7 / T8	Temporal electrodes
TN	True Negative
TP	True Positive
VA	Valence–Arousal Model
VAEs	Variational Autoencoders
VMD	Variational Mode Decomposition
WST	Wavelet Scattering Transformation
XGBoost	Extreme Gradient Boosting

General Introduction

The study of human emotions and the understanding of brain activity are among the most important topics in neuroscience and computer science, as they provide deeper insight into the relationship between neural processes and human behavior. In this context, Electroencephalography (EEG) is considered one of the most significant tools, as it enables the indirect recording of the brain's electrical activity.

In recent years, EEG signals have attracted increasing attention in the field of Affective Computing, particularly for emotion recognition tasks, due to their ability to directly reflect individuals' internal emotional states through neural activity. This growing interest stems from the fundamental role that emotions play in shaping human behavior and influencing interactions with intelligent systems, making automatic emotion recognition an important research topic in Human–Computer Interaction (HCI) and Artificial Intelligence (AI).

To investigate and analyze emotional responses, stimulus-based approaches are commonly employed to elicit natural affective states. In this regard, computer games have emerged as an effective medium for emotion elicitation due to their interactive and immersive nature. By combining challenge, engagement, and competition, they are capable of inducing a wide range of emotional states, including excitement, stress, frustration, and enjoyment.

Given the importance of studying emotions in interactive environments such as computer games, the integration of EEG signals with gaming environments provides an effective framework for affective state analysis. This combination enables the recording of brain activity during gameplay and the extraction of meaningful patterns associated with emotional responses. Compared with other modalities, such as facial expressions and speech signals, EEG offers a more objective measure of emotions because it directly reflects neural activity and is less influenced by external factors. Accordingly, this work focuses on the use of EEG signals in computer gaming environments to improve the accuracy and reliability of emotion classification systems.

Within this context, EEG-based emotion recognition has witnessed significant progress through the adoption of both conventional Machine Learning (ML) and Deep Learning (DL) approaches, including CNNs [**15**], RNNs [**16**], and hybrid architectures, as well as specialized models such as EEGNet [**17**] and similar compact EEG-specific frameworks, all aimed at improving classification performance. However, despite these advancements, important challenges still persist, particularly in achieving robust generalization across subjects and real-world scenarios, which highlights the need for a deeper investigation into the limitations of existing methods.

Problem Statement

Despite significant progress in EEG-based emotion recognition, accurately identifying emotional states in computer gaming environments remains challenging. This is mainly due to the intrinsic properties of EEG signals, including low amplitude, noise sensitivity, non-stationarity, and high inter-subject variability. In addition, emotional responses during gameplay are dynamic and continuously evolving, which further complicates the extraction of stable and discriminative patterns. As a result, many existing approaches struggle to maintain consistent performance, particularly when evaluated on unseen subjects, limiting their generalization capability.

To address these limitations, this study is guided by the following research question:

> *To what extent can EEG signals be leveraged to enable accurate, robust, and highly generalizable emotion recognition in interactive computer gaming environments, and which learning paradigms—including machine learning, ensemble methods, deep learning, and hybrid architectures—are most effective in achieving this goal?*

Research Objectives

General Objective

The main objective of this study is to investigate the effectiveness of EEG-based emotion recognition approaches in computer gaming environments and to evaluate different artificial intelligence techniques for improving both classification performance and model generalization.

Specific Objectives

- Investigate the ability of EEG signals to represent emotional states elicited during gameplay.
- Analyze and compare different EEG feature extraction techniques derived from multiple signal domains.
- Evaluate the performance of machine learning and ensemble learning methods for EEG-based emotion recognition.
- Explore deep learning and hybrid architectures for automatic emotional representation learning from EEG signals.
- Assess the impact of preprocessing and feature extraction techniques on classification performance.
- Compare Subject-Dependent and Subject-Independent evaluation settings to analyze model robustness and generalization capability across different subjects.
- Identify the most effective methodology for EEG-based emotion recognition in the GAMEEMO dataset in both Subject-Dependent and Subject-Independent settings, in terms of classification performance and generalization capability.

This thesis is organized into six chapters:

Chapter 1: ***General Background on Electroencephalography and Emotion Recognition***

presents the theoretical background of electroencephalography (EEG), emotion recognition, and the relationship between EEG signals, emotions, and computer gaming environments.

Chapter 2: ***Literature Review of EEG-Based Emotion Recognition***

reviews the existing literature on EEG-based emotion recognition, including signal acquisition, annotation strategies, datasets, preprocessing methods, feature extraction techniques, machine learning and deep learning approaches, as well as recent related studies.

Chapter 3: ***Research Methodology and Technical Framework***

presents the proposed framework and research methodology, including the dataset, emotion annotation process, preprocessing procedures, feature extraction techniques, and the machine learning, ensemble learning, deep learning, and hybrid architectures investigated in this study.

Chapter 4: ***Experimental Setup***

describes the experimental setup, evaluation protocols, performance metrics, implementation tools, and computational environment used throughout the research.

Chapter 5: ***Results and Discussion for the Subject-Dependent Approach***

reports and discusses the results obtained under the Subject-Dependent evaluation setting and provides comparisons with related studies.

Chapter 6: ***Results and Discussion for the Subject-Independent Approach***

presents and discusses the results obtained under the Subject-Independent evaluation setting, investigates model generalization across unseen subjects, and provides comparisons with related studies.

Finally, a **general conclusion** summarizes the main findings and contributions of the study and outlines potential directions for future research.

General Background on Electroencephalography and Emotion Recognition

1.1 Introduction

The study of human emotions and the understanding of brain activity are among the most important topics in neuroscience and computer science, as they provide a deeper insight into the relationship between neural processes and human behavior. In this context, electroencephalography (EEG) is considered one of the most significant tools used, as it allows for the indirect recording of the brain's electrical activity.

EEG signals, which result from the neural activity of brain cells, reflect the changes occurring in the brain in response to various external and internal stimuli, including emotions. By analyzing these signals, it is possible to detect distinctive neural patterns associated with different emotional states, making them the foundation for many modern applications in fields such as mental health, human-computer interaction, and affective computing.

Based on this foundation, this chapter aims to provide a general background on EEG, including its components, types, and the characteristics of neural signals. It will also address the area of emotion recognition using EEG and examine the relationship between neural signals and various stimuli, such as computer games.

1.2 Electroencephalography (EEG)

1.2.1 What is Electroencephalography (EEG)?

Electroencephalography (EEG) is a method used to record the electrical activity of the brain through the changes in the voltage field that can be observed outside the scalp, produced by neural interactions. This is achieved by placing a set of electrodes on the scalp of the individual, which measure the electrical activity of large groups of neurons in the cortical regions, providing insight into brain activity [**18**].

EEG is a safe and non-invasive technique widely used in both clinical practice and research. One of its main advantages is its excellent temporal resolution, meaning that it can capture very rapid changes in brain activity within milliseconds. This property makes EEG particularly suitable for studying fast cognitive and emotional processes. Moreover, EEG is generally more affordable and flexible than other brain imaging techniques such as functional Magnetic Resonance Imaging (fMRI) or Positron Emission Tomography (PET), which contributes to its extensive use in neuroscience and related disciplines [**18**].

Due to these advantages, EEG has been widely applied in research and clinical contexts to investigate brain functions, including attention, memory, sleep studies, neurological and psychiatric disorders, as well as modern applications such as emotional state recognition and brain-computer interfaces (BCIs) [**18**].

1.2.2 Components of an EEG System

An electroencephalography (EEG) system consists of several components that work together to record and analyze the electrical activity of the brain. The main components of an EEG system include electrodes, an amplifier, a data acquisition system, and specialized software used for signal processing and analysis [**19**].

- **Electrodes:** Electrodes are placed on the scalp to capture the electrical signals generated by neural activity in the brain. These electrodes are typically arranged according to the International 10–20 System, which ensures a standardized and consistent distribution of electrodes across the scalp for reliable measurements [**19**].
- **Amplifier:** The amplifier is responsible for increasing the amplitude of the very small electrical signals recorded from the brain, which are typically in the microvolt range. It also helps reduce noise and external interference, allowing for more accurate signal recording [**19**].
- **Acquisition System / ADC:** The acquisition system, often implemented through an Analog-to-Digital Converter (ADC), converts the analog EEG signals into digital form. It also controls important parameters such as the sampling rate and signal resolution [**19**].
- **Software:** Specialized software is used to visualize, store, and process EEG signals. It allows researchers and clinicians to perform preprocessing, feature extraction, and data analysis for various medical and research applications [**19**].

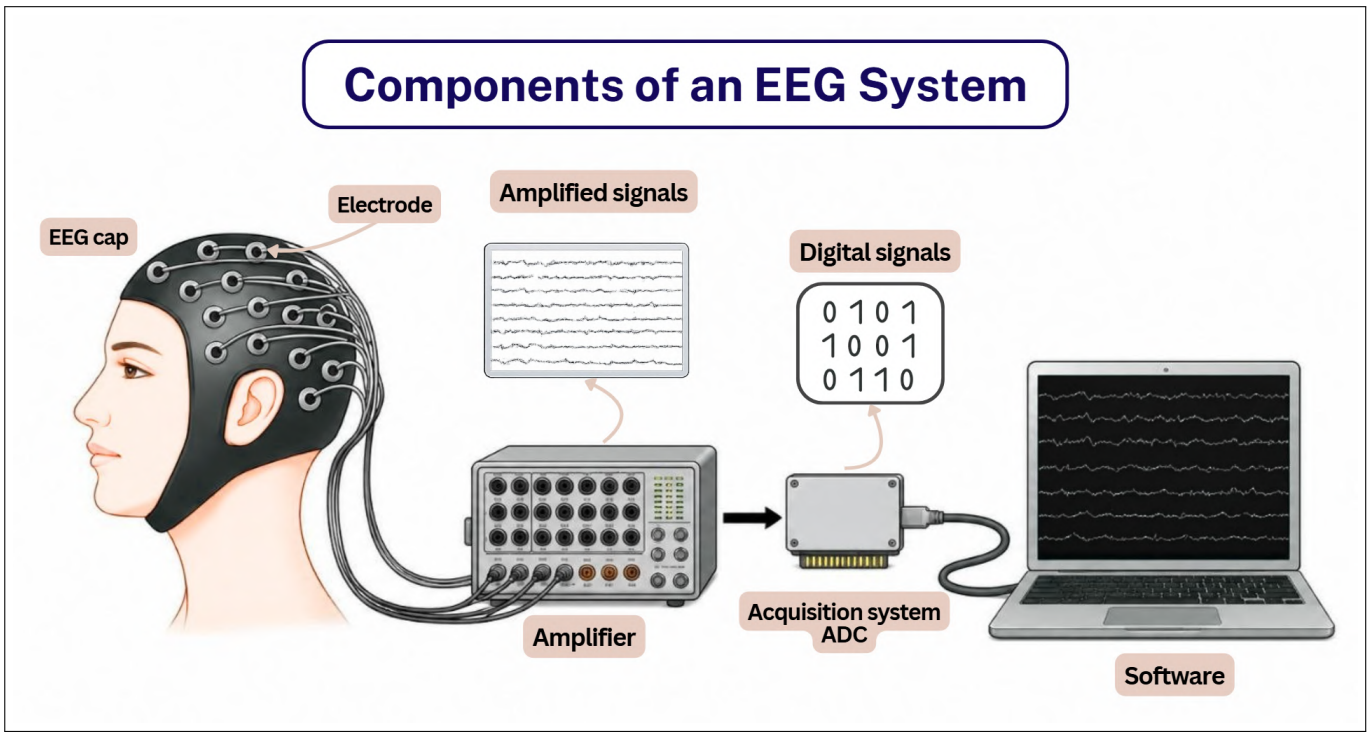


Figure 1.1: Basic EEG System Components.

1.2.3 EEG Electrodes and Distribution Systems

a. Role of Electrodes in the EEG System

The electrode is the fundamental component of an electroencephalography (EEG) system, serving as an interface between the scalp and the recording device. It plays a critical role in capturing the electrical signals generated by neurons and transmitting them safely and accurately to the EEG recorder.

The main functions of electrodes are as follows:

- Capture electrical signals representing voltage differences caused by neural activity in the cerebral cortex [**19**].
- Transmit signals to the amplifier for amplification and processing, as EEG signals usually have amplitudes in the microvolt (μV) range [**19**].
- Provide a safe and non-invasive method to record brain activity since electrodes are placed over the scalp [**19**].
- Deliver reliable EEG signals for the analysis of mental states, including concentration, mood, and cognitive processes [**19**].

b. Types of EEG Electrodes

EEG electrodes differ in design and the way they contact the scalp. The main types are:

- **Wet Electrodes:** These electrodes use a conductive gel or paste to bridge the gap between the electrode and the scalp, ensuring good contact and low impedance. Wet electrodes are considered the gold standard in clinical applications due to the high-quality EEG signals obtained, although they require longer preparation times [**1, 20**].
- **Semi-Dry Electrodes:** These electrodes use a smaller amount of electrolyte solution or a moist sponge to improve conductivity. They provide a balance between signal quality and ease of application compared to wet and dry electrodes [**1, 20**].
- **Dry Electrodes:** Dry electrodes do not require gels or pastes. They are designed to contact the scalp while minimizing impedance, making them quick and convenient to apply. However, they may have higher resistance and signal quality can be affected by head movements [**1, 20**].

All electrode types meet different application needs, and current research focuses on combining high signal quality with user comfort and convenience.

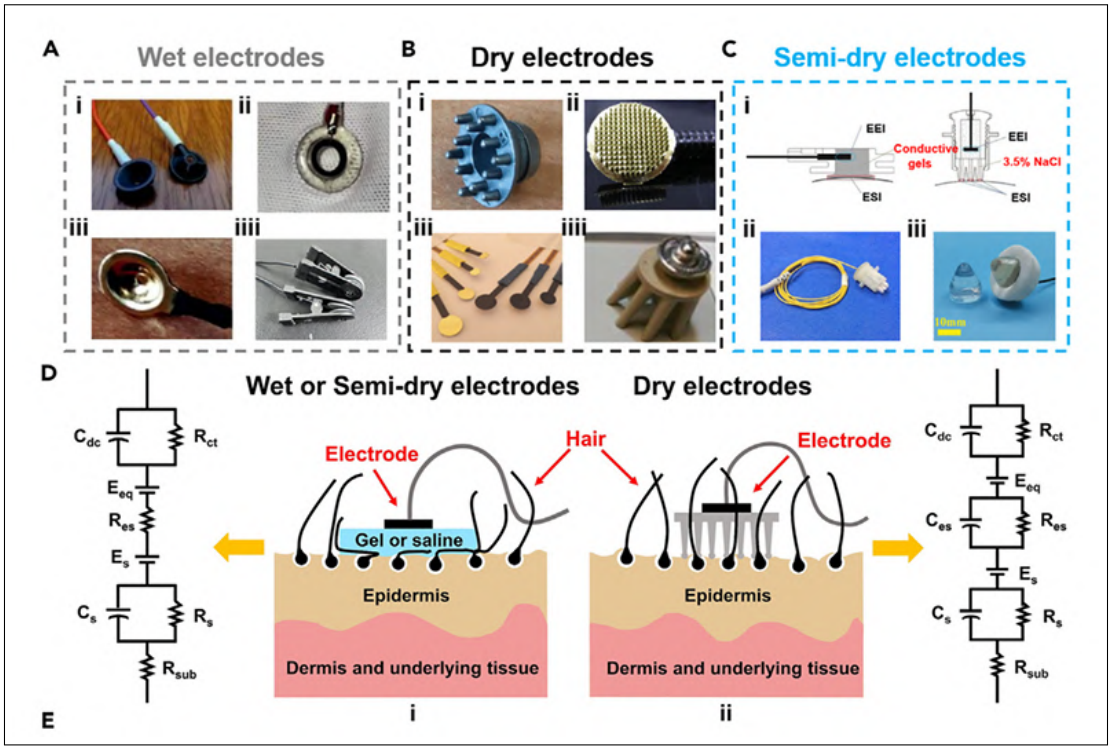


Figure 1.2: Classification of EEG Electrode Types [1].

c. Electrode Distribution Systems

Correct electrode placement is essential for accurate EEG recordings. Standardized placement systems define electrode positions on the scalp:

- **10–20 International System:**

  The most commonly used system, placing electrodes at precise proportions relative to head reference points. It allows acquisition of data from major brain regions, including frontal, parietal, temporal, and occipital areas [**21**].

- **10–10 System and Extended Systems:**

  These involve denser electrode placement at more precise locations, providing higher spatial resolution. They are often used in research or brain-computer interface (BCI) applications [**21**].

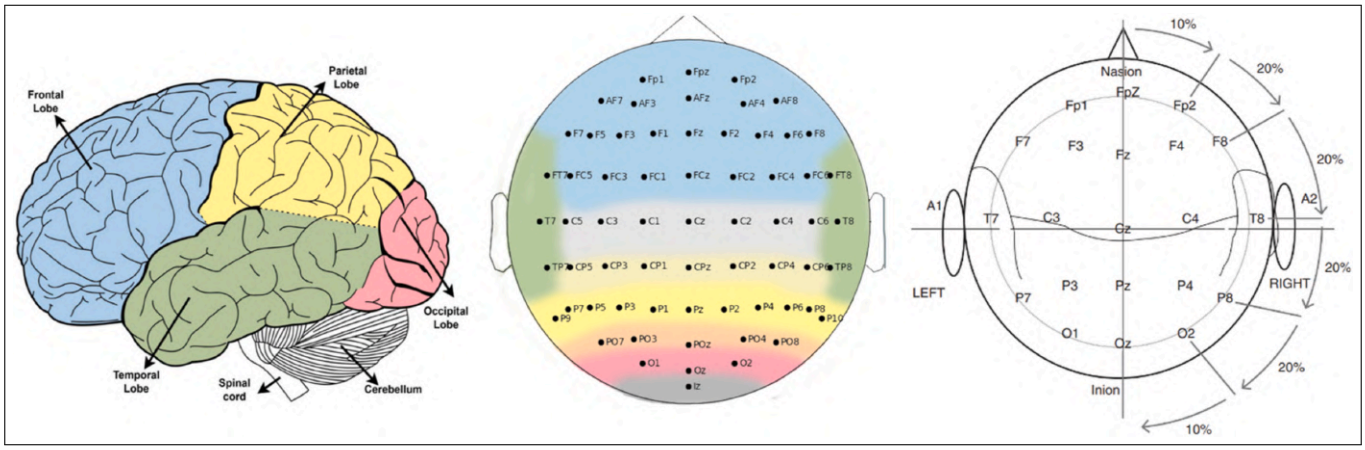


Figure 1.3: The International 10-20 System for EEG and Its Mapping [2].

1.2.4 How EEG Signals are Recorded?

- **Neural Activity:** EEG signals are produced by the synchronous activity of large populations of neurons, mainly in the cerebral cortex. In this region, multiple neural networks interact and generate electrical activity that can be measured and analyzed [**21**].
- **Origin of Electrical Brain Signals:** The electrical signals generated by neural activity propagate through the brain tissue and the skull until they reach the surface of the scalp. Due to this propagation, these signals can be recorded non-invasively using electrodes placed on the scalp [**21**].
- **Recording EEG from the Scalp:** Electrodes positioned on the scalp detect voltage differences resulting from neural activity. Since these signals are extremely small, typically in the microvolt range, they must be amplified before further processing or analysis. After amplification, the signals are converted from analog to digital form using an analog-to-digital converter (ADC). Finally, the digital signals are recorded and processed using specialized software for visualization, storage, and further analysis [**21**].

1.2.5 Characteristics of Electroencephalogram (EEG) Signals

EEG signals exhibit several important characteristics that make their analysis valuable for neurological and medical studies. However, these properties also introduce challenges in biomedical signal processing [**22**].

- **Low Amplitude:** EEG signals have very low amplitudes, typically in the microvolt (μV) range. This is due to the weak electrical activity generated by groups of neurons and the attenuation of the signal as it propagates through brain tissues, cerebrospinal fluid, and the skull before reaching the scalp [**22**].
- **Frequency Nature:** EEG signals are often analyzed in the frequency domain and are generally divided into several main frequency bands, including Delta, Theta, Alpha, Beta, and Gamma. Each frequency band is associated with specific physiological and cognitive states of brain activity [**22**].
- **Temporal Variability (Non-stationarity):** EEG signals are considered non-stationary, meaning that their statistical properties change over time. This temporal variability reflects the dynamic nature of brain activity influenced by cognitive processes, emotional states, and external stimuli [**22**].
- **Sensitivity to Noise and Artifacts:** EEG signals are highly sensitive to noise and artifacts originating from different sources such as eye movements (EOG), muscle activity (EMG), electrode displacement, and electrical interference from the surrounding environment [**22**].
- **Low Signal-to-Noise Ratio (SNR):** Due to the low amplitude of EEG signals and their susceptibility to artifacts, they generally exhibit a low signal-to-noise ratio (SNR). Therefore, preprocessing techniques such as filtering, artifact removal, and signal enhancement are usually required before advanced analysis [**22**].

- **Spatial Dependence:** EEG signals vary according to the locations of electrodes placed on the scalp. Each electrode captures the activity of specific brain regions, enabling the analysis of neural activity across different areas of the brain [**22**].

a. EEG Signal Frequency Bands

EEG (electroencephalography) signals are commonly divided into several main frequency bands, known as brain waves. Each band reflects specific neural activity and is associated with different physiological and cognitive states of the brain. This classification is widely used in neurological and clinical studies to investigate the relationship between the brain's electrical activity and behavioral or psychological states [**23**].

- **Delta Waves (0.5–4 Hz):** Delta waves are the slowest EEG waves and are primarily associated with deep sleep. They reflect a reduced level of consciousness and low neural responsiveness to external stimuli [**23**].
- **Theta Waves (4–7 Hz):** Theta waves typically appear during drowsiness or light sleep and are linked to certain cognitive processes, including memory and attention. They are often observed during the transition from wakefulness to sleep [**23**].
- **Alpha Waves (8–12 Hz):** Alpha waves are associated with mental relaxation and calm wakefulness. They usually appear when an individual is awake with eyes closed and tend to decrease in intensity during tasks requiring focus or mental effort [**23**].
- **Beta Waves (16–31 Hz):** Beta waves are linked to alertness, active thinking, and analytical processes. Beta activity increases during cognitive tasks or periods of heightened mental concentration [**23**].
- **Gamma Waves (36–90 Hz):** Gamma waves represent the highest frequency band in EEG signals. They are associated with higher-order cognitive processes, such as perception, information processing, and integration across different brain areas [**23**].

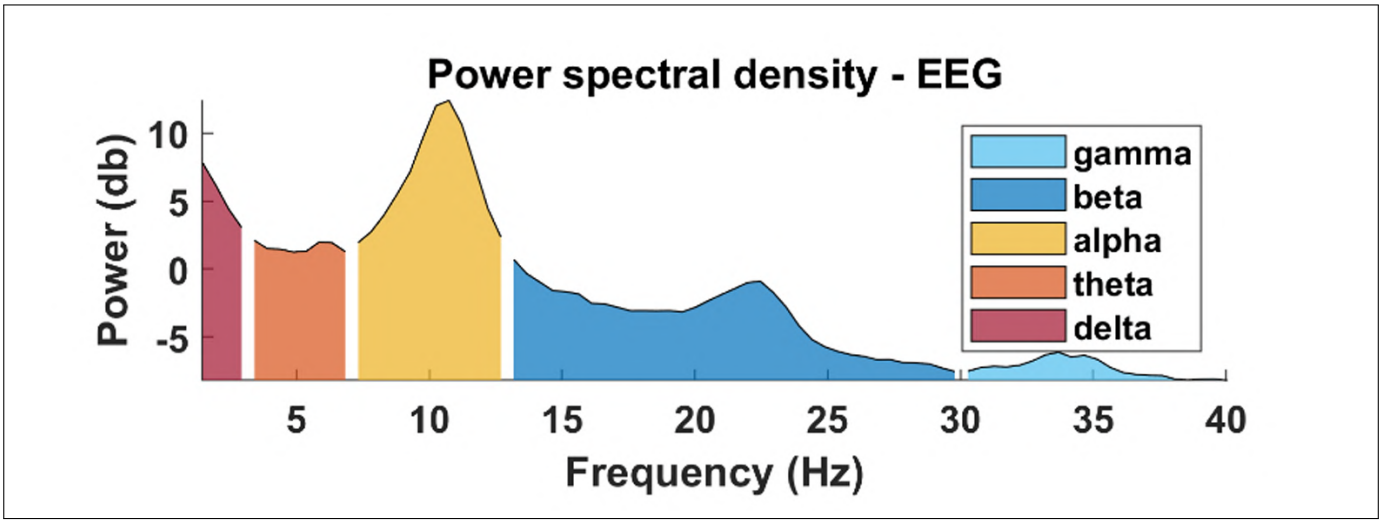


Figure 1.4: Power Spectral Density (PSD) Analysis and EEG Frequency Band Distribution [3].

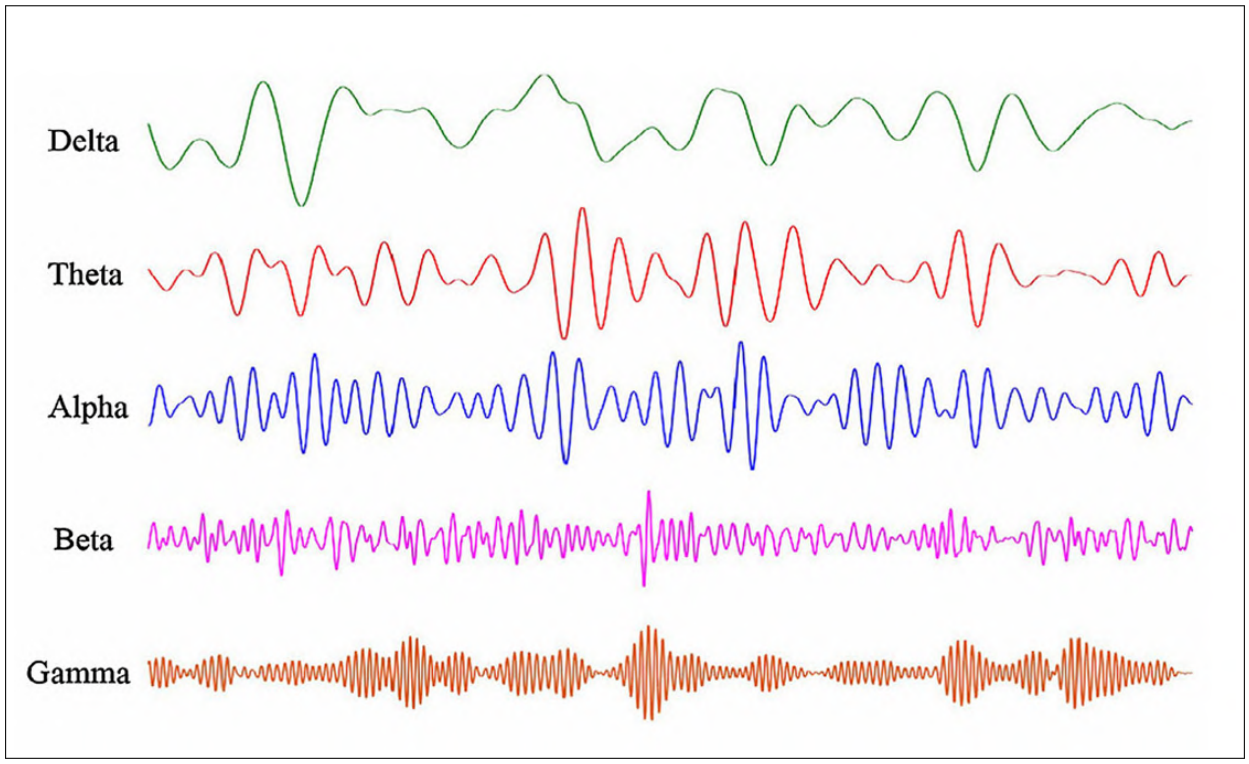


Figure 1.5: Comparison of Standard EEG Frequency Bands [4].

1.2.6 Types of EEG Signals

- **Time-Invariant EEG:** Time-invariant EEG refers to brain signals recorded when neural activity is relatively stable and not associated with specific external stimuli or events. These signals are typically used to evaluate the overall state of the brain and to analyze brain wave activity such as alpha, beta, and delta rhythms. Common examples include EEG recordings obtained when an individual is relaxed with eyes closed, during sleep, or while sitting quietly without performing a particular task. Such recordings are also useful in the clinical diagnosis of several neurological disorders [**18**].

- **Accurate Event-Related EEG:** Accurate event-related EEG refers to signals recorded in response to stimuli that occur at precisely known moments in time. These recordings are widely used to study cognitive processes such as attention and perception, as well as to evaluate neural responses to specific sensory inputs. For instance, EEG signals may be recorded when an image appears on a screen, when a sound is presented at a specific time, or when a participant presses a button in response to a cue. This category also includes event-related potentials (ERPs), such as the P300 component associated with the detection of significant stimuli, and the N200 component related to error processing [**18**].

- **Random Event-Related EEG:** Random event-related EEG refers to signals recorded in response to events that occur unexpectedly or at unpredictable times. These recordings are useful for studying how the brain naturally interacts with the surrounding environment and for analyzing emotional responses such as stress, fear, or joy. They are particularly relevant in interactive environments and virtual simulations, where EEG signals can be measured during gameplay or during spontaneous interaction with simulated environments [**18**].

1.2.7 Advantages, Disadvantages, and Challenges of EEG

a. Advantages

EEG is one of the most widely used methods for monitoring neural activity in both experimental and clinical settings due to several key benefits. It provides a non-invasive measure of neural activity, as EEG signals recorded from the scalp reflect synchronized electrical potentials generated by thousands of pyramidal neurons. Consequently, EEG can be safely applied across all age groups without surgical procedures and has no side effects.

Another advantage is its excellent temporal resolution in the millisecond range, allowing researchers to track rapid neural responses during specific cognitive or affective processes. EEG is also cost-effective compared to neuroimaging techniques such as MEG or fMRI and enables real-time monitoring of brain activity [**24**].

b. Disadvantages

Despite its advantages, EEG has inherent limitations that can affect the accuracy of certain studies. A major drawback is its poor spatial resolution since EEG measures activity only at the scalp surface, making it challenging to localize deep neural sources accurately.

EEG signals are also highly sensitive to artifacts caused by noise, including eye blinks, head movements, and electrical interference from the environment. Consequently, advanced signal processing techniques for noise reduction must be employed prior to analysis [**24**].

c. Challenges

Applying EEG in research and clinical contexts also presents several challenges. Handling non-linear and non-stationary signals is a primary concern, requiring proper filtering and feature extraction methods to maintain signal quality.

Classifying specific mental or psychological states from EEG data is another challenge. Neural patterns often overlap across different states, necessitating robust classification algorithms.

Finally, reproducibility across studies is an ongoing concern. Variations in equipment settings and electrode placement between studies make it difficult to reliably compare results [**24**].

1.2.8 General Applications of Electroencephalography (EEG)

Despite certain limitations and challenges, EEG remains a powerful tool in healthcare, offering practical applications that leverage its non-invasive and real-time monitoring capabilities. These capabilities allow EEG to be used across a wide range of healthcare scenarios [**25**].

a. Diagnosis of Neurological Disorders

EEG is extensively used in diagnosing neurological conditions such as epilepsy, Alzheimer's disease, and Parkinson's disease. It enables earlier detection and monitoring of disease progression over time [**25**].

b. Brain–Computer Interfaces (BCIs)

EEG signals can be converted into commands interpretable by electronically driven devices, including powered wheelchairs and assistive robots. This also enables individuals with physical impairments to interact and communicate with their environment [**25**].

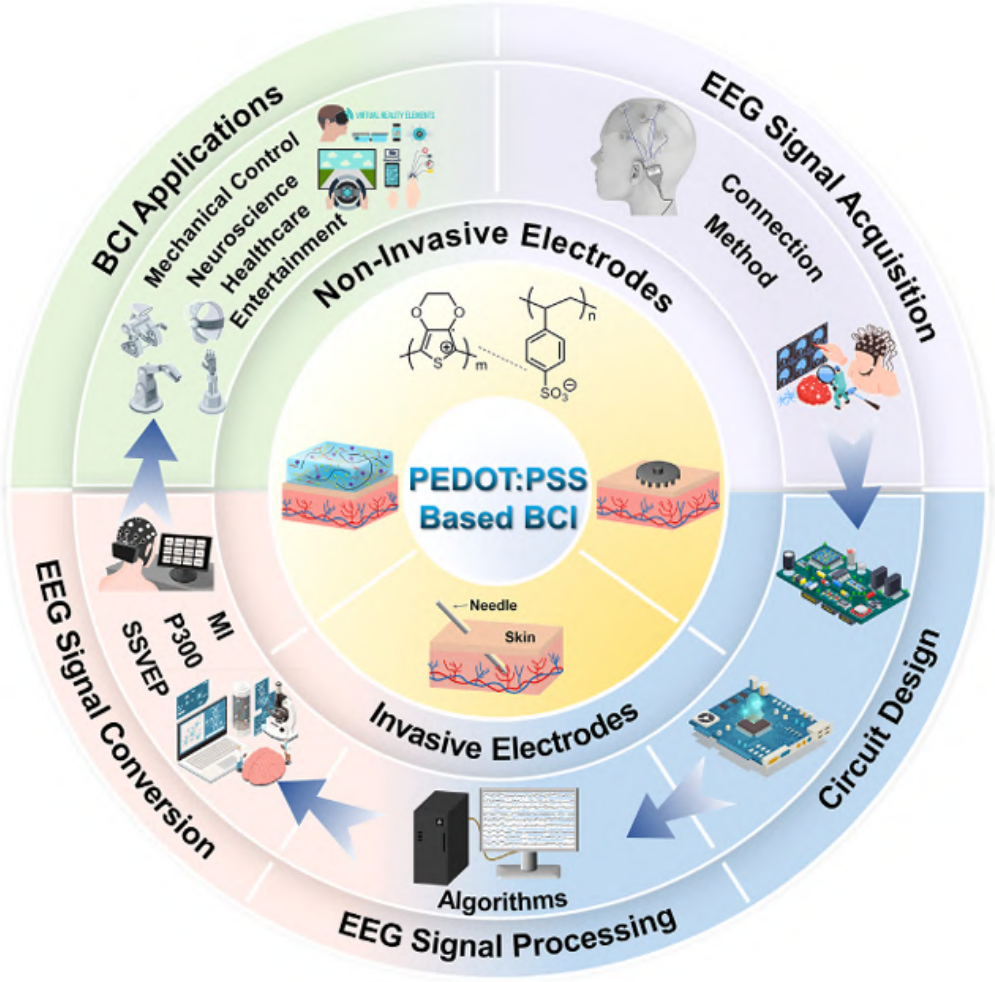


Figure 1.6: An Overview of BCI Systems: Components, Processing, and End-Uses [1].

c. Sleep and Fatigue Monitoring

EEG is utilized to assess various stages of sleep and diagnose sleep disorders such as insomnia and sleep apnea. Additionally, it can evaluate levels of fatigue and exhaustion, supporting preventative healthcare applications [**25**].

d. Neurorehabilitation and Therapy

EEG is applied in rehabilitation after stroke or spinal cord injury, assisting movement therapy and enhancing muscle function through assistive devices. This contributes to regaining motor abilities and improving quality of life [**25**].

e. Mental Health and Emotion Recognition

EEG can evaluate and monitor psychological conditions such as depression, anxiety, and schizophrenia. It also allows real-time assessment of emotional states, supporting correlations between physiological measures and human-computer interactions, and enabling targeted psychological interventions [**25**].

1.3 Emotion Recognition

1.3.1 What is Emotion Recognition?

a. Definition of Emotions

Emotions are multidimensional phenomena that emerge from the dynamic interaction between the body and the mind. Rather than being simple feelings, they reflect coordinated changes across physiological processes, cognitive evaluation, and behavioral reactions. These changes may become observable through physical cues or internal bodily responses. Emotions are generally elicited by external situations and can be communicated through different forms, including verbal expression, written language, or biological signals [**26**].

b. Emotion Recognition

Emotion recognition is the process of identifying and analyzing a person's emotional state using various data sources, including facial expressions, voice tone, gestures, text, and physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), and galvanic skin response (GSR). Automated emotion recognition systems employ artificial intelligence and machine learning techniques to extract features from these signals and predict emotional states accurately [**27**].

c. Importance of Emotion Recognition

Emotion recognition is important in several areas:

- **Marketing and Consumer Interaction:** Understanding emotions helps companies improve advertising responses and personalize offers [**27**].
- **Human-Robot Interaction:** Enables robots and intelligent systems to recognize the user's emotional state and interact naturally and effectively [**27**].
- **Healthcare and Mental Health Monitoring:** Supports the study and diagnosis of neurological and psychological disorders, such as sleep disorders, schizophrenia, depression, and Parkinson's disease, as well as detecting fatigue, drowsiness, and pain [**27**].
- **Study of Psychological and Behavioral Disorders:** Useful for conditions like autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), and panic disorders [**27**].
- **Development of Intelligent Systems:** Facilitates the design of systems capable of interacting with humans more naturally and effectively [**27**].

1.3.2 Emotion Recognition using EEG

Emotions naturally arise from interactions with events and situations in the environment and can be observed through behaviors such as facial expressions, tone of voice, and body movements [**28**]. In affective computing, emotions are commonly classified into basic categories including happiness, sadness, anger, disgust, surprise, and fear. They can also be represented using the Valence–Arousal model, which describes emotions along two dimensions: valence (degree of positivity or negativity) and arousal (level of activation) [**28**].

To objectively assess emotions, EEG is used alongside other physiological signals such as electrodermal activity (EDA), electrocardiogram (ECG), and electrooculogram (EOG). EEG is the most frequently used tool in emotion recognition research due to its ability to directly measure neural activity and involuntary responses, making it more objective than behavioral cues. Its high temporal resolution and ability to analyze spatial patterns across multiple electrodes allow efficient, real-time study and analysis of emotions [**28**].

In practice, researchers extract distinctive temporal and frequency-based features from recorded EEG signals. These features are then used as inputs for machine learning algorithms to associate them with different emotional states, enabling the development of automated emotion recognition systems [**28**].

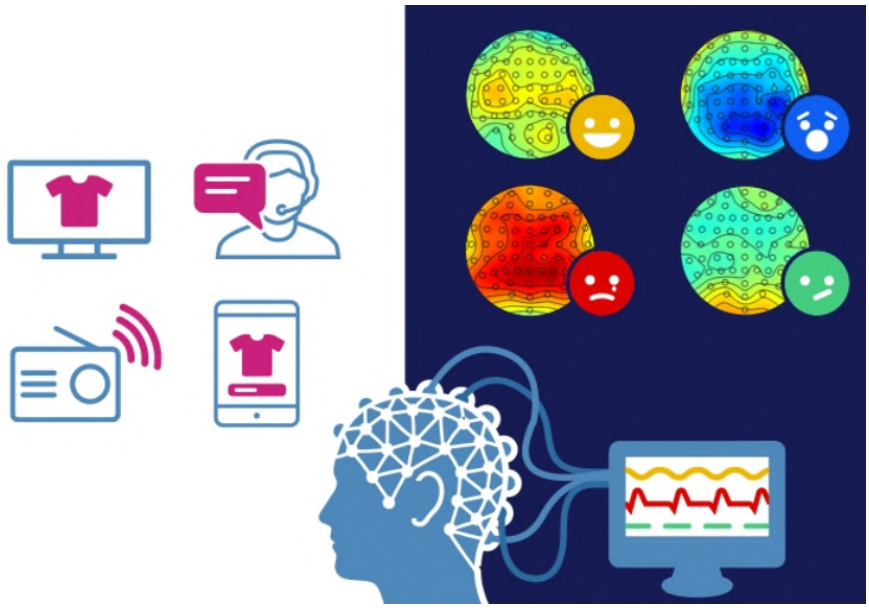


Figure 1.7: Neural Correlates of Different Emotional Responses.[1]

1.3.3 Computer Games as Emotion Elicitation Environments

Video games provide an effective interactive medium for eliciting affective responses from individuals. Through interaction with game events, decision-making, and varying levels of difficulty, players experience emotions that can be captured and analyzed. By employing EEG as a non-invasive tool to measure brain activity in real-time, researchers can obtain accurate and objective insights into the emotional states of gamers [**28**].

[1] `https://www.rsm.nl/discovery/2019/emotions-brain/`

EEG signals allow analysis of neural activity corresponding to emotions during gameplay by examining the characteristics of different frequency bands and patterns of brain activity. This approach helps researchers understand the emotional experiences of video gamers and supports the development of intelligent video games that can adapt to players' emotional states, enhancing immersion and interaction within the gaming environment [**28**].

1.4 Relationship between EEG Signals, Emotions, and Computer Games

Electroencephalogram (EEG) signals are linked to emotions experienced during gameplay in computer games, as the emotional experiences of the player directly reflect neural activity in the brain. When a player interacts with interactive game environments, multiple cognitive and emotional processes are activated, such as attention, excitement, stress, and concentration, leading to measurable changes in EEG signals. This provides an objective means for analyzing emotional states [**29**].

Computer games are considered an ideal environment for studying emotions using EEG due to their interactive and dynamic nature. They rely on continuous interaction, decision-making, and immediate feedback, which generates realistic emotional responses that are reflected in brain activity in real time. Incorporating EEG-based emotion recognition systems also enables the development of intelligent games that can adapt to the player's emotional state by adjusting the difficulty level, game content, or pace of events, thereby enhancing immersion, reducing frustration, and improving the user experience [**29**].

In general, the integration of EEG signals, emotion recognition, and computer games creates a comprehensive framework in the field of affective computing. Here, games are used not only to elicit emotions, but also as a platform for analyzing them and developing intelligent human-computer interaction systems, such as adaptive games and serious games [**29**].



Figure 1.8: Experimental Setup for EEG-Based Emotion Recognition during Gaming.[2]

[2] https://www.nature.com/articles/d42473-022-00218-7

1.5 Conclusion

In light of what has been presented in this chapter, it is clear that EEG signals are an important source for understanding emotional neural activity, owing to their ability to directly and accurately reflect brain changes. Furthermore, studying and analyzing these signals provides a strong foundation for developing intelligent systems capable of recognizing human emotional states.

Accordingly, this chapter serves as an important theoretical introduction to the following chapters by clarifying fundamental concepts in EEG and establishing an initial understanding of the relationship between neural signals and emotions, thereby supporting the transition to the study's practical and methodological aspects.

Literature Review of EEG-Based Emotion Recognition

2.1 Introduction

Accurately identifying human emotional states remains a significant challenge in neuroscience and artificial intelligence, particularly when relying solely on facial expressions or behavioral cues, which may not reliably reflect an individual's internal state. This has motivated the development of more objective and physiological approaches for emotion assessment.

Among these approaches, electroencephalogram (EEG) signals have emerged as a promising modality for analyzing emotion-related neural activity, as they enable direct observation of brain dynamics with reduced subjectivity compared to conventional methods. Consequently, EEG-based emotion recognition has attracted increasing attention in recent research.

Existing studies in this domain present a wide range of methodologies, spanning signal preprocessing, feature extraction, and the application of machine learning and deep learning models for classification. However, several challenges remain, including signal noise and artifacts, inter-subject variability, and the difficulty of developing models that generalize effectively across different individuals and experimental settings. Given this diversity, this chapter provides an analytical overview of the principal approaches employed in EEG-based emotion recognition.

2.2 EEG Signal Acquisition Methods

In order to build any system capable of understanding human emotions based on brain signals, it is first necessary to provide real data that represents neural activity in different states. Therefore, the data acquisition process is an essential and indispensable step in this type of study, as it involves recording the initial EEG signals that will later be used for analysis and feature extraction. This process takes place through a series of sequential, methodical stages aimed at ensuring data quality, starting from designing the experiment and selecting participants, to recording the signals under controlled conditions to ensure precision and stability [7].

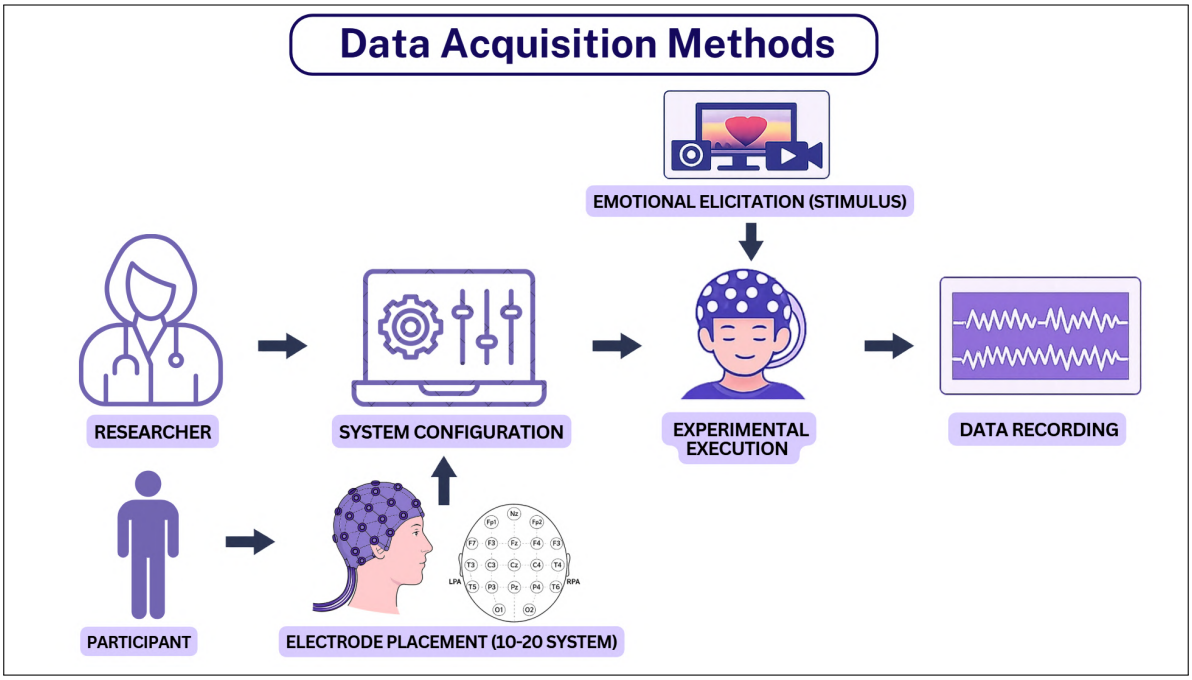


Figure 2.1: Illustration of the Baseline EEG Data Acquisition Framework.

2.2.1 Electrode Placement and Configuration

The data acquisition process begins by attaching electrodes to the scalp according to internationally recognized systems, such as the 10-20 system, which ensures systematic coverage of different brain regions associated with emotional responses, particularly the frontal lobe and the temporal lobe.

Once the electrodes are positioned, the recording system settings are adjusted according to technical standards, which may vary depending on the device used or the adopted database, such as the number of channels or sampling frequency, to ensure the acquisition of precise EEG signals suitable for subsequent analysis stages [**7**].

2.2.2 Experimental Protocol and Stimulus Environment

After preparing the recording system, the process moves to the implementation of the experimental protocol, where emotions are elicited in participants using a variety of stimuli.

These stimuli include emotionally charged images, video clips used to provoke more pronounced emotional responses, or audio clips presented in a calm environment to obtain varied and more stable emotional responses [**7**].

2.2.3 Signal Recording and Data Quality Control

Once the recording system is set up and the experimental conditions are established, the process moves to recording the EEG signals within a controlled laboratory environment. Signals are captured during specific time intervals, reflecting participants' responses to different stimuli.

Throughout this process, recording conditions are continuously monitored to ensure that the recorded signals accurately represent genuine brain activity.

This careful organization helps provide stable raw data, forming the basis for subsequent processing and feature extraction stages [**7**].

2.3 Annotation for EEG-Based Emotion Recognition

Annotation is a crucial step in emotion recognition research based on EEG recordings, through which emotional tags are assigned to each signal segment. Thus, machine learning algorithms can be trained to recognize emotional patterns with high accuracy.

Most researchers use the Valence-Arousal (VA) model to represent human emotions in terms of two main components:

- **Valence (pleasant versus unpleasant):** Measures the affective tone of the emotion, whether it is a positive one like happiness or a negative one like sadness, and demonstrates how different emotions are processed by the brain [**30**].
- **Arousal (energy or activation):** Measures the degree of emotional intensity, whether it is calmness (low) or excitement/fear (high), and shows the changes in brain activity levels in response to the emotion [**30**].

In addition to the VA model, some studies propose extending the emotional space with additional dimensions:

- **Liking (preference/acceptance):** Indicates the extent to which an individual is satisfied with or attracted to an emotional stimulus, thus distinguishing between two emotions that are equal in valence and arousal but differ in personal liking [**30**].
- **Dominance (power/influence):** Indicates the individual's level of control or submission during the emotional moment, which helps in mapping more precisely the brain's reaction to the emotional stimulus. To illustrate, fear is normally characterized by low dominance, while happiness is perceived as high dominance [**30**].

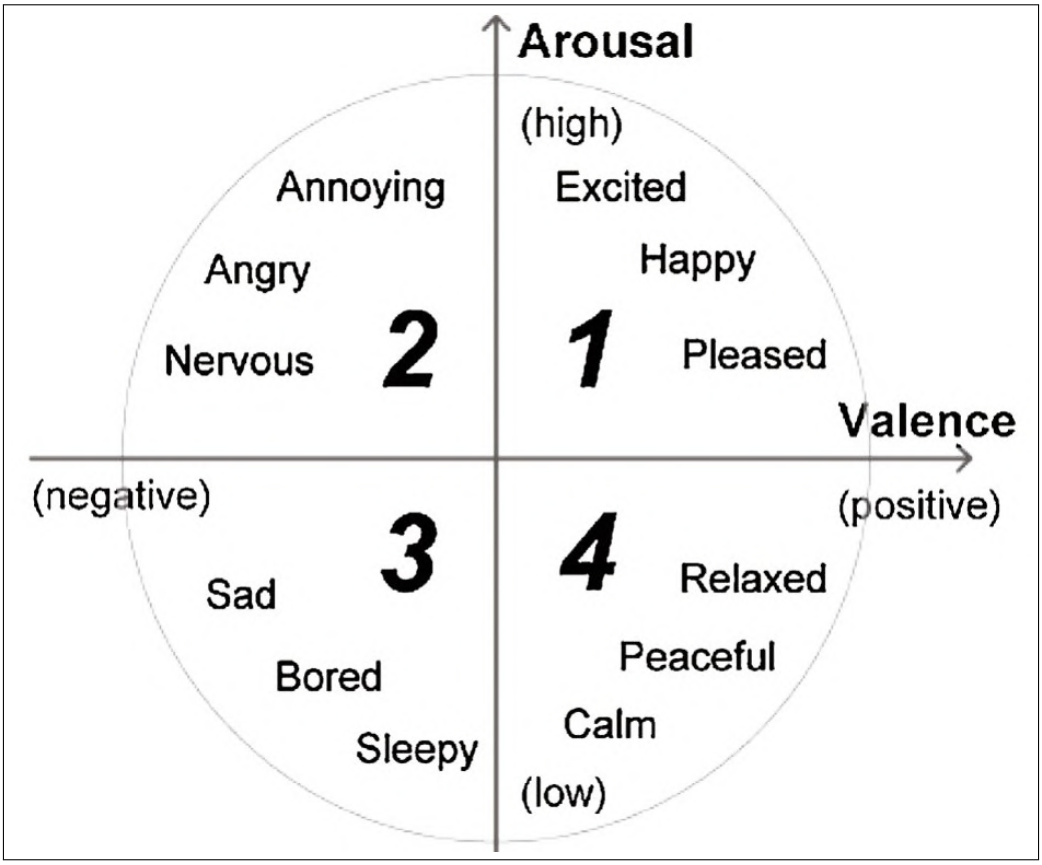


Figure 2.2: The 2D Valence-Arousal Space for Dimensional Emotion Representation [5].

One of the most common methods for obtaining these labels is self-assessment using the Self-Assessment Manikin (SAM) scale, where participants evaluate the dimensions of emotion (valence, arousal, liking, and dominance) on a Likert scale ranging from 1 to 9.

Low values (1–3) indicate a weak or negative level, medium values (4–6) indicate a moderate state, while high values (7–9) represent a strong or positive emotional level.

These values are then converted into categories used in classification models (e.g., high or low valence) [**30**].

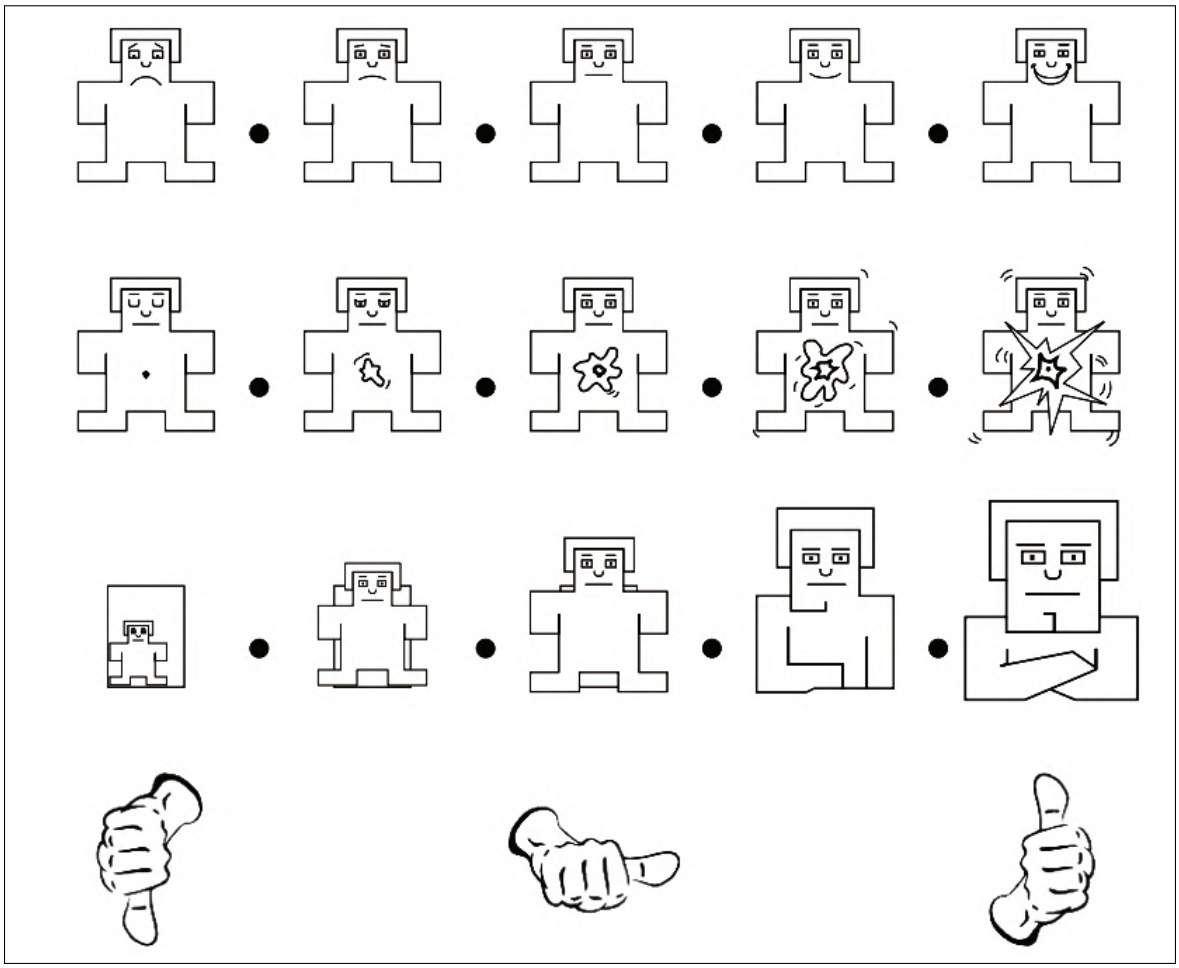


Figure 2.3: Self-Assessment Manikin (SAM) Scales for Subjective Emotion Labeling [6].

On the other hand, some studies rely on pre-classified emotional stimuli (such as video clips), where the same emotional label is assigned to EEG signals recorded during the presentation of the stimulus without depending on self-assessment, which helps simplify the labeling process even though it may not always reflect individual differences in emotional response [**31**].

Some advanced research also adopts a multi-perspective approach to labeling, in which multiple sources of evaluation are combined, such as self-assessment, partner assessment, and external observer comments, with the aim of reducing bias and improving data reliability [**31**].

This approach is supported by the use of psychological models and tools, such as the Big Five Personality Traits to identify personality characteristics, and the PANAS scale to assess basic mood states, which contributes to providing a more accurate and comprehensive interpretation of emotional responses [**32, 33**].

2.4 Datasets for EEG-Based Emotion Recognition

Research on EEG-based emotion recognition commonly relies on several publicly available datasets for training and evaluating machine learning models. The most widely used datasets include:

Table 2.1: Comparison of Common EEG Datasets for Emotion Recognition.

Dataset	Subjects	Duration	Stimulation	Channels	Annotation
DEAP [**6**]	32	40 × 1 min	Music videos	32	SAM (1–9)
SEED [**34**]	15	∼4 min	Movie clips (15)	62	Pre-labeled emotions
SEED-IV [**34**]	15	∼2 min	Movie clips (72)	62	Pre-labeled emotions
SEED-V [**34**]	16	∼2–4 min	Movie clips (45)	62	Pre-labeled emotions
GAMEEMO [**35**]	28	20 min	Gameplay	14	SAM (1–9)
EEG Gaming Dataset [**29**]	12	2.5 h	2D space game	66	SAM (1–9) / Event states
Graz data set A [**36**]	9	6 s × 2 runs	Tasks/ Neutral/ visual cues/ warning tone	22	Task-based labels
MAHNOB-HCI [**37**]	30	40 min	Video clips	32	SAM
FACED [**38**]	123	86–129 s	Video clips	32	SAM (0–7)
K-EmoCon [**33**]	32	2 h	Debate	N/A	Self / partner / 5 external observers
MPED [**39**]	23	Varies	Video stimuli (28)	62	Psychological methods
PhyMER [**40**]	30	Varies	Video	N/A	SAM
SAFE [**41**]	6	Varies	Audio (IADS)	5	SAM
ASCERTAIN [**42**]	58	Varies	Movie clips	N/A	Pre-labeled emotions
AMIGOS [**32**]	40 / 37	20 min / 86 min	Short & long videos	14	SAM / Psychological profiling(PANAS,Big-Five)

2.5 Signal Preprocessing for EEG Signals

Preprocessing is an essential stage that precedes the analysis of EEG signals. Its purpose is to purify the raw signals from various forms of noise and interference introduced during recording, whether physiological (e.g., eye and muscle movements) or environmental and technical.

Since these signals are not directly suitable for analysis or classification, this stage helps improve their quality and ensure the reliability of the results obtained later. According to the literature, various signal processing techniques are employed to preprocess EEG signals [**43**].

Table 2.2: Signal Preprocessing for EEG Signals.

Methods	**Description**	**Techniques Used**
Filtering	This method cleans and enhances signal quality by removing noise and interference caused by unwanted frequencies. This is achieved using specialized filters to isolate specific frequency bands such as Alpha, Beta, Gamma, and Theta, which are associated with mental and cognitive activity [**44**].	**Band-pass Filter:** Retains a specific frequency range from the EEG signal, such as 0.3–50 Hz [**31**]. **Median Filter:** Smooths the signal by removing outlier values or sharp peaks that may appear in it [**45**]. **Notch Filter:** Eliminates interference caused by mains power frequency (usually 50 Hz) [**45**]. **Moving Average Filter:** Removes baseline drift caused by poor electrode contact and smooths values while reducing fluctuations [**45**]. **5th order Sinc Filter:** Analyzes the signal in the frequency domain using a fifth-order Sinc function [**5**].
Segmentation	This process divides continuous EEG signals into small, equal time segments known as windows. It helps extract the signal's dynamic features over time, with each window analyzed independently rather than the entire signal being processed at once [**44**].	**Sliding Window:** Extracts consecutive segments by sliding a time window of a specified length over the EEG signal, allowing analysis over different time intervals [**46**]. **Non-overlapping Windows:** Segment the signal into consecutive time windows without any overlap, ensuring each window represents an independent part of the signal [**46**]. **Overlapping Windows:** Segment the EEG signal into consecutive time windows with overlap between adjacent windows, allowing continuity of information between segments and improving the representation of signal variations over time [**30**].

Table 2.2: Signal Preprocessing for EEG Signals.

Methods	Description	Techniques Used
Normalization	This process rescales signal values to a uniform and specified range to reduce differences between values and improve their comparability, enhancing the performance of machine learning models and data analysis [**34**].	**Min-Max Scaling:** Rescales the EEG signal values to a range between 0 and 1, where the minimum value becomes 0 and the maximum becomes 1 [**34**]. **Z-score (Standardization):** Centers the signal values around zero by subtracting the mean from each value and dividing by the standard deviation [**46**].
Signal Decomposition	This process analyzes the complex EEG signal by breaking it down into a set of simpler fundamental components. It aims to represent the internal characteristics of the signal more accurately and facilitate feature extraction [**47**].	**VMD (Variational Mode Decomposition):** Decomposes the EEG signal into multiple intrinsic mode functions (IMFs) using an optimization framework, where each component is separated within a specific frequency range [**47**]. **EMD (Empirical Mode Decomposition):** Decomposes the signal based on its local characteristics to extract intrinsic mode functions (IMFs) without using a predefined mathematical model [**47**].
Artifact Removal	This procedure purifies EEG signals by removing non-cerebral activities such as eye movement (EOG) and muscle activity (EMG) [**44**].	**Independent Component Analysis (ICA):** Decomposes the signal into its fundamental components to isolate unwanted components and obtain a cleaner and more accurate brain signal [**44**]. **Manual Artifact Removal:** Inspects EEG signals manually to identify and remove noisy or unwanted components, ensuring that the resulting data is clean and suitable for further analysis [**31**].
Downsampling	This method reduces the amount of data by decreasing the number of samples recorded per second. It helps speed up computation and signal analysis while preserving essential signal characteristics when an appropriate reduction rate is applied [**44**].	**Resampling:** Reduces the EEG signal sampling rate from 1000 Hz to 200 Hz [**34**]. **EEGLAB Toolbox:** Performs resampling of the original signal at regular time intervals to generate a representation with fewer samples [**30**].

Table 2.2: Signal Preprocessing for EEG Signals.

Methods	Description	Techniques Used
Smoothing	This process applies feature smoothing after feature extraction to reduce random fluctuations in the data and improve its stability [44].	**Moving Average:** Calculates the average of values within a sliding time window that moves across the data, then replaces each value with this average to reduce noise and smooth the signal [44]. **Linear Dynamical Systems (LDS):** Models the signal over time and estimates its underlying state, separating useful information from noise using a mathematical dynamic model [44].

2.6 Feature Extraction for EEG Signals

The process of feature extraction from electroencephalogram (EEG) signals is a pivotal stage in analysis and classification systems, wherein the raw signal is transformed into a compressed, more meaningful numerical representation, which facilitates algorithmic processing.

This stage reduces the complexity of the signal while preserving the essential information that reflects crucial neural patterns. As a result, the performance, accuracy, and reliability of classification models are positively impacted.

Given the non-stationary and multidimensional nature of EEG signals, and the distribution of their information across both time and frequency, various methods are employed for feature extraction, with each method focusing on a specific aspect of signal representation.

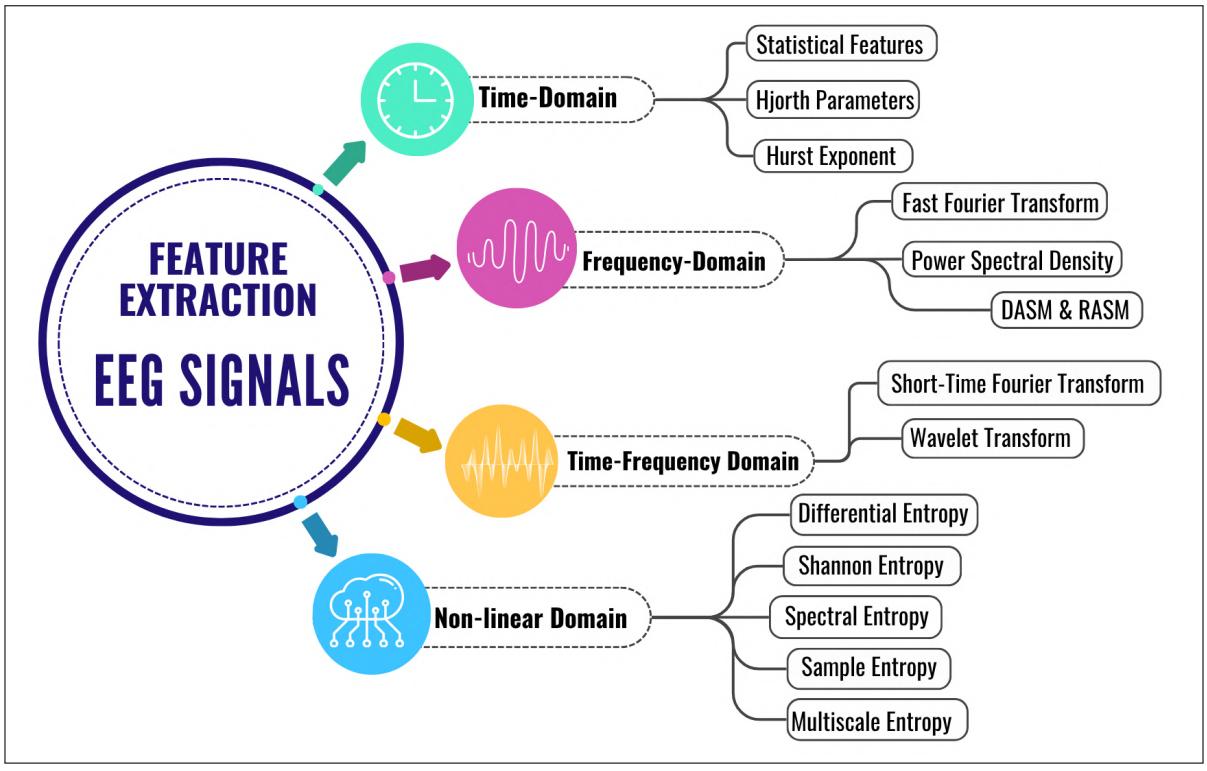


Figure 2.4: Overview of Feature Extraction for EEG Signals.

2.6.1 Time-Domain Features

The Time Domain focuses on analyzing EEG signals in their original form, considering them as a function of time without any transformation into other domains. This approach is used to study the direct changes in signal values over time by tracking their behavior as recorded, in order to extract information that reflects the instantaneous dynamics of brain activity.

Table 2.3: EEG Time-Domain Feature Extraction Methods and Objectives.

Feature Category	**Description**	**Scientific Objective**
Statistical Features [**44**]	A set of characteristics used to describe the distribution of EEG signal values over time, based on statistical measures such as mean, variance, and standard deviation, providing a general representation of the signal's behavior.	Capture and analyze variability patterns in EEG signals to summarize temporal changes.
Hjorth Parameters [**43**]	A set of statistical features in the time domain of EEG signals, consisting of three main indicators: Activity, Mobility, and Complexity, used to describe the signal's dynamic characteristics in a simplified manner.	Quantify EEG signal dynamics using activity, mobility, and complexity descriptors.
Hurst Exponent [**48**]	A statistical measure used in the analysis of EEG time series to quantify the degree of persistence in the signal over time by studying the effect of previous values on subsequent ones (long-term memory).	Measure signal persistence to identify long-range temporal dependencies in EEG time series.

2.6.2 Frequency-Domain Features

The Frequency Domain represents the signal using its frequency components rather than its time representation. This approach enables the analysis of the distribution of signal energy across different frequencies by studying its spectral components and identifying prominent frequency bands, allowing for a more precise and clear description of its frequency content.

Table 2.4: EEG Frequency-Domain Feature Extraction Methods and Objectives.

Feature Category	Description	Scientific Objective
Fast Fourier Transform (FFT) [43]	A mathematical transformation used to convert an EEG signal from the time domain to the frequency domain, allowing for the analysis of its frequency components, such as Delta, Alpha, and Beta, and understanding its spectral properties.	Analyze and extract the frequency components of EEG signals by transforming them from the time domain to the frequency domain.
Power Spectral Density (PSD) [44]	A metric used to estimate how the energy of an EEG signal is distributed across various frequencies, reflecting the signal's intensity within each frequency band.	Quantify spectral energy distribution across EEG frequency bands.
DASM & RASM [44]	A group of features used to analyze asymmetry between the two hemispheres in EEG signals. DASM represents the difference in activity between the right and left sides, while RASM represents the ratio of this activity between them, highlighting spatial differences associated with brain responses.	Identify hemispheric asymmetry patterns to better understand emotional and behavioral responses.

2.6.3 Time-Frequency Domain Features

The Time–Frequency domain represents signals by combining temporal and frequency information simultaneously, allowing the tracking of frequency evolution over time through instantaneous variations. This provides a more comprehensive representation of dynamic signals compared to time-domain or frequency-domain representations alone.

Table 2.5: EEG Time-Frequency Domain Feature Extraction Methods and Objectives.

Feature	Description	Scientific Objective
Short-Time Fourier Transform (STFT) [44]	A technique for analyzing EEG signals in both the time and frequency domains. This is achieved by dividing the signal into short time windows and then applying the Fast Fourier Transform (FFT) to each window to track frequency changes over time.	Analyze non-stationary EEG signals by tracking frequency distribution across consecutive time windows.
Wavelet Transform [44]	An advanced analysis method used to decompose EEG signals into components at different time scales, enabling the analysis of rapid temporal changes while preserving frequency information. Unlike the STFT, it adapts to signal characteristics by using variable analysis windows, providing high temporal resolution for rapid changes and better frequency resolution for stable regions, making it particularly suitable for analyzing non-stationary EEG signals.	Analyze time–frequency characteristics of non-stationary EEG signals using multi-scale representation.

2.6.4 Non-Linear Domain Features

The Non-Linear domain analyzes EEG signals based on their Non-Linear properties to capture their complexity and irregularity. It examines temporal change patterns and extracts indicators that reflect the signal's underlying dynamical behavior, which cannot be described using conventional linear methods.

Table 2.6: EEG Non-Linear Domain Feature Extraction Methods and Objectives.

Feature	Description	Scientific Objective
Differential Entropy (DE) [44]	A measure used to express the degree of randomness or irregularity in EEG signals, where higher values are associated with increased complexity of brain activity and its fluctuation between different emotional states.	Quantify EEG signal complexity by measuring variability in amplitude distribution and use it to distinguish emotional states.

Table 2.6: EEG Non-Linear Domain Feature Extraction Methods and Objectives.

Feature	Description	Scientific Objective
Shannon Entropy [5]	A quantitative measure based on the probability distribution of EEG signal values, used to estimate the amount of information or how difficult it is to predict these values. Higher values indicate increased randomness and irregularity in brain activity across different emotional states.	Estimate information content of EEG signals based on amplitude probability distribution and level of uncertainty.
Logarithmic Energy Entropy [5]	A measure based on the energy of the EEG signal in the frequency domain. It describes how this energy is distributed across different frequencies in a logarithmic form, helping to reduce the impact of large values and highlight subtle differences between frequency components.	Analyze EEG signal energy distribution in the frequency domain using a logarithmic representation to highlight spectral differences.
Spectral Entropy [35]	A measure used to assess the dispersion of energy in an EEG signal across the frequency domain, based on how energy is distributed among different frequencies. The more evenly the energy is distributed among frequencies, the higher the spectral entropy.	Measure the uniformity of energy distribution across frequency bands to assess brain activity regularity or disorder.
Sample Entropy [5]	A measure used to assess the degree of complexity and regularity of a signal by evaluating the likelihood of pattern repetition within the time series. Higher values indicate more difficulty in predicting the signal's behavior.	Evaluate EEG signal regularity by measuring the likelihood of repeating temporal patterns.
Multiscale Entropy [5]	An extension of Sample Entropy used to analyze the complexity of the EEG signal across multiple temporal scales. The signal is reconstructed at various temporal levels, and then Sample Entropy is calculated for each level. This allows the study of pattern changes over both short and long time intervals.	Analyze EEG complexity across multiple temporal scales to capture short and long-term dynamics.

2.7 Data Augmentation for EEG Signals

Data scarcity is considered one of the main challenges facing machine learning models, especially in the field of electroencephalogram (EEG) signal analysis. This shortage can lead to the phenomenon of overfitting, a situation in which the model learns the details and noise of the training data instead of extracting generalizable patterns.

To address this issue and enhance the generalization ability of models, data augmentation techniques are employed. These techniques aim to expand the training set either by generating modified versions of the original data or by creating synthetic samples that mimic their statistical and spectral properties.

The following table presents the most important techniques used in this context:

Table 2.7: EEG Data Augmentation Strategies.

Method Category	**Technique**	**Description**	**Scientific Objective**
Re-sampling Technique	SMOTE (Borderline-SMOTE) [**49**]	A technique used to generate synthetic samples for underrepresented classes based on their nearest neighbors, allowing for increased representation of these classes within the data.	Improve class balance and enhance model performance on underrepresented classes.
Signal Transformation	Random Channels Exchange (RCE) [**49**]	A technique that relies on randomly re-arranging or exchanging EEG channels within the same sample, creating new variations of the original signal.	Increase EEG data diversity while preserving the overall signal structure.
Generative Models	GANs & VAEs [**49**]	A technique that uses generative neural networks to learn the statistical distribution of the original data and then generate synthetic EEG signals that resemble real data.	Generate realistic synthetic EEG samples to expand the dataset and address data scarcity.
Noise Injection	Gaussian Noise [**30**]	A type of statistical noise randomly added to the data according to the normal distribution, with the aim of generating samples similar to the original data with slight variations.	Enhance EEG data diversity and improve model generalization to signal variations and noise.

2.8 Background of Machine Learning and Deep Learning for EEG Analysis

2.8.1 Traditional Machine Learning Models

a. Linear and Probabilistic Models

- **Logistic Regression** [**50**] is a probabilistic classification algorithm that estimates class probabilities from numerical features and assigns each sample to the category with the highest probability using a decision boundary. It can be used for both binary and multiclass classification tasks [**51**].

  In EEG signals, it is used to classify neural signals based on features extracted from the recorded signals, enabling the identification of the most likely emotional state for each signal.

- **Naive Bayes** [**52**] is a probabilistic classification model that assumes the independence of features within each class and calculates the contribution of each feature separately when estimating probabilities. It is characterized by its effectiveness with high-dimensional data, as well as its simplicity and speed, while achieving good performance in many applications [**51**].

  In EEG signals, it is used to classify emotional states based on features extracted from the recorded signals, relying on probabilistic inference to assign each signal to its corresponding class.

- **Linear Discriminant Analysis (LDA)** is a statistical classification method that aims to distinguish between two or more groups based on multiple features. It constructs a linear combination of the input variables to maximize class separation while minimizing within-class variance [**53**].

  In EEG signals, it is used to transform extracted features into a lower-dimensional space that maximizes the separation between different states, improving classification performance [**54**].

b. Distance and Margin-Based Models

- **Support Vector Machines (SVM)** [**55**] is a supervised non-linear classification algorithm used to handle data that are not linearly separable. It finds the optimal hyperplane with the largest margin between categories in a higher-dimensional feature space, as illustrated in Figure **2.5** [**7**].

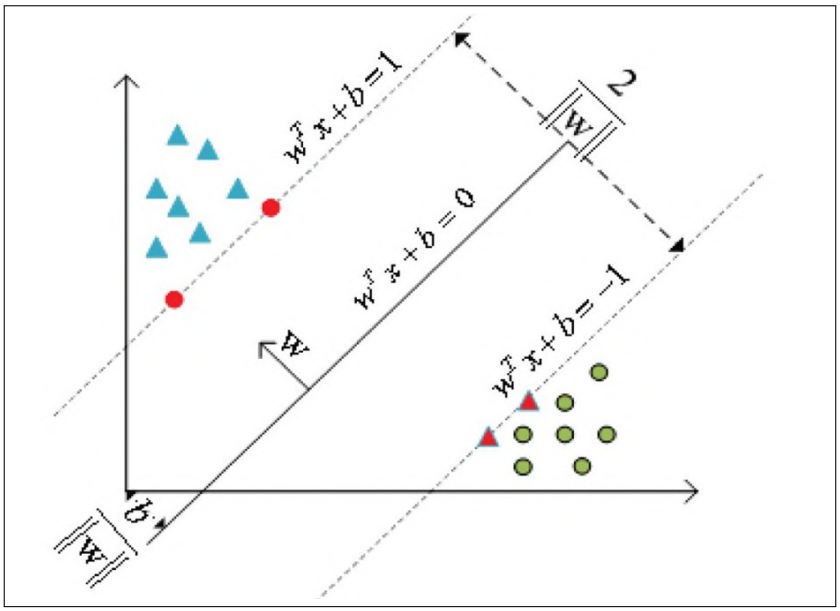


Figure 2.5: Illustration of How the Support Vector Machines (SVM) Algorithm Works [7].

In EEG signals, it is used to classify brain activity based on features extracted from multiple channels, assigning each signal to the appropriate emotional state according to its position relative to the separating hyperplane [**51**].

- **k-Nearest Neighbors (k-NN)** [**56**] is a supervised learning method where a test sample is classified based on its nearest training samples using distance measures. The classification is obtained by selecting the k closest samples after sorting distances, and assigning the sample to the most frequent class among them. The results vary depending on the value of k, as shown in Figure **2.6** [**7**].

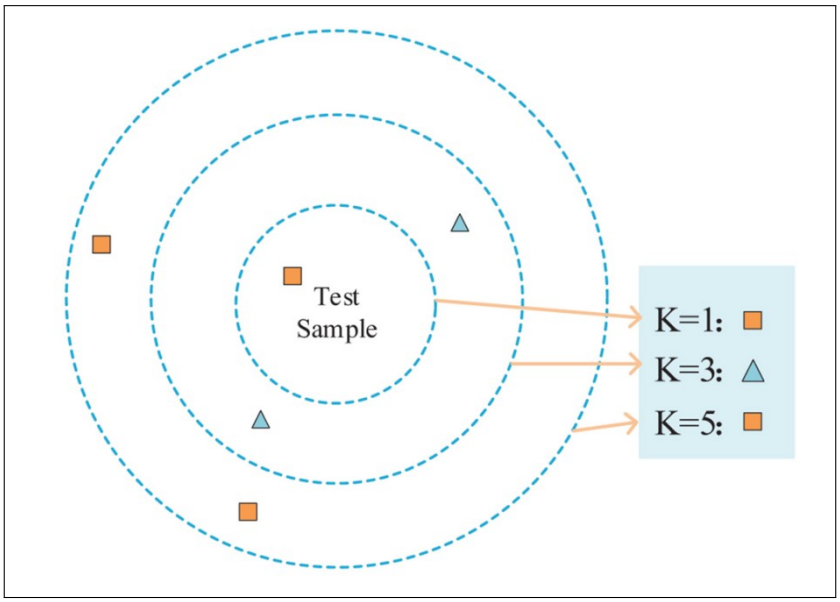


Figure 2.6: Illustration of How the K-Nearest Neighbors (K-NN) Algorithm Works [7].

In EEG signals, it is used to classify a new signal by comparing it with previously recorded signals and assigning it to the most similar emotional state based on distance similarity.

c. Tree-Based and Structural Models

- **Decision Trees** [**57**] are a classification algorithm based on a hierarchical tree-like structure where decisions are made through a sequence of simple rules on data features (if–then rules). Each sample is passed through successive branches until reaching a final node that represents the predicted class or decision. It is characterized by its interpretability and its ability to handle non-linear relationships [**51**].

  In EEG signals, it is used to classify signals based on extracted features through a series of sequential decisions, leading to the final predicted state for each signal.

d. Bagging Methods

- **Random Forest (RF)** [**58**] is an ensemble classification model based on multiple decision trees, where the final prediction is obtained through majority voting. It reduces overfitting and improves stability by introducing randomness in feature selection during tree construction [**7**].

  In EEG signals, it is used to classify states based on features extracted from the signal, and the final decision is obtained by combining the outputs of several trees.

- **Extra Trees** [**58**] is an ensemble learning method that builds multiple decision trees on the same dataset. It introduces randomness by selecting random split points for each feature instead of searching for the optimal one, which increases model diversity. The predictions of all trees are then aggregated to produce the final output. Unlike Random Forest, it does not rely on bootstrapping, as each tree is trained on the full data [**59**].

  In EEG signals, this method improves classification stability by using highly randomized trees, which helps better capture variability in neural signal features.

e. Boosting Methods

- **XGBoost** is a boosting-based machine learning algorithm that builds decision trees sequentially, where each tree corrects the errors of the previous ones. It is effective in modeling complex and non-linear relationships and is widely used for high-performance classification in complex data [**60**].

  In EEG signals, it is applied to extract discriminative patterns from extracted features and improve classification accuracy across different mental and emotional states [**61**].

- **LightGBM** [**62**] is a gradient boosting framework based on the Gradient Boosting Decision Trees (GBDT) technique, designed for efficiency and scalability, where trees are built sequentially using an improved learning process. Unlike standard boosting implementations, it enables faster training and lower computational cost, making it suitable for large, high-dimensional EEG datasets with extensive recordings, while maintaining strong classification performance [**63**].

2.8.2 Deep Learning Models

a. CNN-Based Models

- **Convolutional Neural Network (CNN)** [**15**] is a deep learning model originally developed by *LeCun et al.* It is designed to automatically extract features and patterns from raw data without manual intervention. It relies on convolutional layers containing filters (kernels) that analyze the input data and detect important features within it.

  For EEG signal analysis, the signal is first divided into small segments (epochs) and then passed through the convolutional neural network to extract local temporal and spatial patterns associated with brain activity. Each filter generates a feature map representing the distribution of these patterns within the signal. The extracted features are subsequently integrated through fully connected layers to classify the emotional state (see Figure **2.7**) [**64**].

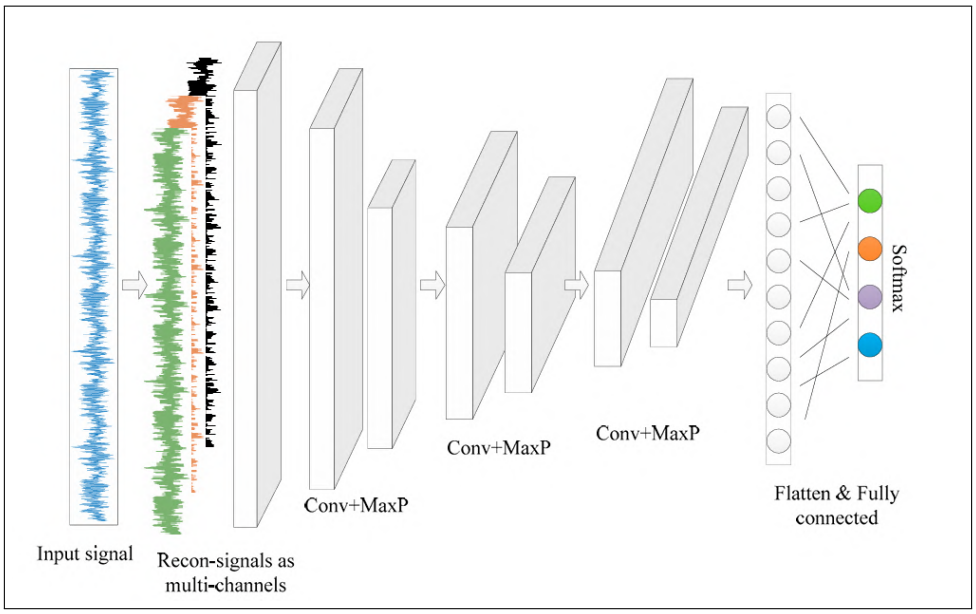


Figure 2.7: 1D Deep Convolutional Neural Network (CNN) Structure [**8**].

- **EEGNet** is a compact convolutional neural network (CNN) specifically designed for the analysis of electroencephalogram (EEG) signals in brain–computer interface (BCI) applications. Originally proposed by Lawhern et al., it employs specialized operations, including depthwise and separable convolutions, to efficiently extract spatial and temporal features from EEG data while maintaining a simple architecture and reducing the number of trainable parameters.

  For EEG-based applications, EEGNet is known for its strong generalization across different tasks and its ability to achieve good performance even with limited training data. It also provides an interpretable framework that enables the visualization and analysis of features extracted from brain signals [**17**].

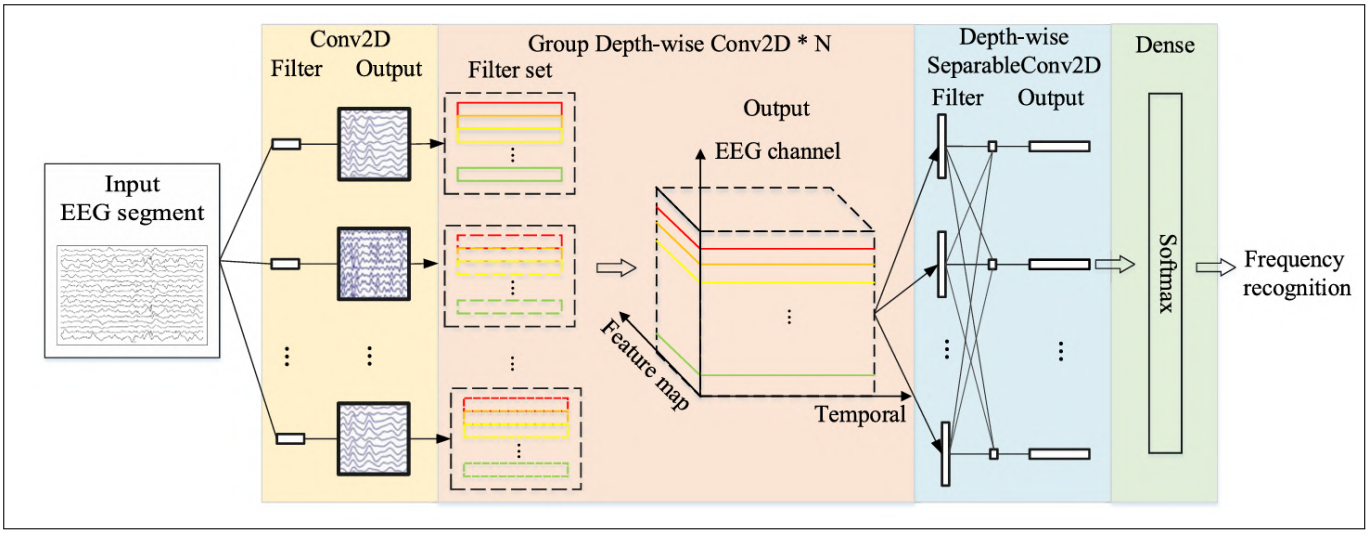


Figure 2.8: The Architecture of the EEGNet [9].

b. RNN-Based Models

- **Recurrent Neural Network (RNN)** is a deep learning model designed to process sequential data over time. It relies on a hidden state that connects current inputs with previous ones, enabling the network to capture contextual information, learn relationships between consecutive time points, and handle sequences of varying lengths.

  In the context of EEG signals, RNNs are used to analyze brain activity over time through their hidden states, which transfer information between successive time points. They are commonly employed for classification tasks based on extracted EEG features, enabling the accurate identification of emotional states [**10**].

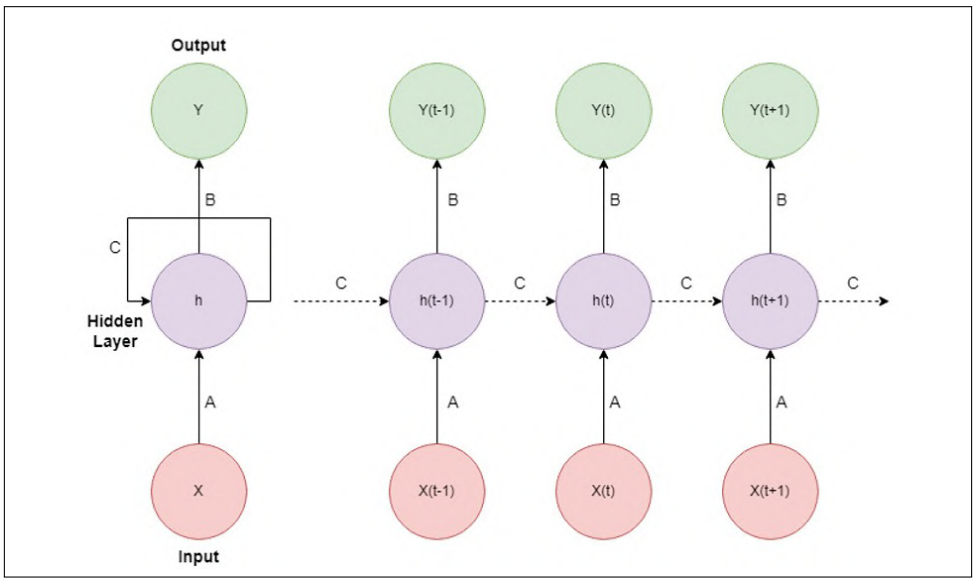


Figure 2.9: Basic Architecture of a Recurrent Neural Network [10].

- **Long Short-Term Memory (LSTM)** [**65**] is an advanced type of recurrent neural network (RNN) designed to process sequential data over time. It was developed by *Hochreiter* and *Schmidhuber* in 1997. It is characterized by its ability to overcome the vanishing gradient prob-

lem and long-term dependency issues. It achieves this by preserving relevant past information and using it to influence current outputs.

This model is based on three main gates: the input gate, which controls what new information should be stored in memory; the forget gate, which determines which information from the previous state should be removed; and the output gate, which decides which part of the stored information is passed as output at each time step.

In the context of EEG signals, LSTM networks are used to analyze temporal variations in brain activity and to better capture the evolution of emotional states over time. By exploiting the sequential nature of EEG signals, they improve classification performance [**66**].

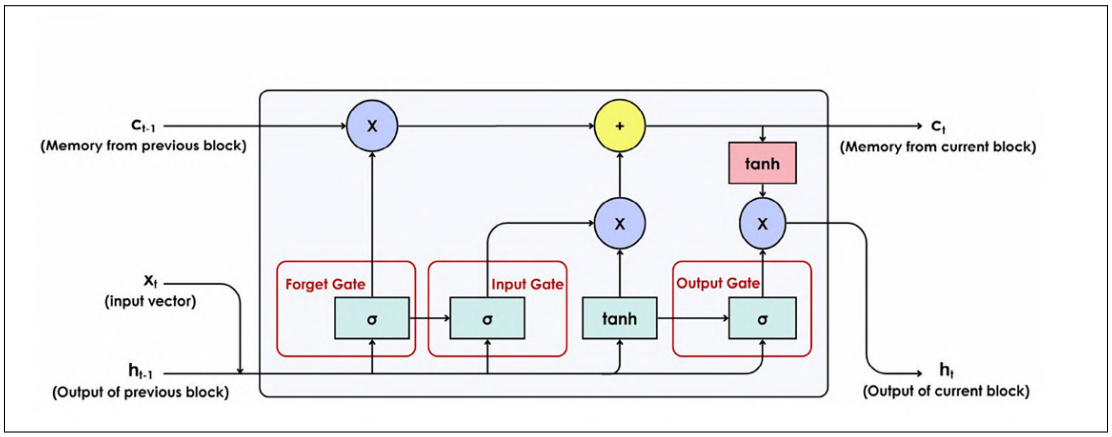


Figure 2.10: Structure of an LSTM Cell [11].

- **Bidirectional LSTM (BiLSTM)** [**16**] is an extension of the conventional LSTM model, developed by *Graves and Schmidhuber in 2005*. It processes data in both forward and backward directions, allowing the model to utilize past and future information simultaneously. This enables better modeling of temporal dependencies and improves classification accuracy in time series and signal analysis tasks.

  In EEG-based sequence modeling, BiLSTM enhances feature learning by capturing information from both past and future time steps, which provides a more complete representation of brain activity dynamics and improves emotion recognition performance [**35**].

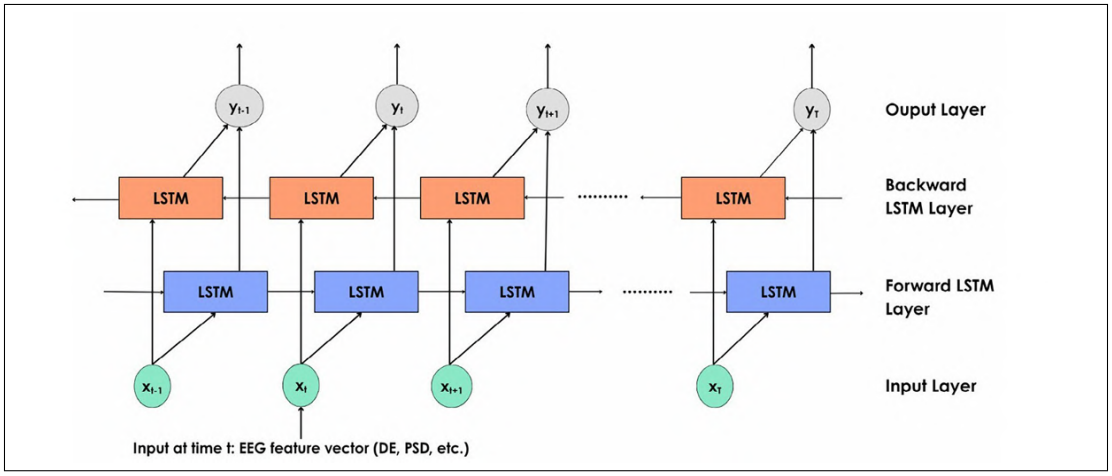


Figure 2.11: Bidirectional LSTM (BiLSTM) Architecture [11].

- **Deep BiLSTM:** is an extension of the standard BiLSTM model composed of multiple stacked BiLSTM layers. This deeper architecture enables the model to learn hierarchical representations of sequential data while capturing richer temporal dependencies. As a result, it improves pattern modeling and enhances classification performance compared to shallow architectures [**47**].

- **Gated Recurrent Unit (GRU)** [**67**] is an advanced type of recurrent neural network (RNN) introduced by *Cho et al. in 2014*. It is designed as a simplified alternative to LSTM while maintaining strong performance in sequential data processing. GRU addresses limitations of traditional recurrent networks, such as the vanishing gradient problem and the difficulty of capturing long-term dependencies in time-series data, through two main mechanisms: the update gate and the reset gate. These gates allow the model to selectively control the flow of information by deciding when to update or reset the hidden state.

  In EEG-based modeling, GRU is favored for its simpler architecture and lower computational complexity compared to LSTM, while still maintaining strong performance in capturing relevant patterns for emotion classification [**68**].

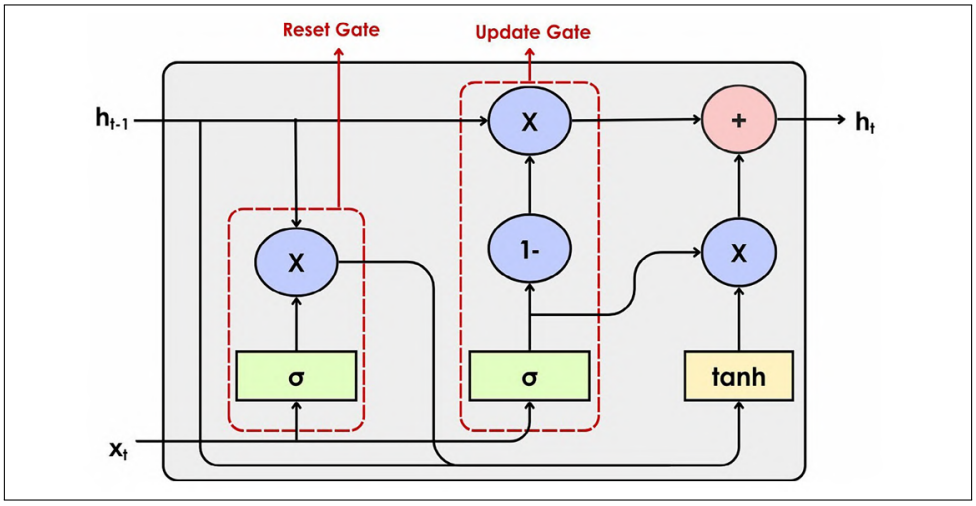


Figure 2.12: Architecture of a Gated Recurrent Unit (GRU) [11].

- **Bidirectional GRU (BiGRU)** is an extension of the standard GRU model that consists of two GRU layers processing the input sequence in opposite directions: forward and backward. This architecture enables the model to capture contextual information from both past and future time steps simultaneously, leading to a more comprehensive representation of temporal dependencies. As a result, it improves the learning of sequential patterns and enhances performance in time-series classification tasks, particularly in emotion recognition from EEG signals [**69**].

c. Feedforward Neural Network Models

- **Multilayer Perceptron Neural Network (MLPNN)** is a type of neural network consisting of multiple interconnected layers of neurons, trained using the Back-propagation algorithm to minimize the error between predicted and actual values.

This network is distinguished by its ability to learn new representations and features within the hidden layers, which helps it capture complex patterns in the data and improve classification accuracy [**70**].

For this reason, it is widely used in emotion classification in EEG signal analysis systems [**5**].

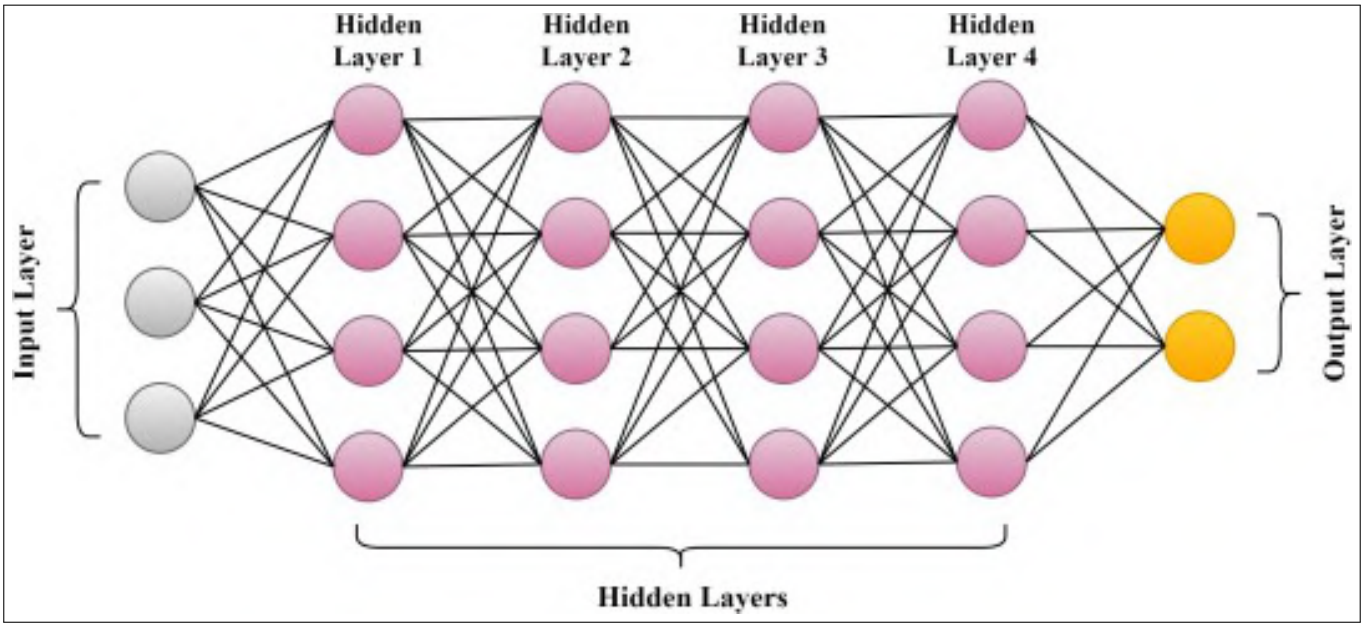


Figure 2.13: Architecture of a Deep Multilayer Perceptron (MLP) Neural Network [12].

d. Attention and Adversarial Models

- **Deep Attention Network (DAN)** [**71**] is a deep learning model proposed by *Ganin and Lempitsky in 2015*. It is designed to learn domain-invariant and shared representations across different datasets, which improves generalization and reduces sensitivity to variations between subjects within a transfer learning framework. This enables the model to focus on the most relevant features while ignoring domain-specific noise.

  In EEG-based applications, DAN is used to address inter-subject variability by reducing the influence of individual-specific patterns and focusing instead on emotion-related neural representations in brain signals, leading to more robust and improved emotion classification performance [**72**].

e. Graph-Based Models

- **Dynamic Adaptive Multigraph Convolutional Network (DAMGCN)** is a graph neural network model that employs a dual-attention mechanism to enhance data analysis by focusing on the most informative features and relationships. It represents data through multiple adaptive graphs, enabling the model to capture complex interactions and dependencies more effectively.

  In EEG-based applications, DAMGCN models brain activity as a graph connecting different EEG channels and utilizes attention mechanisms to identify the most relevant brain regions and frequency bands. This improves feature representation, enhances interpretability, and increases emotion classification accuracy [**73**].

f. Hybrid Deep Learning Models

Hybrid models are considered among the modern approaches in deep learning, as they rely on combining two or more models to leverage the strengths of each within a unified framework. This integration enables simultaneous learning of different types of features, such as temporal, spatial, and contextual features, thereby enhancing the model's ability to represent data more comprehensively.

In the field of emotion recognition from EEG signals, these models are employed by integrating different architectures depending on the nature of the signal for example, combining LSTM with GRU to capture temporal dynamics, or combining CNN with LSTM to simultaneously extract spatial and temporal features. Additionally, GCN can be merged with the Attention mechanism to improve the representation of relationships between brain channels and focus on the regions most influential in the signal.

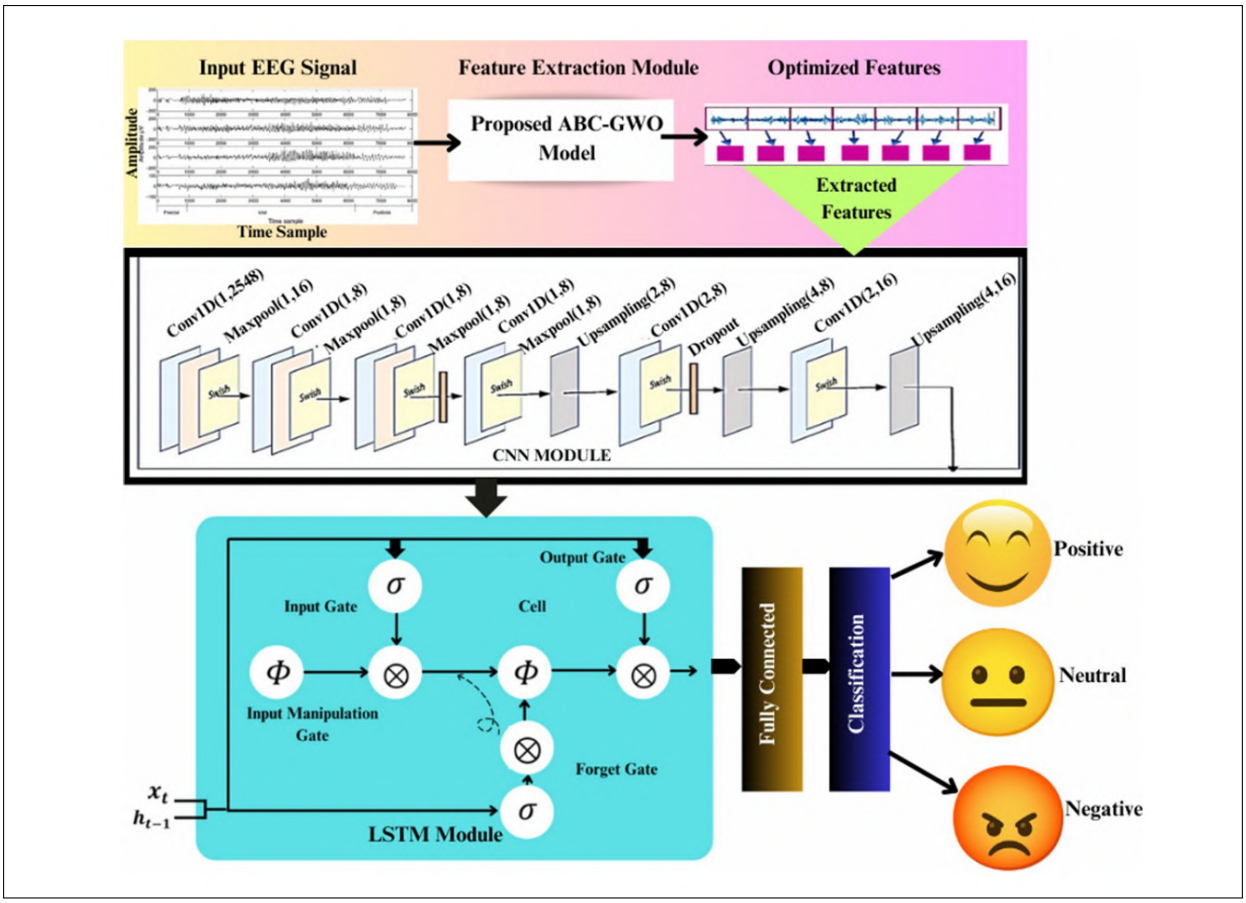


Figure 2.14: Hybrid Deep Learning Architecture Integrating CNN and LSTM Modules for Emotion Recognition [13].

2.9 Related Work

This section reviews the most important previous studies related to emotion classification from EEG signals, focusing on the differences in adopted evaluation strategies, which can be categorized into three main types: Random Split, Subject-Dependent, and Subject-Independent.

2.9.1 Random-Split Evaluation

The Random Split approach involves mixing EEG data from all individuals and dividing it randomly into training and testing sets without considering participant independence. In this evaluation method, data from the same individual may appear in both the training and testing stages, often resulting in high reported performance.

In the context of developing methods for emotion classification from EEG signals, studies have varied between traditional approaches based on feature extraction and machine learning and modern approaches based on deep learning. Within this framework, some works have adopted EEG signals from the GAMEEMO and DEAP datasets:

- A study using GAMEEMO [**28**] applied preprocessing with a Sinc filter to denoise the signals, followed by the extraction of frequency-domain features using Fast Fourier Transform (FFT) and spectral density via Welch's PSD, and then selected the most important features using the mRMR method to reduce redundant dimensions. Subsequently, classification algorithms such as RCAR, RIPPER, and SVM were utilized, with the RCAR model achieving the best performance at an accuracy of 94.69%.

- Another study using DEAP [**51**] data carried out preprocessing that included segmenting the EEG signals into short time windows, followed by feature extraction and dimensionality reduction using Principal Component Analysis (PCA). Multiple machine learning models were then applied, including SVM, KNN, LDA, Logistic Regression, and Decision Trees. The SVM model showed the best performance among these, reaching an accuracy of 73.75%.

Conversely, other studies adopted deep learning models to process EEG signals directly:

- A BiLSTM model [**35**] with the spectral entropy feature was used on GAMEEMO data after preprocessing, achieving an accuracy of 76.91% in binary emotion classification.

- In other works, the BiLSTM model [**66**] was enhanced with an attention mechanism to improve the sequential representation of EEG signals. The model was trained and evaluated on DEAP data after preprocessing steps that included downsampling, noise removal, signal filtering, and segmentation into fixed time windows. The results showed that the attention-based model outperformed the standard model, achieving accuracies of 90.1% and 87.9% for the binary case, compared to 89.7% and 85.7% without the attention mechanism.

Additionally, the Deep BiLSTM model was adopted as a deep extension of BiLSTM architectures.

- One study on GAMEEMO [**47**] data involved preprocessing based on signal decomposition using EMD and VMD to extract intrinsic mode functions (IMFs). Statistical features such as maxima, minima, and mean were then extracted to represent the signal before being fed into a Deep BiLSTM model composed of three BiLSTM layers to learn temporal relationships, achieving 70.89% accuracy in binary classification and 90.33% in multiclass classification.

- Another study using DEAP [**48**] data included preprocessing steps of signal filtering, downsampling, and data segmentation, followed by the extraction of temporal and spectral features such as mean, variance, standard deviation, homogeneity, and kurtosis, along with dimensionality reduction and selection of the most important features. A Stacked BiLSTM model was then used to handle temporal dependencies in both directions, achieving very high performance with accuracies of 99.45% for valence, 96.87% for arousal, and 99.68% for liking.

Table 2.8: Summary of Random-Split Evaluation for EEG-Based Emotion Recognition Studies.

Year	Contributions	Dataset	Classification	Evaluation	Metrics
2020 [**51**]	KNN, SVM, DT, LR, NB, LDA.	DEAP	`Binary` (Valence, Arousal, Dominance, Liking).	50-fold CV	Valence: 70.41% (KNN); Arousal: 73.75% (SVM); Dominance: 71.73% (DT); Liking: 69.64% (LR); Liking: 67.50% (LDA).
2020 [**66**]	BiLSTM	DEAP	`Binary` (Valence, Arousal); `Multi-class` (Low / Mid / High).	4-fold CV	`Binary`: Valence 89.7%, Arousal 85.7%; `Multi-class`: Valence 50.9%, Arousal 74.0%.
2020 [**66**]	BiLSTM + ATTENTION	DEAP	`Binary` (Valence, Arousal); `Multi-class` (Low / Mid / High).	4-fold CV	`Binary`: Valence 90.1%, Arousal 87.9%; `Multi-class`: Valence 83.5%, Arousal 82.6%.
2020 [**35**]	BiLSTM	GAMEEMO	`Binary` (Positive/Negative , Valence).	Hold-out 80/20	Acc: 76.91%.
2022 [**48**]	Stacked BiLSTM	DEAP	`Binary` (Valence, Arousal, Liking).	Hold-out 70/30	Valence: 99.45%; Arousal: 96.88%; Liking: 99.69%.
2022 [**47**]	Deep BiLSTM	GAMEEMO	`Binary` (Positive/Negative, Valence); `Multi-class`.	10-fold CV	`Binary`: 70.89%; `Multi-class`:90.33%.
2024 [**28**]	RCAR; RIPPER; SVM.	GAMEEMO	`Multi-class` (4 classes).	Hold-out 70/30	RCAR: 94.69%; RIPPER: 94.06%; SVM: 94.49%.

2.9.2 Subject-Dependent Evaluation

The Subject-Dependent split method relies on training and evaluating the model using each participant's data independently, without merging it with data from other individuals. This approach enables the extraction of participant-specific brain activity patterns, often referred to as emotional brain fingerprints, and provides a personalized evaluation for each participant.

Within this approach, several deep models have been proposed to improve emotion classification accuracy from EEG signals:

- Among these is the DSSA-Net (Directional Spatial and Spectral Attention Network) [**74**], which relies on merging spatial and spectral information using attention mechanisms that focus on the most important channels and bands. The process included signal segmentation, Butterworth filtering, and temporal segmentation. This was followed by feature extraction and organization using differential entropy (DE) across five frequency bands with spatial representation before the final classification stage. The model achieved high performance on several databases.

- Another proposed model is DE-CNN-BiLSTM [**75**], which is based on a hybrid methodology combining spatial and temporal learning. The process began by dividing the signal into frequency bands and temporally segmenting it using a Hanning window. DE features were then extracted and converted into a spatial representation reflecting the distribution of brain electrodes. This representation was then fed into a CNN to extract deep spatial features, followed by a BiLSTM to capture bidirectional temporal relationships. This enabled the model to achieve high classification performance, reaching 94.86% on DEAP and 94.82% on SEED.

- By contrast, the DAMGCN model [**73**] relies on integrating graph convolutional networks (GCN) with an attention mechanism. Features were first extracted from EEG signals, and a matrix of relationships between electrodes was constructed. GCN was then used to capture the spatial structure, while the attention mechanism directed focus toward the most influential channels and frequency bands before passing the final representation to the classifier for emotion classification.

- As for the CIT-EmotionNet model [**76**], it relies on a dual architecture combining CNN and Transformer, where the CNN extracts local features while the Transformer processes global relationships. After signal segmentation, PSD features were extracted and represented in a spatial-spectral form. The resulting data was then simultaneously fed into both branches, and the extracted representations were merged through the CIT unit, which enables interaction between local and global features. Finally, the merged representation was passed to a classifier to produce the final emotion classification result.

To facilitate comparison, the following table **2.9** summarizes the most important results achieved in previous studies:

Table 2.9: Summary of Subject-Dependent Evaluation for EEG-Based Emotion Recognition Studies.

Year	Contributions	Dataset	Classification	Evaluation	Metrics
2022 [**75**]	DE+CNN+ BiLSTM	DEAP SEED	`DEAP/Binary` (Valence, Arousal). `SEED/Multi-class` (Positive/Negative, neutral)	5-fold CV	`DEAP`: Valence 94.02%, Arousal 94.86% `SEED`: 94.82%.
2023 [**76**]	CIT-EmotionNet	SEED SEED-IV	`SEED/Multi-class` (3 classes); `SEED-IV/Multi-class` (4 classes).	Hold-Out 60/40	`SEED`: 98.57%; `SEED-IV`: 92.09%
2024 [**28**]	RCAR; RIPPER; SVM.	GAMEEMO	`Multi-class` (4 classes).	Hold-out 70/30	RCAR: 87.89%; RIPPER: 79.51%; SVM: 87.29%.
2024 [**73**]	DAMGCN	SEED SEED-IV DEAP	`SEED/Multi-class` (3 classes); `SEED-IV/Multi-class` (4 classes); `DEAP/Binary` (Valence, Arousal, Dominance).	LOTO	`SEED`: 90.4%; `SEED-IV`: 83.55% DEAP: Valence 78.21%, Arousal 78.21%, Dominance 78.21%.
2024 [**73**]	DAMGCN	SEED SEED-IV DEAP	`SEED/Multi-class` (3 classes); `SEED-IV/Multi-class` (4 classes); `DEAP/Binary` (Valence, Arousal, Dominance).	10-FOLD CV	`SEED`: 99.42%; `SEED-IV`: 96.86%; `DEAP`: Valence 96.96%, Arousal 97.17%, Dominance 97.50%.
2025 [**74**]	DSSA-Net	SEED SEED-IV DEAP	`SEED/Multi-class` (3 classes); `SEED-IV/Multi-class` (4 classes); `DEAP/Binary` (Valence, Arousal, Dominance).	SEED (9Train-6Test); SEED-IV (16Train-8Test); DEAP (10-FOLD-CV).	`SEED`: 96.61%; `SEED-IV`: 85.07% `DEAP`: Valence 94.97%, Arousal 94.73%.

2.9.3 Subject-Independent Evaluation

The Subject-Independent splitting approach relies on training the model using data from a group of participants and testing it on data from new participants who were not involved in the training phase. This enables the assessment of the model's ability to generalize and recognize emotional states in unseen individuals. It is considered one of the most challenging evaluation strategies due to inter-subject variability in EEG signals.

- In this study [**77**], a unified framework is adopted for EEG signal processing and feature extraction for both machine learning and deep learning models. The signals underwent the same preprocessing stages by segmenting them into time windows using a Sliding Window approach. Statistical and frequency-based features, such as mean, standard deviation, entropy, and spectral band powers, were then extracted using FFT, and normalization was applied. Once these unified representations were obtained, the same features were used as inputs for all models, while the processing strategy differed according to the classifier. Random Forest treated the features as fixed vectors, whereas LSTM-GRU processed them as temporal sequences to capture temporal dependencies. Random Forest achieved an accuracy of 79.6%, compared to 94.5% for LSTM-GRU.

- Another study [**78**] relies on GAMEEMO and DEAP data. The signals were processed without segmentation into fixed windows and were then passed through the Wavelet Scattering Transformation (WST) to extract multi-level representations reflecting the deep structure of the signal. A set of features, including statistical features, entropy measures, Hjorth parameters, and Higuchi's fractal dimension, was subsequently extracted from both the raw signals and the WST representations. These features were merged into a comprehensive feature vector and reduced using Linear Discriminant Analysis (LDA) to improve class separability and simplify the representation before the decision stage. Classification was then performed using KNN with majority voting across channels to determine the final decision. The approach achieved an accuracy of 100% with voting, compared to 79.02% without voting. It also achieved accuracies of 98.44%, 99.22%, and 97.58% on DEAP under different evaluation scenarios.

- In a multimodal context, study [**30**], utilizes DEAP data, which include both EEG signals and facial expression videos recorded from the same participants. The EEG signals underwent preprocessing steps including downsampling, frequency filtering, noise removal, and segmentation into time windows. Meanwhile, the facial data were prepared by extracting frames, standardizing their dimensions, and removing unusable samples. Features covering temporal, frequency, and wavelet domains were extracted from EEG signals and then combined with numerical representations of facial expressions. The resulting brain and facial features were fused into a unified feature-level representation before being passed to the model. The fused representation was then processed using a CNN-Transformer architecture equipped with an attention mechanism to emphasize the most relevant attributes from each modality. Finally, emotion classification was performed based on the fused brain and facial representations.

Table 2.10: Summary of Subject-Independent Evaluation for EEG-Based Emotion Recognition Studies.

Year	Contributions	Dataset	Classification	Evaluation	Metrics
2024 [**73**]	DAMGCN	SEED SEED-IV DEAP	`SEED/Multi-class` (3 classes); `SEED-IV/Multi-class` (4 classes); `DEAP/Binary` (Valence, Arousal, Dominance).	LOSO	`SEED`: 73.21%; `SEED-IV`: 68.22%; DEAP: Valence 64.61%, Arousal 64.61%,Dominance 64.61%.
2025 [**74**]	DSSA-Net	SEED SEED-IV	`SEED/Multi-class` (3 classes); `SEED-IV/Multi-class` (4 classes);	LOSO	`SEED`: 87.03%; `SEED-IV`:75.86%.
2025 [**30**]	EMOTRANS	DEAP	`Binary` (Valence, Arousal, Dominance, Liking).	LOSO	Valence 87.8%, Arousal 89.3%; Dominance 88.9%; Liking 89.1%.
2025 [**78**]	WST+LDA+ KNN+Voting	GAMEEMO DEAP	`GAMEEMO/Multi-class` (4 classes); `DEAP/Binary` (Valence, Arousal); `DEAP/Multi-class`	LOSO	`GAMEEMO`: without Voting 79.02%, with voting 100 ; `DEAP`: Valence (without Voting 77.83% , with voting 98.43%); Arousal (without Voting 79.19% , with voting 97.58%); `DEAP/Multi-Class` (without Voting 65.67% , with voting 97.58%).
2025 [**77**]	RF LSTM-GRU	GAMEEMO	`Multi-class` `Binary` (Positive/Negative, Valence).	(22train / 6test)	`Binary`: RF 85% , LSTM-GRU 93.3%; `multi-class`: RF 79.6%, LSTM-GRU 94.5%.

2.10 Discussion on Emotion Recognition through EEG Analysis

The results reported by state-of-the-art literature indicate that the performance of emotion recognition models varies significantly across datasets. The SEED dataset achieves the best performance compared to the others, followed by DEAP, while GAMEEMO shows the lowest performance. This variability reflects differences in the nature of the data, including stimulus type, number of sessions, and the stability of the recorded signals. Performance is also directly influenced by the adopted evaluation method, where Random Split yields the highest results, while performance decreases under Subject-Dependent evaluation and reaches its lowest level in Subject-Independent (LOSO) settings.

Overall, these results highlight that the main challenge in emotional state classification lies not only in achieving high accuracy but also in the ability of models to generalize to new individuals. In Subject-Dependent settings, models achieve better performance because they can capture individual EEG signatures. However, performance decreases when moving to more realistic settings such as LOSO, where models are tested on unseen participants whose data were not used during training. This decline reflects a generalization gap caused by natural variability in EEG signals between individuals, making the development of robust and generalizable models to handle this variability one of the key challenges in this field.

2.11 Conclusion

In summary, this chapter analyses previous research on emotion recognition using EEG signals. It explains the different methodologies and machine learning and deep learning models used, as well as the strengths and limitations of each study. The review emphasises the necessity of developing more effective and accurate methodologies, which will underpin the next chapter's presentation of the proposed methodology, data processing, and model selection for reliably classifying emotional states.

Research Methodology and Technical Framework

3.1 Introduction

This chapter presents the methodology adopted in this study for developing an effective emotion recognition system based on electroencephalogram (EEG) signals. The proposed methodology focuses on the development of an integrated framework capable of handling the complex characteristics of EEG signals and addressing associated challenges, including noise, data scarcity, and significant inter-individual variability.

To achieve this objective, various preprocessing, data augmentation, and feature extraction techniques are applied. In addition, multiple classification approaches are investigated, including machine learning, deep learning, hybrid models, ensemble learning methods, and transfer learning strategies based on fine-tuning and intra-subject adaptation. These approaches are designed to exploit complementary information from EEG signals, improve classification accuracy, and enhance the generalizability and robustness of the developed models.

3.2 Problem Statement

Despite the significant advances in EEG-based Emotion Recognition systems, developing accurate and generalizable models still faces a number of key challenges. These challenges are not limited to just the data itself but also extend to the characteristics of the signal, the nature of data collection, and differences in neural responses among individuals.

In this context, the most prominent of these issues can be summarized as follows:

3.2.1 Data Scarcity and Accessibility

EEG databases are generally limited in size due to the complexity and high cost of data collection, in addition to ethical constraints related to participant privacy. Publicly available open-access datasets are often relatively small in terms of the number of participants and recordings, while many other databases are either unavailable or require special licenses for use. This limited data size reduces sample diversity, which directly affects models' ability to generalize to different real-world environments.

3.2.2 EEG Noise Susceptibility

EEG recordings are highly sensitive and are affected by various types of noise, such as muscle activity (EMG artifacts), eye movements (EOG artifacts), as well as electrical and environmental interferences. As a result, raw signals become non-stationary and difficult to interpret, which directly impacts the performance of classification models. Therefore, handling these signals requires precise pre-processing due to their complex and unstable nature compared to more structured data types.

3.2.3 Signal Amplitude Variability

EEG signals exhibit inter-individual variability in amplitude even when exposed to the same stimuli, resulting in emotional responses that are similar in type but vary in intensity. This difference is mainly due to physiological factors such as the level of neural activity and variability in response to stimuli. This variation leads to differences in the representation of recorded signals between individuals despite similar emotional states.

3.2.4 Inter-Subject Variability and Generalization Gap

EEG signals exhibit notable inter-individual differences due to variations in brain structure, cognitive patterns, and emotional responses. As a result, each individual tends to present a unique neural signature, which reflects their specific brain activity patterns. This inherent variability makes it difficult to develop universal models that can generalize across all individuals with the same level of accuracy. Consequently, models trained on a specific group often experience a noticeable drop in performance when evaluated on unseen subjects. This issue highlights one of the main challenges in EEG-based emotion recognition, namely building robust and generalizable systems.

3.3 Motivation

Given the challenges associated with emotion recognition systems based on EEG signals, it is evident that improving model performance requires adopting a comprehensive approach that takes into account data limitations, the sensitivity of signals to noise, and significant inter-individual variability. Therefore, it is not sufficient to rely solely on conventional models; rather, it is necessary to develop an integrated framework that combines multiple techniques aimed at improving representation quality, boosting model performance, and enhancing their generalization and stability. Within this framework, each approach is directed toward addressing a specific aspect of these challenges.

3.3.1 Addressing Data Scarcity

To address data scarcity in EEG systems, strategies are adopted that aim to increase data volume and diversity and improve data representation through techniques such as Windowing, Overlapping, and Feature Extraction from raw signals. In addition, Ensemble Learning is used to enhance model stability and generalization performance. Moreover, knowledge transfer strategies such as Intra-Fine Tuning are applied to improve adaptability across individuals and reduce the impact of data scarcity.

3.3.2 Handling EEG Noise Sensitivity

To reduce the impact of noise in EEG signals, Bandpass Filtering is relied upon to improve signal purity and eliminate unwanted frequencies, while benefiting from Processed Data previously available in the database, which exhibit lower levels of interference. In addition, Feature Extraction allows for the extraction of more indicative features that reduce the effect of random fluctuations and improve signal representation, ultimately leading to enhanced data quality and improved model performance.

3.3.3 Addressing Signal Amplitude Variability

To address the problem of amplitude variation in EEG signals across individuals, Normalization is applied to standardize the signal value range for all samples, thereby reducing the effect of differences in signal strength resulting from physiological factors across individuals. This makes signals more comparable by reducing reliance on absolute values and focusing on relative patterns within the signal, resulting in improved representation stability and making models less sensitive to amplitude changes.

3.3.4 Managing Inter-Subject Variability and Generalization Gap

In EEG studies, inter-individual variability is addressed using the Subject-Dependent setting, which involves training and testing the model on the same individuals within the same data distribution. In contrast, the Subject-Independent setting is used to evaluate the model's ability to generalize to new individuals, where the Generalization Gap problem emerges due to differences in signal characteristics between individuals. To improve this generalization, knowledge transfer through Fine-Tuning is used to adapt pretrained models with limited data, thereby contributing to enhanced performance stability and reducing the impact of individual variability.

3.4 Dataset Overview

3.4.1 Description of the Dataset

This study relies on the GAMEEMO dataset, a public database of electroencephalogram (EEG) signals provided by Alakus et al. in 2020 and published in the journal Biomedical Signal Processing and Control, which is also available on the Mendeley Data and Kaggle platforms. Unlike traditional emotion elicitation approaches that rely on images or video clips, the GAMEEMO dataset uses interactive games to induce emotions, resulting in a more realistic and interactive environment and more natural and complex emotional responses.

3.4.2 Participants and Recording Procedures

- The dataset includes 28 healthy participants aged between 20 and 27 years, including both males and females, providing diversity in neural and emotional responses.
- EEG signals were recorded using the EMOTIV EPOC+ wireless device, which features 14 channels according to the international 10–20 system.
- Each participant took part in an experiment involving four different games, each lasting 5 minutes, resulting in a total recording time of 20 minutes of continuous EEG signals per participant.
- Accordingly, each participant completed four recording sessions (one session per game). Since the dataset includes 28 participants, the total number of recording sessions was 112.

- The games were designed to elicit four different emotional states as follows:
  - Boring game → to elicit feelings of boredom and frustration.
  - Calm game → to elicit a relaxed state.
  - Horror game → to elicit fear and increase attention.
  - Funny game → to elicit joy and positive energy.

3.4.3 Data Structure and Technical Details

- The data is organized within a main folder named GAMEEMO, which contains subfolders for each participant. Each folder contains CSV files representing recordings of different games, with both raw and pre-processed versions available.
- Each CSV file represents a time series, where the rows correspond to time points and the columns represent the following EEG channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4.
- Each recording session lasts 5 minutes at a sampling rate of 128 Hz, resulting in approximately 38,400 time samples per channel per file.
- Self-assessments are used as labels in the classification process.

3.5 Emotion Representation and Annotation

3.5.1 Emotion Representation using the Self-Assessment Manikin (SAM)

In this study, the Valence–Arousal model is adopted for emotion representation, where valence indicates whether the feeling is positive or negative, while arousal indicates the level of activity from low to high. SAM (Self-Assessment Manikin) forms are used, in which each participant assesses their emotional state directly after each five-minute gaming session by entering numerical values from 1 to 9 for both valence and arousal.

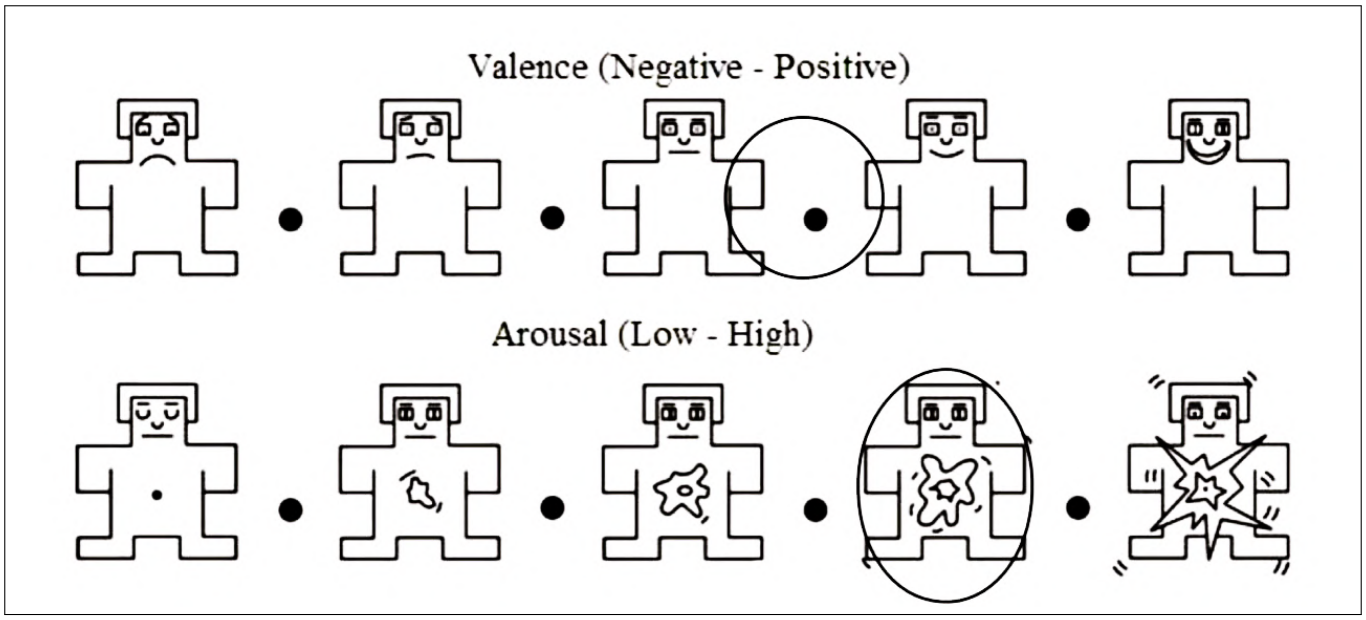


Figure 3.1: SAM Scale Used by Subject 16 to Label Valence and Arousal Levels after Game 4.

3.5.2 Annotation Strategy

The valence and arousal values extracted from the SAM forms are used to generate emotional labels based on a threshold (Threshold = 5), where the states are classified as follows:

- If Valence $> 5 \rightarrow$ Positive Valence.
- If Valence $\leq 5 \rightarrow$ Negative Valence.
- If Arousal $> 5 \rightarrow$ High Arousal.
- If Arousal $\leq 5 \rightarrow$ Low Arousal.

Accordingly, four emotional categories representing all possible combinations of valence and arousal are defined:

- Positive Valence – High Arousal (HAPV).
- Positive Valence – Low Arousal (LAPV).
- Negative Valence – High Arousal (HANV).
- Negative Valence – Low Arousal (LANV).

The resulting emotional labels obtained after the annotation process across all subjects and game sessions are presented in Table **3.1**.

Table 3.1: Emotional Labels Distribution Across Game Sessions per Subject.

Subjects	G1	G2	G3	G4	Labels N°
Subject 1	LANV	LANV	HANV	HANV	2
Subject 2	LANV	LANV	HANV	LAPV	3
Subject 3	LANV	LAPV	HANV	HAPV	4
Subject 4	LANV	HAPV	HANV	HAPV	3
Subject 5	LANV	HAPV	HANV	HAPV	3
Subject 6	LANV	HAPV	LANV	HAPV	2
Subject 7	LANV	HAPV	HANV	HAPV	3
Subject 8	LANV	HAPV	LANV	HANV	3
Subject 9	LANV	LANV	HANV	LANV	2
Subject 10	LANV	LANV	HANV	LAPV	3
Subject 11	LANV	LAPV	LANV	HAPV	3
Subject 12	LANV	LAPV	LANV	LANV	2
Subject 13	LAPV	LANV	LANV	LAPV	2
Subject 14	LANV	LAPV	HANV	LAPV	3
Subject 15	LANV	LAPV	HANV	HAPV	4
Subject 16	LANV	LAPV	HANV	HAPV	4
Subject 17	LANV	LAPV	HANV	HANV	3
Subject 18	LAPV	LANV	HANV	HAPV	4
Subject 19	LANV	LANV	LANV	LAPV	2
Subject 20	LANV	LAPV	HANV	LAPV	3
Subject 21	LANV	HAPV	HANV	HAPV	3
Subject 22	LANV	LAPV	HANV	HANV	3
Subject 23	LANV	LANV	LANV	HAPV	2
Subject 24	LANV	HAPV	LANV	HAPV	2
Subject 25	LANV	LANV	HANV	HAPV	3
Subject 26	LAPV	LANV	HANV	LANV	3
Subject 27	LANV	LAPV	HANV	LANV	3
Subject 28	LANV	LAPV	LANV	LAPV	2

Table **3.2** shows the distribution of emotional labels across the four game sessions for all 28 participants. It represents the number of subjects in each emotional category per game, highlighting the most frequent emotional response in each session.

Table 3.2: Frequency Distribution of Emotional Labels Across Game Sessions for All Participants.

Label	G1	G2	G3	G4
LAPV	3	11	0	7
LANV	25	10	9	4
HAPV	0	7	0	13
HANV	0	0	19	4

To complement the tabular representation, the following Figure **3.2** provides a visual illustration of the same distribution, enabling clearer comparison of emotional label frequencies across the four game sessions.

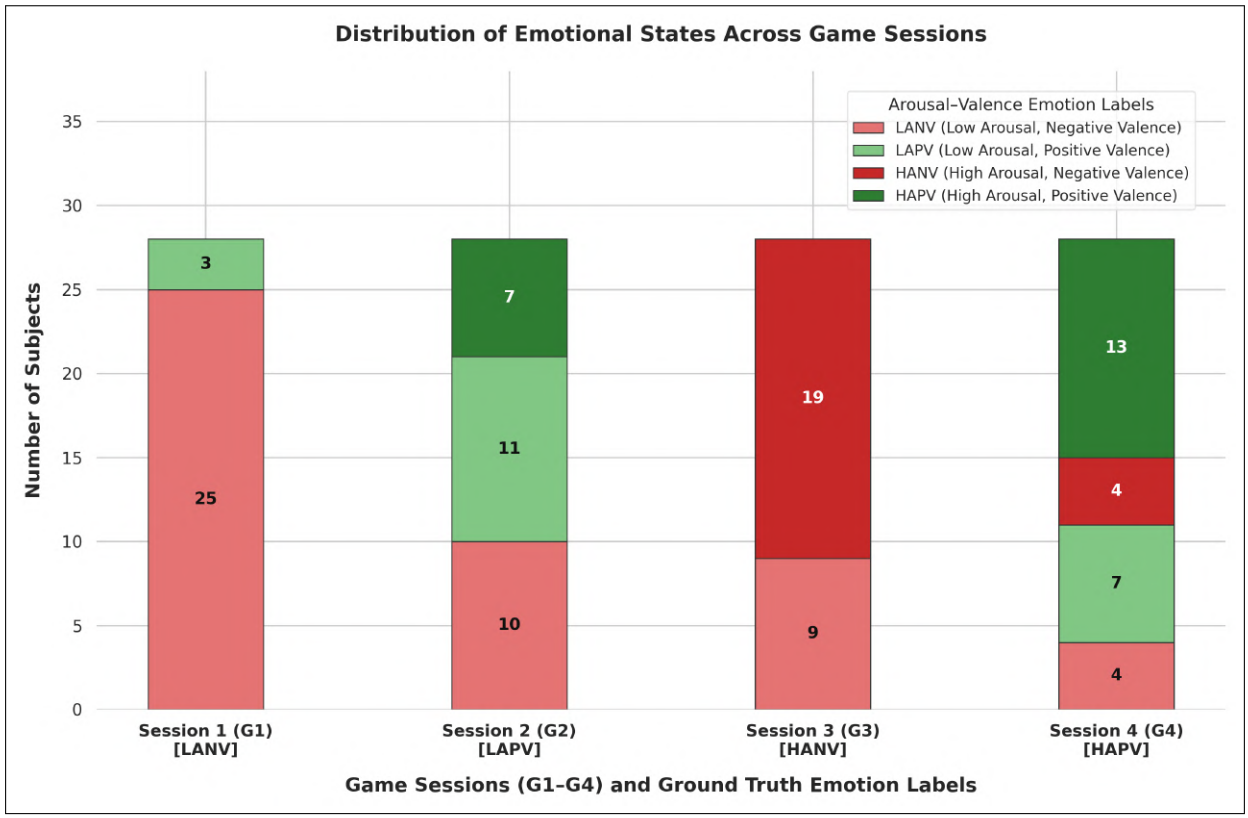


Figure 3.2: Visualization of the frequency distribution of emotional labels across game sessions (G1–G4) for all participants.

3.6 Proposed Method

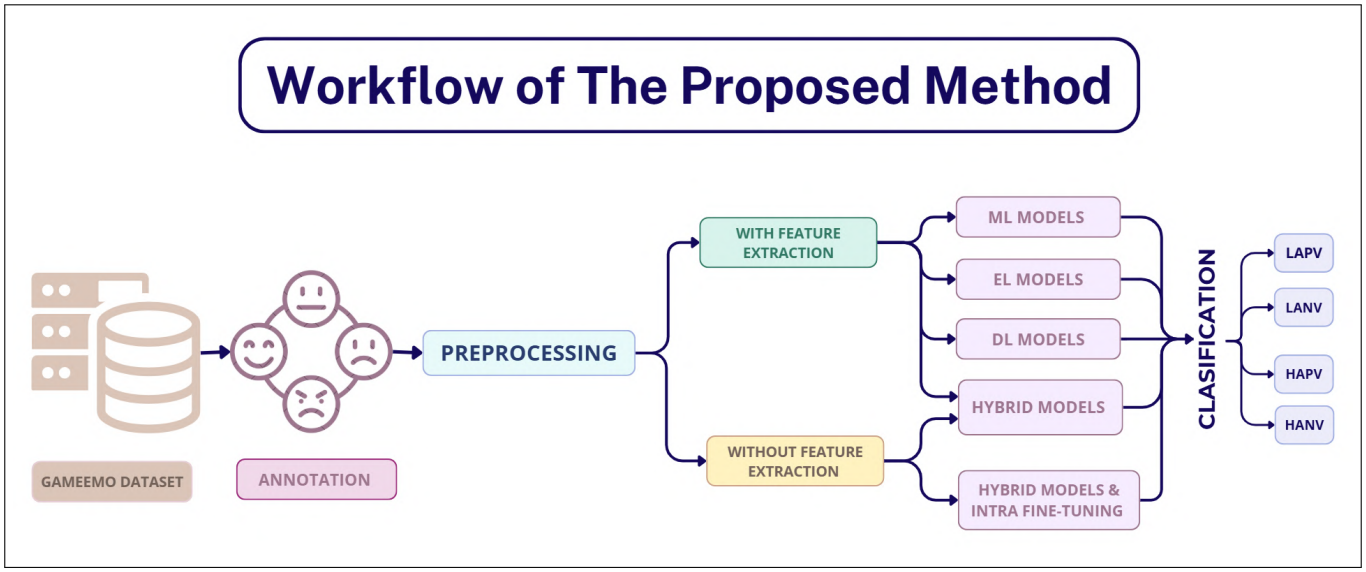


Figure 3.3: Overview of The proposed Method for Emotion Recognition Using EEG Signals.

Figure **3.3** illustrates the overall framework of the proposed EEG-based emotion classification methodology. The pipeline begins with annotation and preprocessing stages, followed by a classification phase based on two complementary strategies: feature-based representation and direct time-window-based learning. This dual-stream framework aims to improve the robustness and generalization capability of emotional state representation.

3.6.1 Preprocessing of EEG Signals

In this study, EEG signals are preprocessed through four consecutive stages, as illustrated in Figure **3.4**: Bandpass Filtering, Windowing, Normalization, and Label Encoding. This process aims to enhance signal quality and properly organize the data for classification.

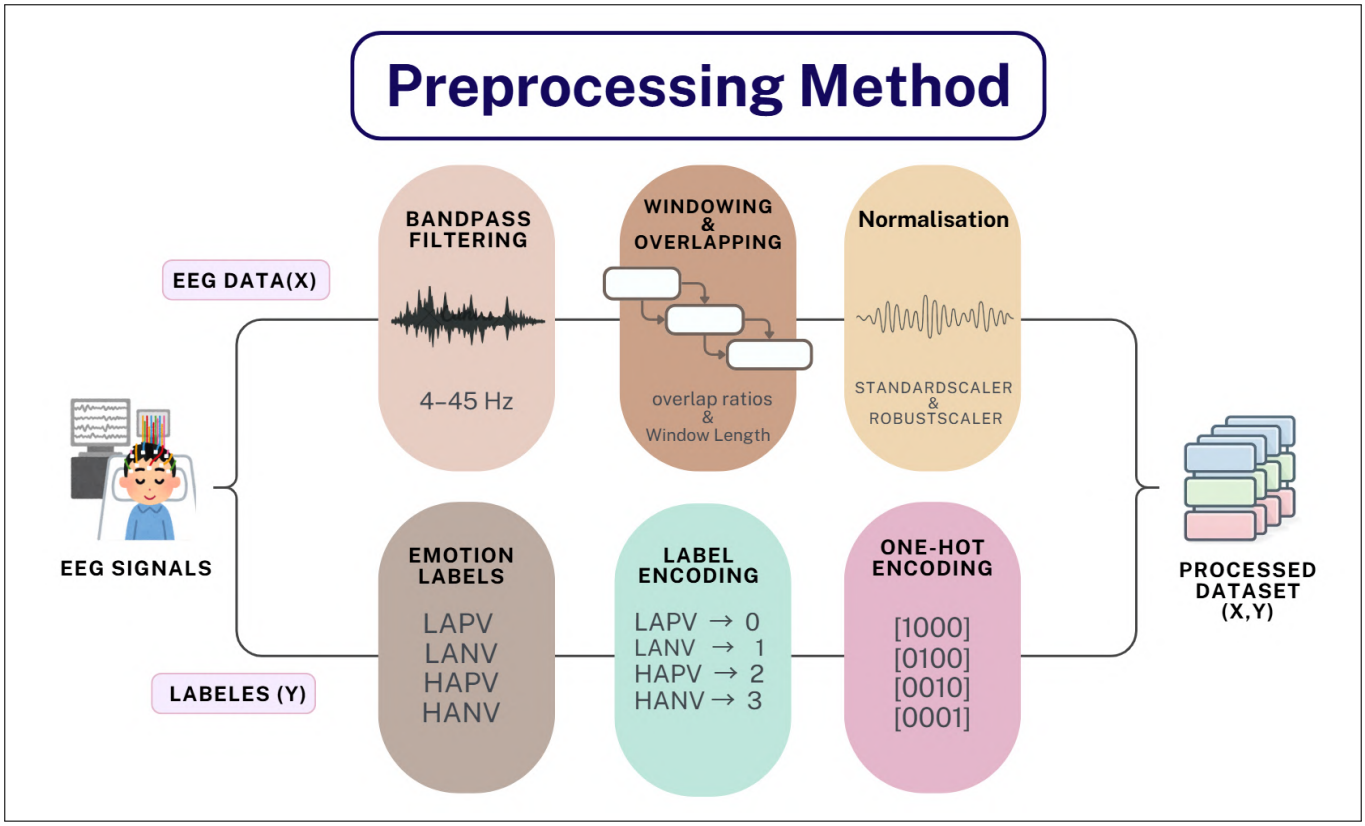


Figure 3.4: Preprocessing Pipeline for EEG Signals.

a. Bandpass Filtering

A fourth-order Butterworth bandpass filter in the 4–45 Hz frequency range is applied to the EEG signals to remove very low-frequency components (below 4 Hz) and high-frequency components (above 45 Hz), which are associated with noise, baseline drift, and muscle artifacts, while preserving the relevant brainwave components.

b. Windowing

After the filtering stage, EEG signals are divided into short time windows to analyze each segment separately rather than processing the signal as a whole. Different window lengths (1 s, 2 s, 8 s, and 10 s) are adopted to investigate the effect of window size on signal representation. After selecting the window length T, the number of samples corresponding to it is determined using the sampling rate F_s, as follows:

$$N = F_s \times T \tag{3.1}$$

Subsequently, the overlap technique is applied using different overlap ratios (50%, 75%, and 90%) to improve information continuity between consecutive windows and reduce the loss of temporal information. The overlap ratio is calculated as:

$$\text{Overlap} = 1 - \frac{\text{StepSize}}{N} \tag{3.2}$$

Finally, after determining the window size and step size, the number of windows is calculated from a signal of total length L (in samples) using the following equation:

$$N_{\text{windows}} = \frac{L - N}{\text{StepSize}} + 1 \tag{3.3}$$

Example:

Let the sampling rate be $F_s = 128$ Hz, window length $T = 2$ s, step size StepSize = 64 samples, and total signal length $L = 38358$ samples.

- **Number of samples per window (N):**

$$N = F_s \times T = 128 \times 2 = 256 \text{ samples}$$

- **Overlap calculation:**

$$\text{Overlap} = 1 - \frac{\text{StepSize}}{N} = 1 - \frac{64}{256} = 1 - 0.25 = 0.75 \ (75\%)$$

- **Total number of windows (N_{windows}):**

$$N_{\text{windows}} = \frac{L - N}{\text{StepSize}} + 1 = \frac{38358 - 256}{64} + 1 = \frac{38102}{64} + 1 \approx 595.34 + 1 \approx 596 \text{ windows}$$

Hence, the signal is divided into approximately 596 windows.

c. Normalization

Normalization is a step aimed at standardizing the scale of data by bringing values into a common range, thereby reducing the effect of differences in magnitude on the learning process. In this work, normalization is applied either after the windowing stage on the segmented signals or after the feature extraction stage on the resulting feature vectors. Two normalization methods are used:

- The first method is `StandardScaler`, and is calculated according to the following equation:

$$Z = \frac{x - \mu}{\sigma} \tag{3.4}$$

  where x is the original value, μ is the arithmetic mean, and σ is the standard deviation.

- The second method is `RobustScaler`, which is used to reduce the effect of outliers within the data. It is based on the median and interquartile range (IQR) and is calculated according to the following equations:

$$x' = \frac{x - \text{median}}{IQR} \tag{3.5}$$

  where:

$$IQR = Q_3 - Q_1 \tag{3.6}$$

d. Label Encoding

Label Encoding is used to convert categorical labels into unique numerical values for each class in the target variable, allowing them to be represented numerically for use in machine learning and deep learning models. Subsequently, One-Hot Encoding is applied to transform these numerical values into binary vectors (0 and 1) without imposing any ordinal relationship between the classes.

3.6.2 Feature Extraction for Machine Learning and Ensemble Learning Models

a. Feature Extraction

In the context of machine learning models, a set of features is extracted from each time window. These features include Time-Domain, Frequency-Domain, Time-Frequency Domain, and Non-Linear features.

- In **Non-Linear Features**, entropy-based features such as Shannon Entropy, Spectral Entropy, Log Energy, and Differential Entropy are extracted to characterize the degree of randomness and complexity of the signal.
- In the **Time-Domain**, statistical features such as minimum and maximum values (min, max) and standard deviation are calculated, in addition to Hjorth parameters (Activity, Mobility, and Complexity).
- In the **Frequency-Domain**, features are extracted using FFT and PSD, including dominant frequency, mean amplitude, and signal energy within frequency bands (Theta, Alpha, Beta, Gamma).
- As the **Time-Frequency Domain**, the Discrete Wavelet Transform (DWT) is used to decompose the EEG signal into an approximation component (A), representing the overall pattern, and detail components (D1 to D4), representing signal details at different levels. This decomposition enables a more comprehensive representation of the signal in both the time and frequency domains.

b. Feature Grouping Strategy

After extracting the features, they are organized into different groups to investigate the effect of each type of representation on the performance of machine learning models. The number of features in each group is summarized in Table **3.3**, where this value represents the length of the feature vector used as input to the machine learning models.

Table 3.3: Feature Grouping Strategy and Number of Features per Group.

Group Name	Description	Num_Features
Non-Linear-Mix	Shannon Entropy, Spectral Entropy, Log Energy, Differential Entropy (DE)	56
Time-Domain-Mix	Statistical (min, max, std), Hjorth (Activity, Mobility, Complexity)	84
Frequency-Mix	FFT, PSD	84
Wavelet-Energy	DWT energy features (A, D1, D2, D3, D4)	70
Elite-Melange	Best feature from each group (DE, Hjorth, PSD)	112
Total-AIO (All In One)	All extracted features	308

3.6.3 Feature Extraction for Deep Learning and Hybrid Models

A set of features is extracted from EEG signals to represent them at different levels, including the global level, the local level, and the spatial-spectral level.

a. Global Feature Extraction

This level relies on representing global characteristics of the EEG signal using spectral entropy and basic statistical features such as maximum (Max), minimum (Min), and mean (Mean) values. In addition, power spectral density (PSD) is used in the computation of differential entropy (DE).

b. Local Feature Extraction using EMD & VMD

At this level, EEG signals are decomposed into intrinsic components using empirical mode decomposition (EMD) and variational mode decomposition (VMD). Statistical features (Max, Min, Mean) are then extracted from each component to describe the signal's behavior more precisely at the local level.

c. Spatial–Spectral Representation (HVFlow)

This representation divides EEG signals into different frequency bands (delta, theta, alpha, beta, gamma). Within each band, differential entropy (DE) is computed for each channel and projected onto a two-dimensional grid representing the spatial distribution of EEG electrodes on the scalp, in order to preserve the spatial structure of the signal. The horizontal and vertical flows are then computed to capture spatial variations, and all bands are integrated into a unified representation called HV-Flow, integrating spectral, spatial, and local information.

3.6.4 Machine Learning-Based Models

As shown in the Figure **3.5**, several machine learning algorithms are used for EEG signal classification, using the same set of features previously extracted. This study includes several models: KNN, SVM (LinearSVC), Random Forest, XGBoost, LDA, ExtraTrees, and LightGBM.

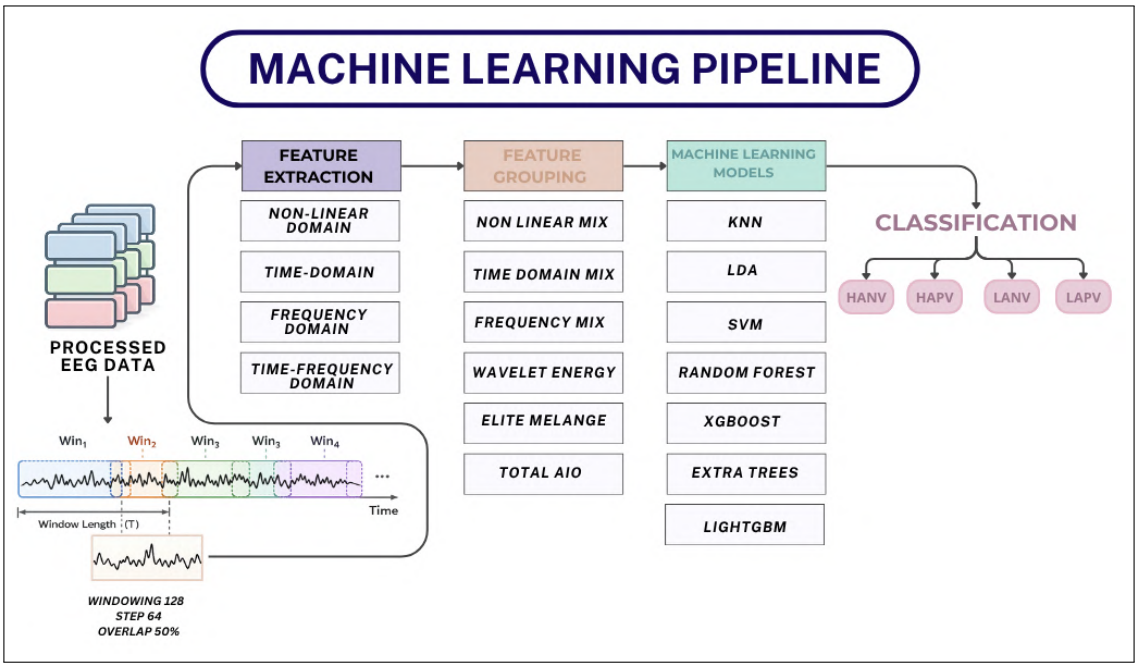


Figure 3.5: Proposed Machine Learning Pipeline Based on Feature Extraction.

3.6.5 Deep Learning-Based Models

A set of deep learning models was adopted for EEG signal classification, as illustrated in Figure **3.6**. Pre-extracted features are used as direct inputs instead of time windows, providing the models with information-rich representations. This approach enables the models to focus on learning complex patterns within the data and improve classification performance.

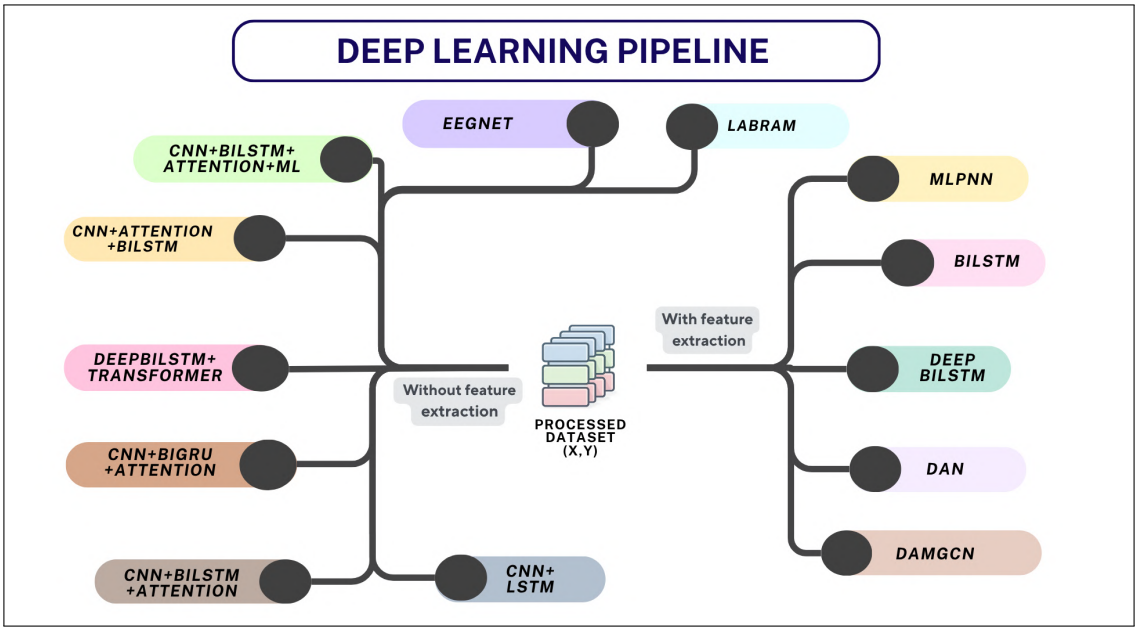


Figure 3.6: Overview of the Proposed Deep Learning and Hybrid Models.

a. MLPNN

The MLPNN model relies on EEG features extracted using PSD, where the input consists of 70 features representing channel responses across different frequency bands. These features are passed through two hidden layers composed of 10 neurons per layer with ReLU activation to extract non-linear patterns and more discriminative representations. Finally, the outputs are passed to an output layer based on Softmax to determine the final emotional class, as shown in Figure **3.7**.

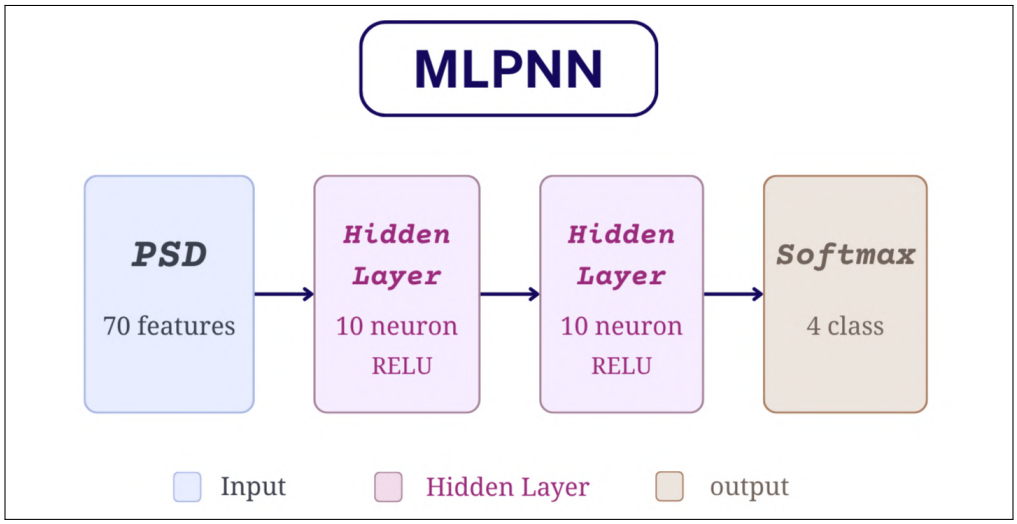


Figure 3.7: Illustrated Architecture of the Proposed MLPNN Model.

b. BiLSTM

The BiLSTM model consists of an input vector composed of 14 features extracted using Spectral Entropy. These features are passed to a BiLSTM layer with return_sequences=False to learn richer and more discriminative representations of the input features through bidirectional processing. This is followed by Batch Normalization to enhance the stability of the representation, then Dropout to reduce overfitting, and then a Dense layer to extract deeper and more distinctive representations. Finally, classification is performed via the output layer to determine the final emotional class, as shown in Figure **3.8**.

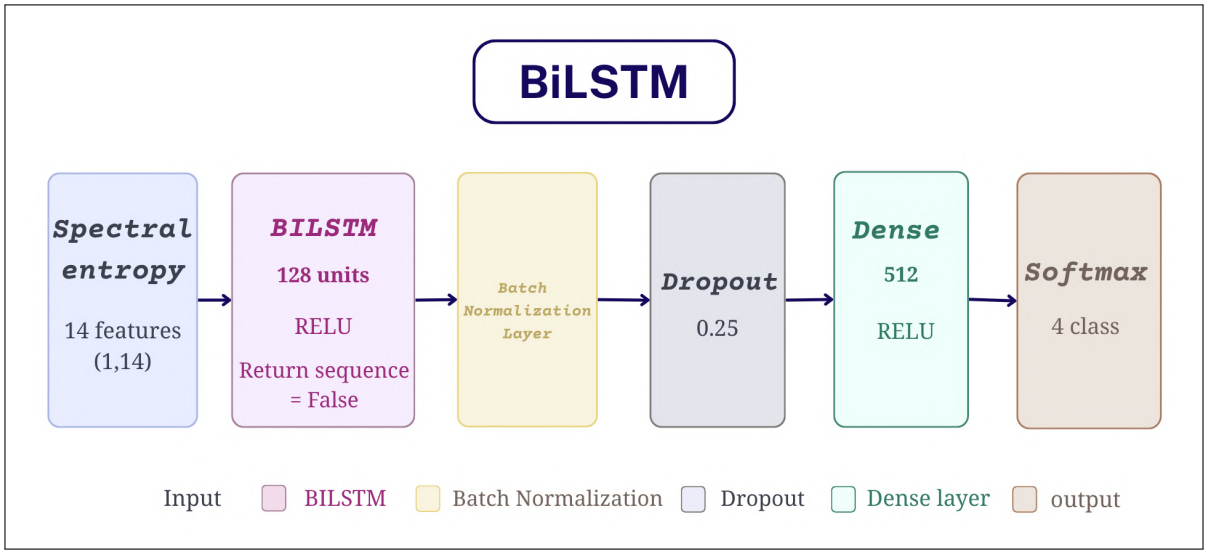


Figure 3.8: Illustrated Architecture of the Proposed BiLSTM Model.

c. Deep BiLSTM

As shown in Figure **3.9**, the Deep BiLSTM model relies on three different feature representations, including 14 features extracted using Spectral Entropy, 42 simple statistical features, and a rich representation combining signal analysis using EMD and VMD with extracted statistical features, totaling 588 features. Each representation is independently passed to the same model, which consists of three sequential bidirectional (Bidirectional LSTM) layers. The first and second layers use return_sequences=True to preserve intermediate representations and enable progressive learning of richer and more discriminative feature representations, while the third layer uses return_sequences=False to generate a compact representation of the input features. This is followed by Batch Normalization and Dropout, then two Dense layers to extract deeper and more discriminative representations. Finally, the output layer is used to determine the final emotional class.

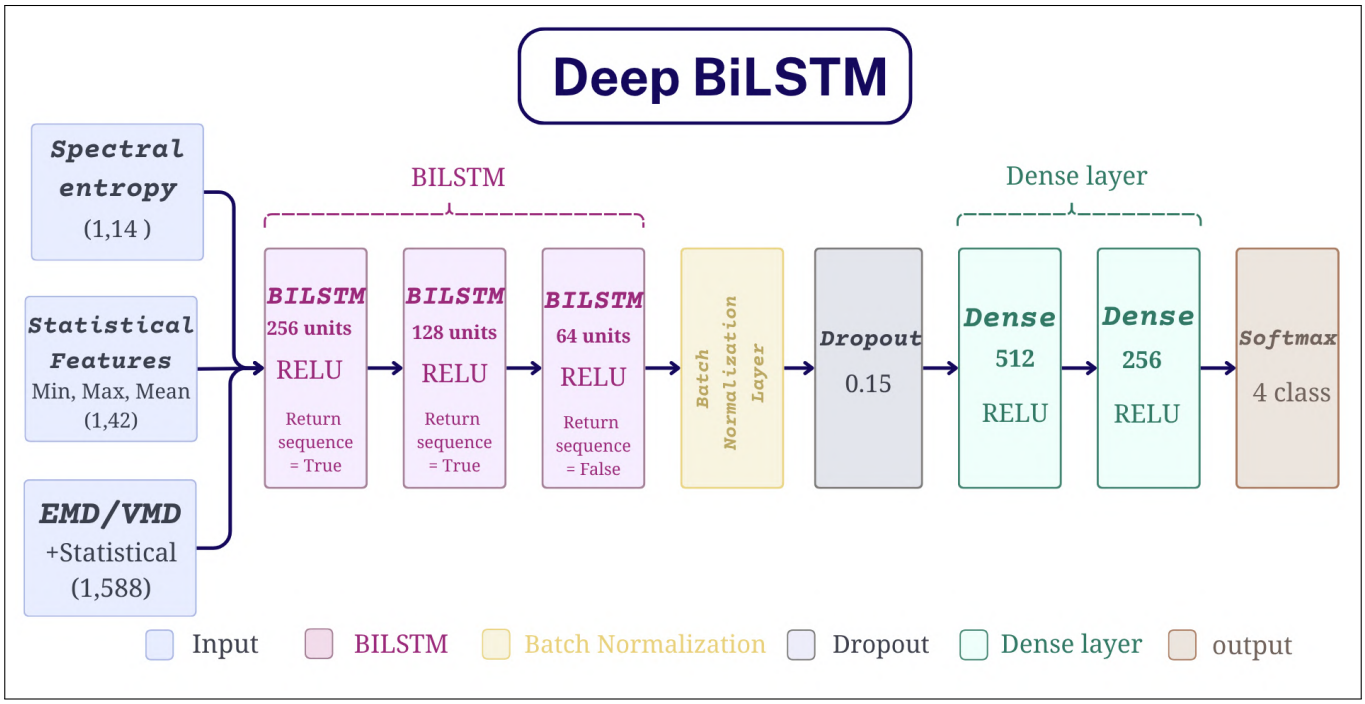


Figure 3.9: Illustrated Architecture of the Proposed Deep BiLSTM Model.

3.6.6 Ensemble Learning-Based Models

In order to improve the performance and reliability of EEG signal classification, ensemble learning techniques are adopted to combine the outputs of several machine learning models trained on different sets of features extracted from the signals. This approach aims to leverage the complementary strengths of various classifiers while reducing the limitations of individual models and enhancing their generalization ability. This is achieved through a set of main techniques illustrated in the following Figure **3.10**.

a. Local Stacking

The local stacking strategy is applied by integrating four machine learning models: XGBoost, LightGBM, Random Forest, and ExtraTrees. Each feature group is handled independently, where the four models are trained and their prediction probabilities (predict_proba) are extracted and combined

into a single representation specific to each group. Subsequently, a final meta-classifier based on logistic regression is trained to make the final decision for each group.

b. Global Stacking (Fusion)

This technique involves combining the prediction probabilities resulting from Local Stacking for each feature group into a single representation. A final model based on logistic regression is then trained to produce the system's final classification.

c. Majority Voting

This technique combines the final decisions (labels) produced by Local Stacking for each feature group, utilizing a voting mechanism among them, where the category obtaining the highest number of votes is selected as the final classification for each sample, without requiring any additional training.

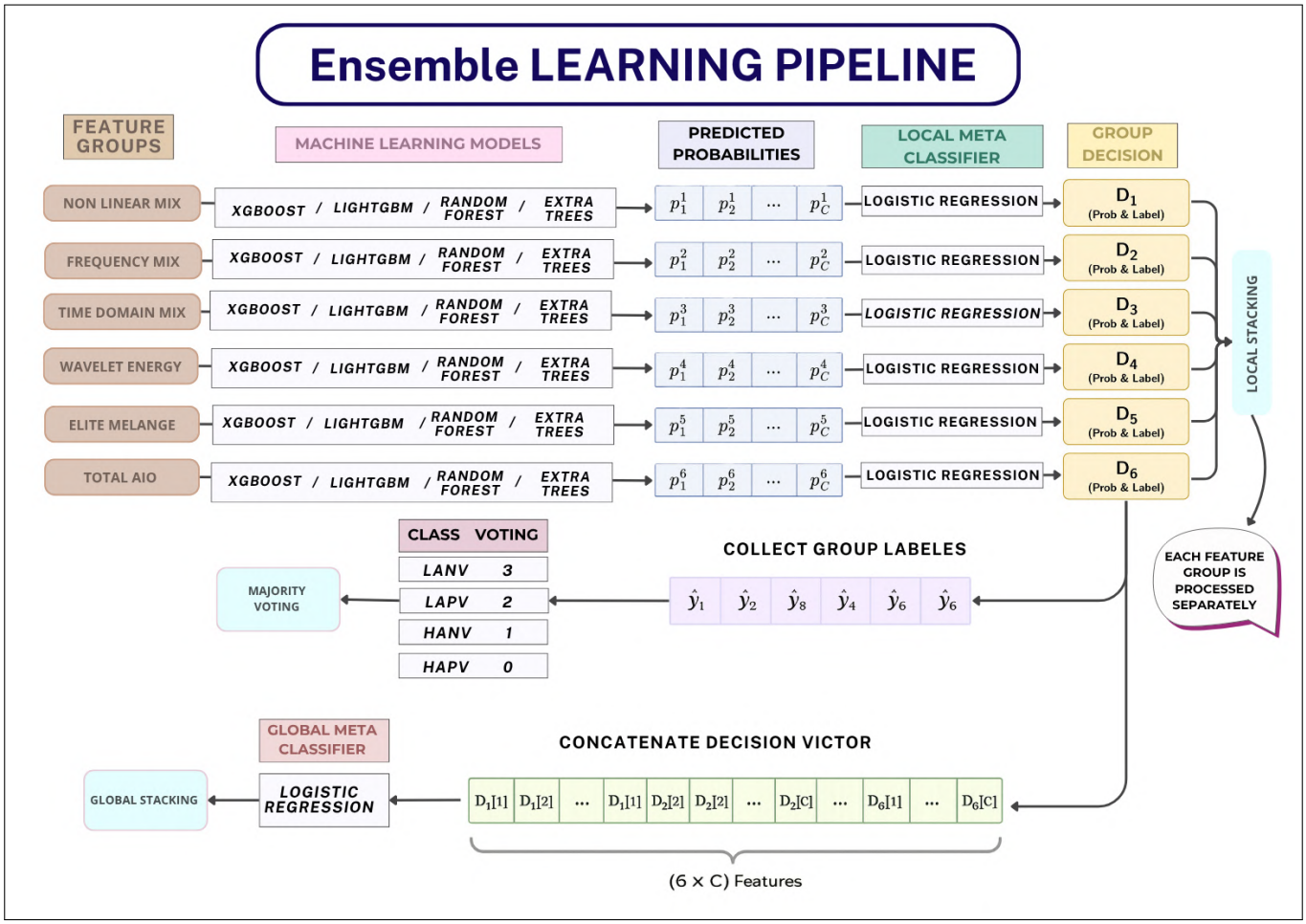


Figure 3.10: Overview of Ensemble Learning Approaches: Local Stacking, Global Stacking (Fusion), and Majority Voting.

3.6.7 Hybrid Deep Learning and Ensemble Learning Models

As illustrated in Figure **3.6**, hybrid deep learning models are used for EEG signal classification. Some models directly use time windows as input to deep neural networks due to their ability to automatically learn representations from raw data and extract non-linear patterns. In contrast, models such as DAN and DAMGCNN rely on pre-extracted features as input, aiming to improve data representation and enhance classification accuracy across different emotional classes.

a. CNN + LSTM

The CNN-LSTM model, as shown in Figure **3.11**, relies on input data in the form of time windows with dimensions (512 × 14). The data is first passed through two Conv1D layers to extract local features from the signal, with each layer followed by Batch Normalization to improve training stability, followed by MaxPooling1D to reduce dimensionality while retaining the most important features. After that, the extracted representations are then passed to LSTM layers to capture long-term temporal dependencies in the signal, with Dropout applied after each layer to reduce overfitting. Finally, the features are passed through two Fully Connected (Dense) layers, ending with a Softmax layer to produce the final multi-class prediction.

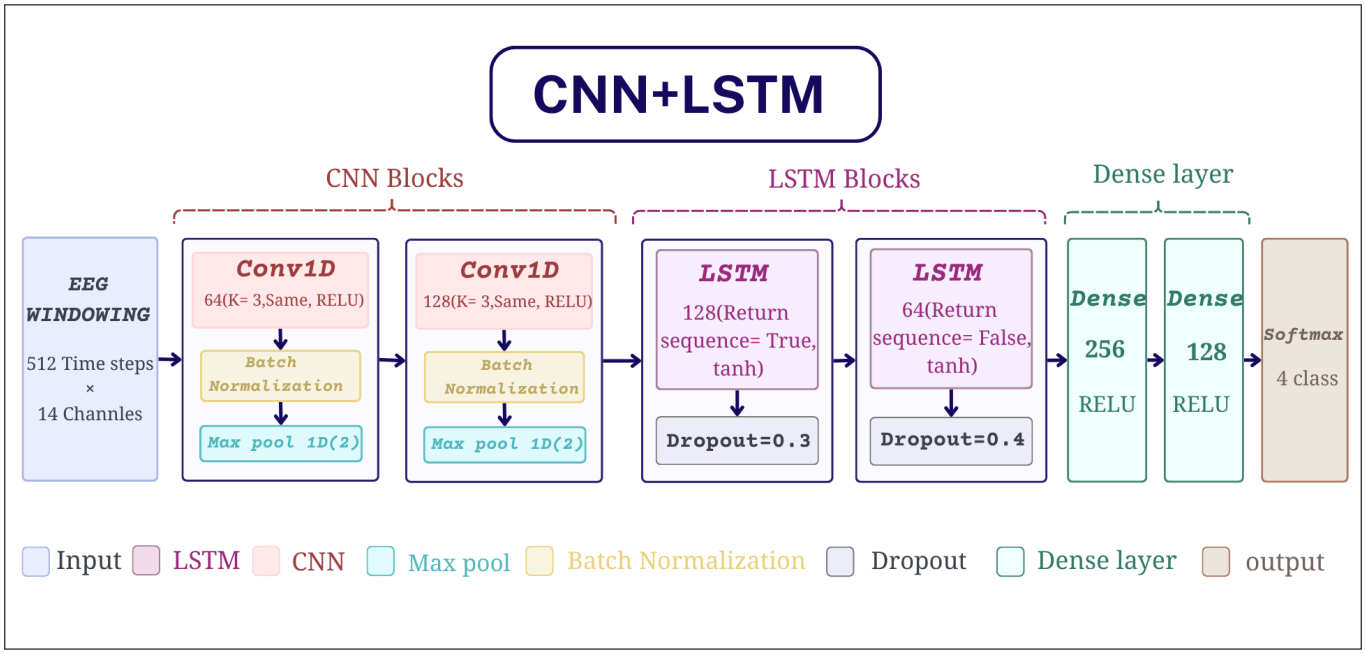


Figure 3.11: Illustrated Architecture of the Proposed CNN + LSTM Model.

b. CNN + BiLSTM + MHA

The CNN-BiLSTM-MHA model consists of two sequences of consecutive Conv1D layers, with each layer followed by Batch Normalization and then MaxPooling1D. The extracted representations are then passed to a Bidirectional LSTM layer to capture temporal dependencies in both forward and backward directions, with Dropout applied. Then, a Multi-Head Attention mechanism is applied to reweight temporal relationships within the sequence and enhance important representations. Its outputs are merged with the BiLSTM outputs via an addition operation (Residual Connection), followed by Layer Normalization to improve representational stability. In the final stage, Global

Average Pooling is used, followed by a Dense layer for classification with Softmax to produce the final outputs, as shown in Figure **3.12**.

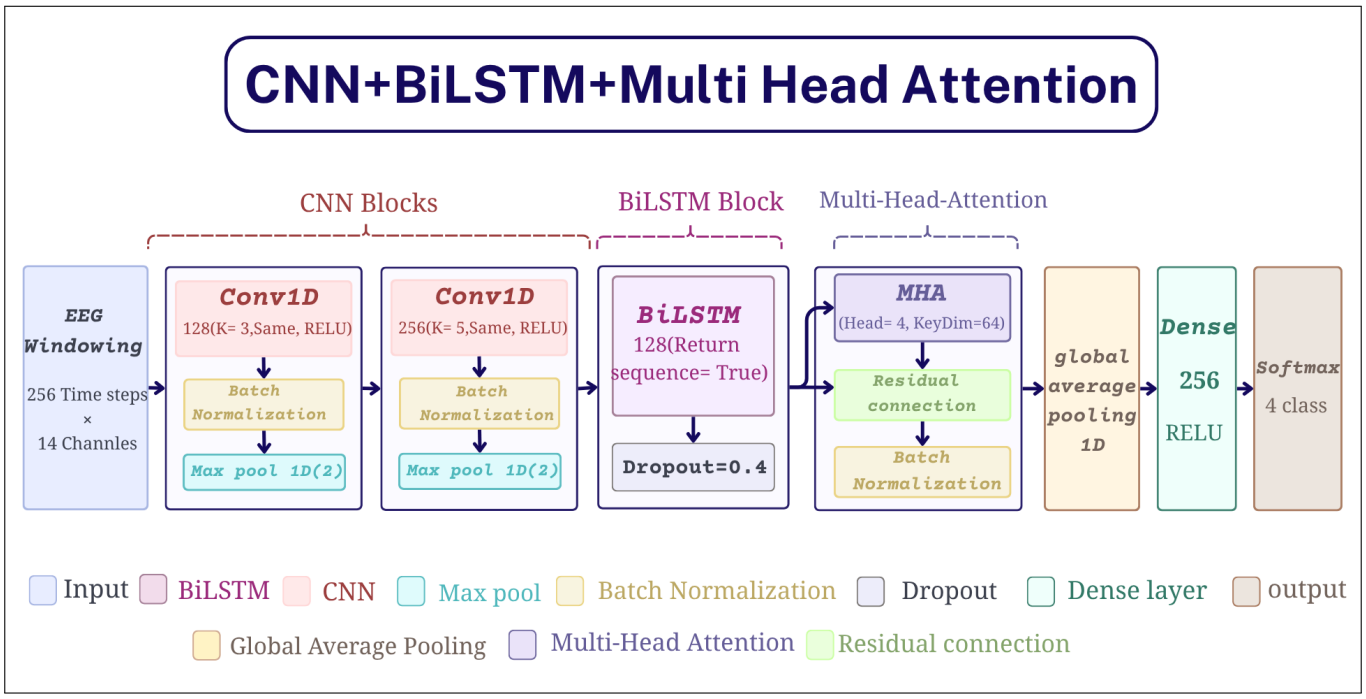


Figure 3.12: Illustrated Architecture of the Proposed CNN + BiLSTM + MHA Model.

c. CNN + BiGRU + MHA

The CNN-BiGRU-MHA model follows the same general architecture based on CNN and the Multi-Head Attention mechanism, with a modification in the temporal modeling unit, where the Bidirectional LSTM layer is replaced by a Bidirectional GRU layer composed of 128 units with return_sequences=True (See figure **3.13**).

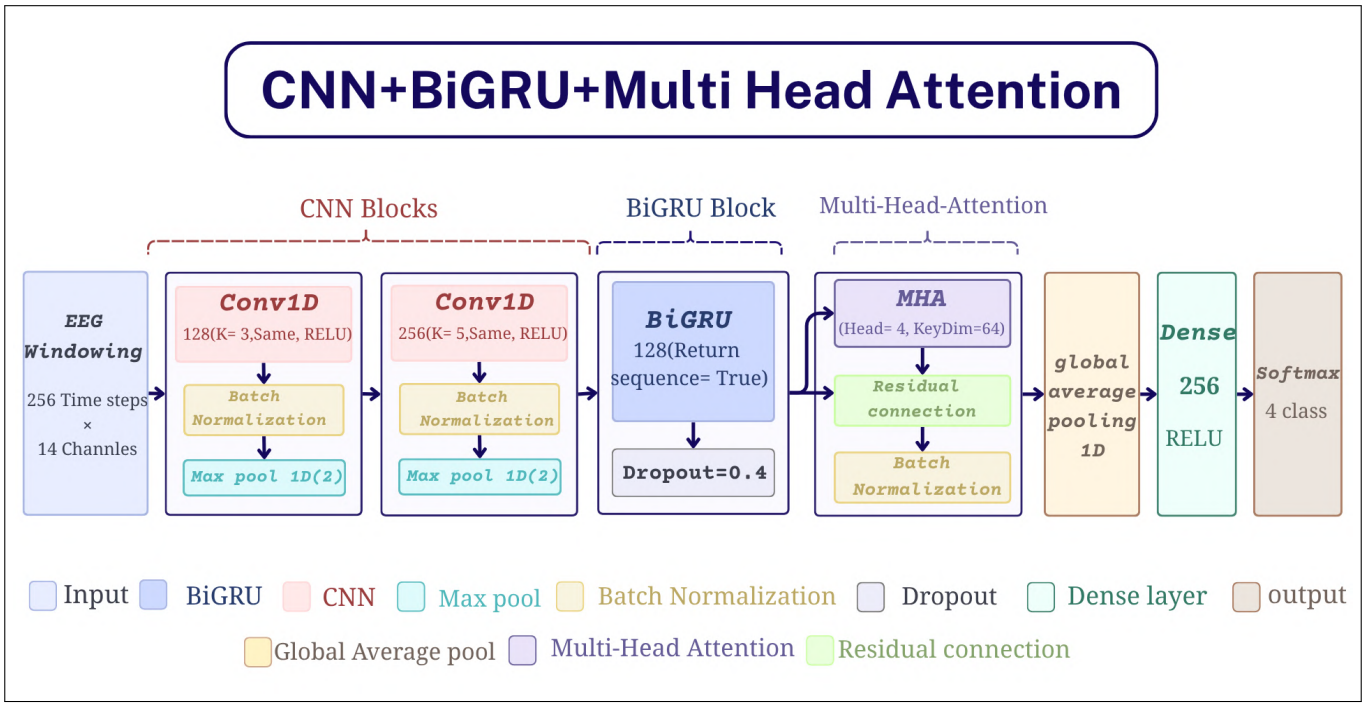


Figure 3.13: Illustrated Architecture of the Proposed CNN + BiGRU + MHA Model.

d. Transformer + DeepBiLSTM

The proposed model consists of a hybrid architecture combining Transformer and BiLSTM to process temporal data and extract deep representations. Time-series data with dimensions (256 × 14) are first transformed into a vector representation through a Dense Projection layer of dimension 128, where the features are mapped into a richer and more discriminative feature space, enabling the model to better capture relationships within the data. This representation is then combined with Positional Encoding to incorporate temporal information. Next, the data passes through two Transformer Encoder layers, which rely on the Multi-Head Attention mechanism to capture global dependencies between time steps, along with Feed Forward Networks, Residual Connections, and Layer Normalization to improve learning stability. The resulting representation is subsequently passed through three BiLSTM layers to capture temporal dependencies in both forward and backward directions. Finally, Fully Connected layers with Softmax activation are used to produce the final multi-class classification.

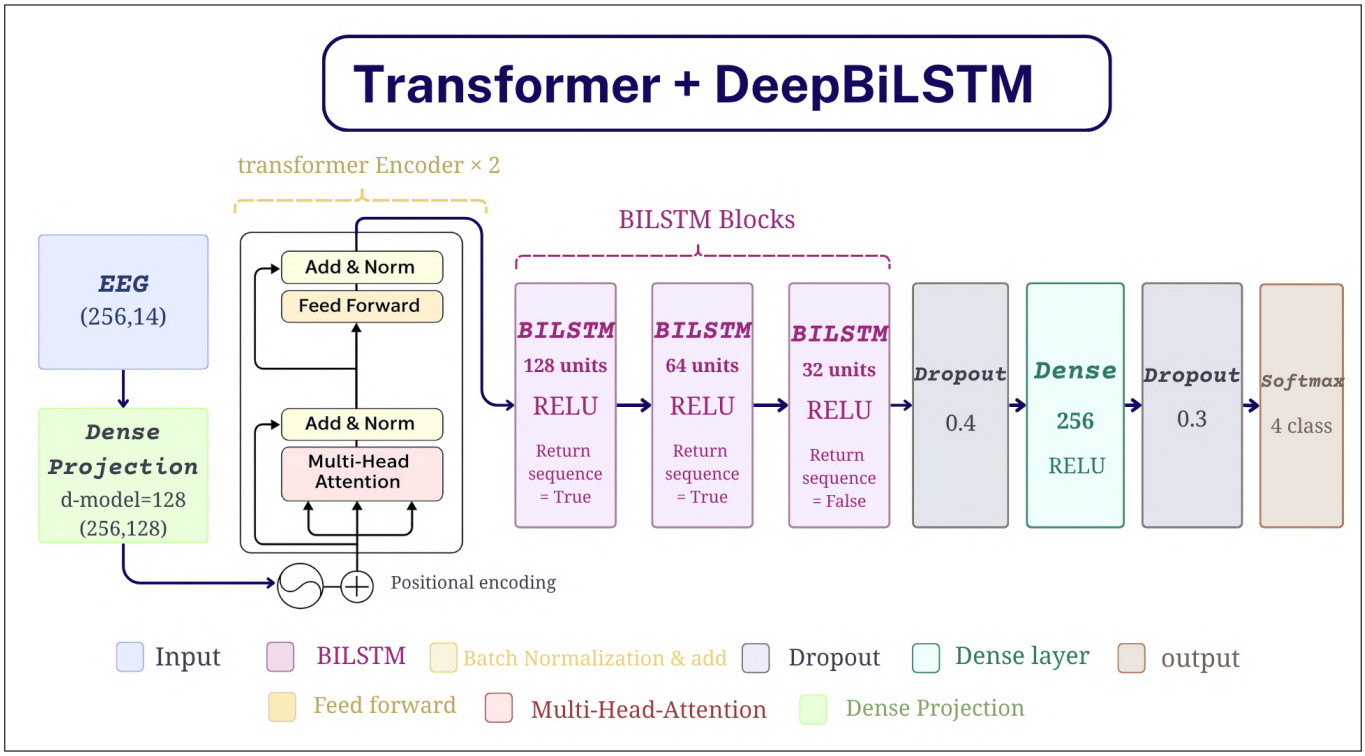


Figure 3.14: Illustrated Architecture of the Proposed Transformer + Deep BiLSTM Model.

e. CNN + MHA + BiLSTM

The proposed model consists of a hybrid architecture combining CNN, Multi-Head Attention, and BiLSTM to extract deep representations from temporal data. The data are first passed through three Conv1D layers, followed by MaxPooling, Batch Normalization, and Dropout layers to extract initial features. The resulting representation is then passed through two consecutive attention blocks. Within each block, Multi-Head Attention is applied to the same representation, and a Residual Connection is added between the block input and the attention output, followed by Layer Normalization. A Dense layer with Dropout is then applied to further enhance the representation. Afterwards, the resulting representation is passed through two consecutive BiLSTM layers with Dropout to capture temporal dependencies in both forward and backward directions. Finally, a Softmax layer is used to produce the final multi-class classification.

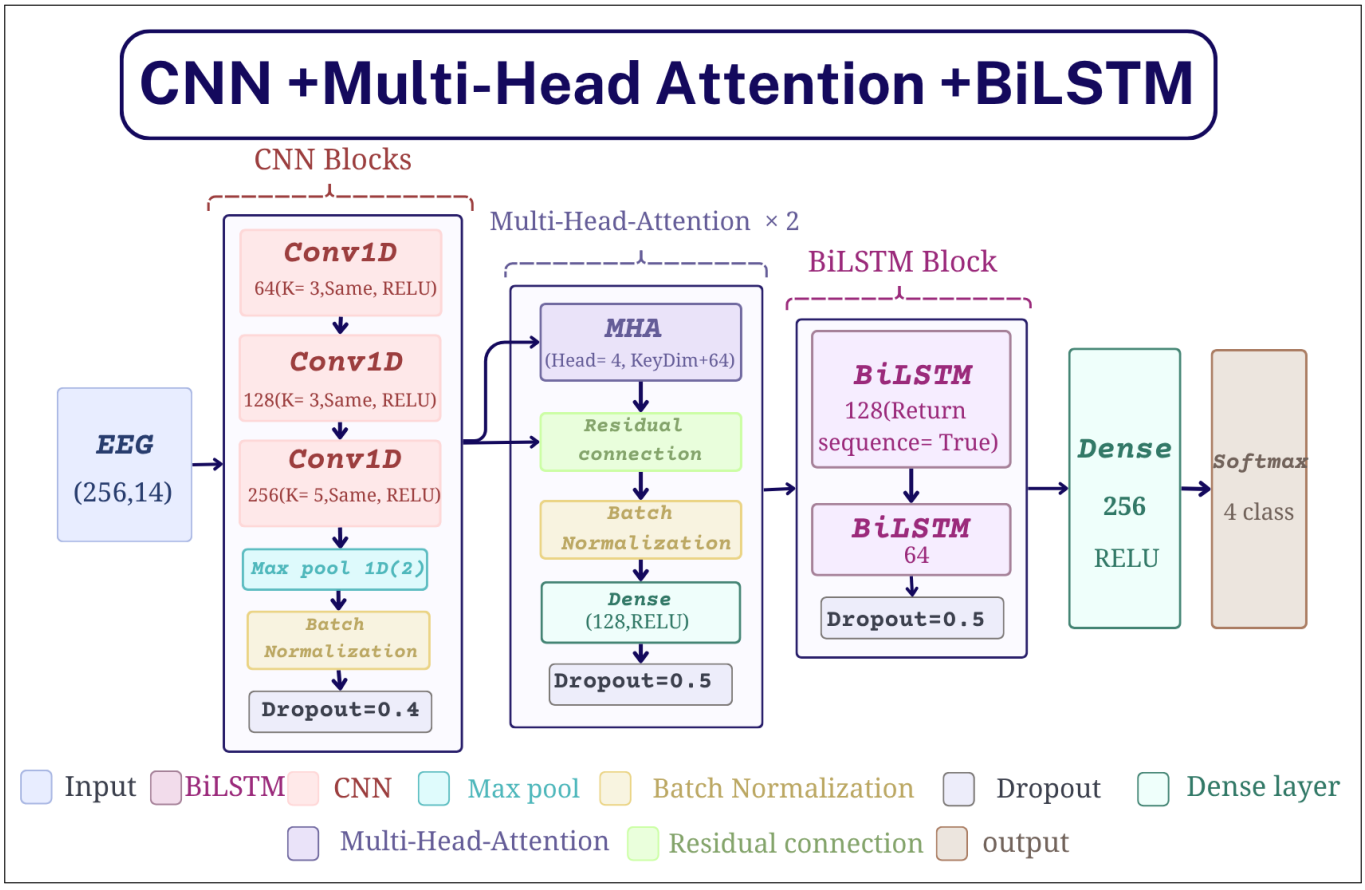


Figure 3.15: Illustrated Architecture of the Proposed CNN + MHA + BiLSTM Model.

f. CNN + MHA + BiLSTM + Ensemble EL

The proposed methodology is based on a hybrid framework that combines deep learning and machine learning to improve EEG signal classification. In the first stage, the previously proposed CNN–MHA–BiLSTM model is adopted as a fixed feature extractor using the Freezing Strategy, allowing the extraction of deep feature representations for each time window that capture spatial and temporal characteristics as well as complex relationships within the signals. In parallel, Differential Entropy (DE) features are extracted to represent the spectral characteristics of the signal across different frequency bands. The deep representations extracted by the model are then combined with the DE features into a unified feature vector using concatenation. Finally, the resulting feature vector is passed to an ensemble machine learning classifier composed of SVM, XGBoost, and LightGBM models. A Soft Voting strategy is then employed to produce the final classification decision.

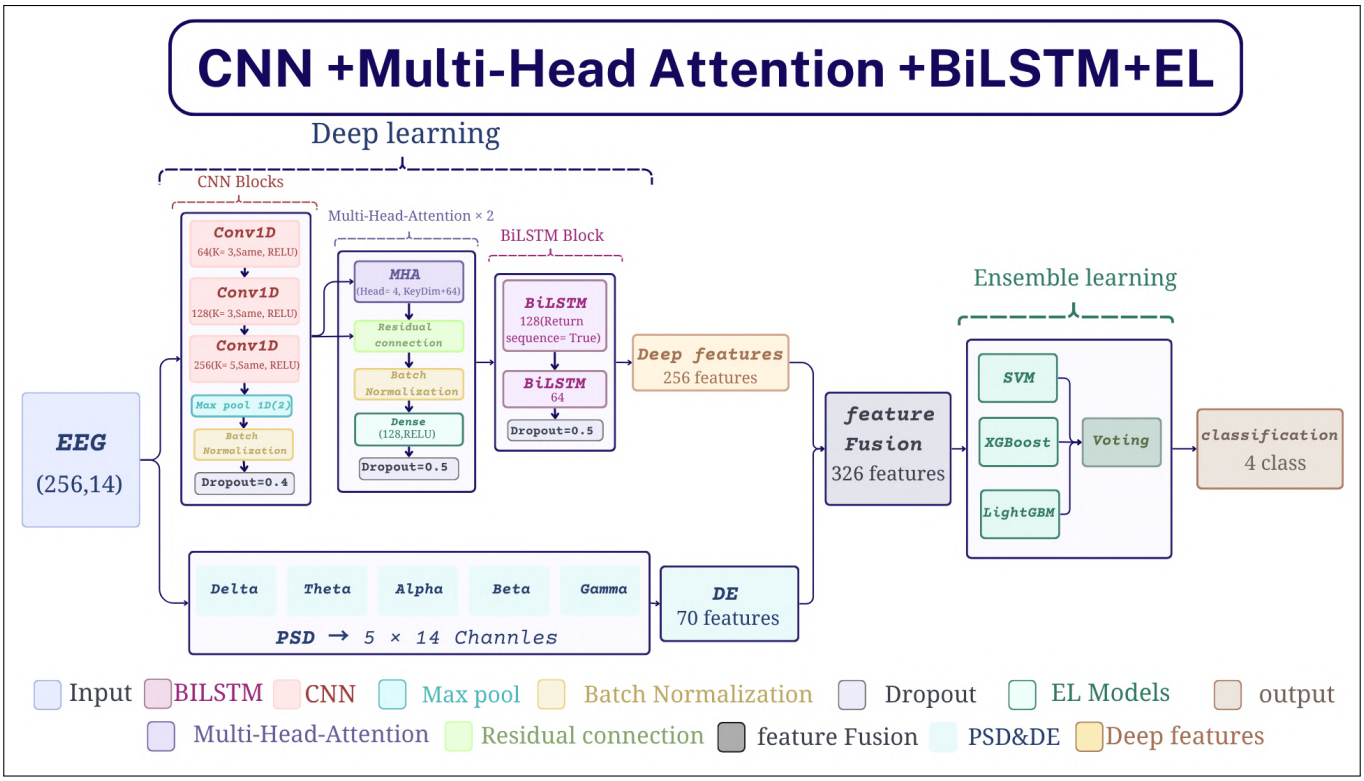


Figure 3.16: Illustrated Architecture of the Proposed CNN + MHA + BiLSTM + Ensemble Learning Model.

g. EEGNet + Attention

The EEGNet-Enhanced model consists of a convolutional neural network (CNN) for analyzing EEG signals, and comprises three main blocks followed by a final classification layer. It begins with a Block Convolution that relies on Temporal Conv2D to extract temporal patterns, followed by Batch Normalization and Depthwise Convolution to learn spatial relationships between channels, then a second Batch Normalization and the ELU activation function to introduce non-linearity, with Average Pooling to reduce dimensions and Dropout to limit overfitting. This is followed by a Block Separable Convolution that uses SeparableConv2D to separate the feature extraction process into two parts: the first part (Depthwise) to extract temporal features more precisely, and the second part (Pointwise 1×1 Convolution) to combine features across channels and improve representation while reducing the number of parameters. Then, Batch Normalization and ELU are applied, followed by Average Pooling and Dropout. As for Block Attention, it is a stage for re-weighting feature maps instead of extracting them, where Conv2D (1×1) is used to produce channel-wise feature recalibration weights that represent the importance of channels and the spatial representations resulting from feature fusion, after which these values pass through the Sigmoid function to transform them into weights between 0 and 1 representing the importance level of each feature. Subsequently, the attention map is multiplied with the outputs of the Block Separable Convolution via element-wise multiplication to re-weight features, enhance important channels, and reduce less important features. Then, the output is passed to a Flatten layer and then a Dense layer with Softmax to produce the probability distribution of the final classes.

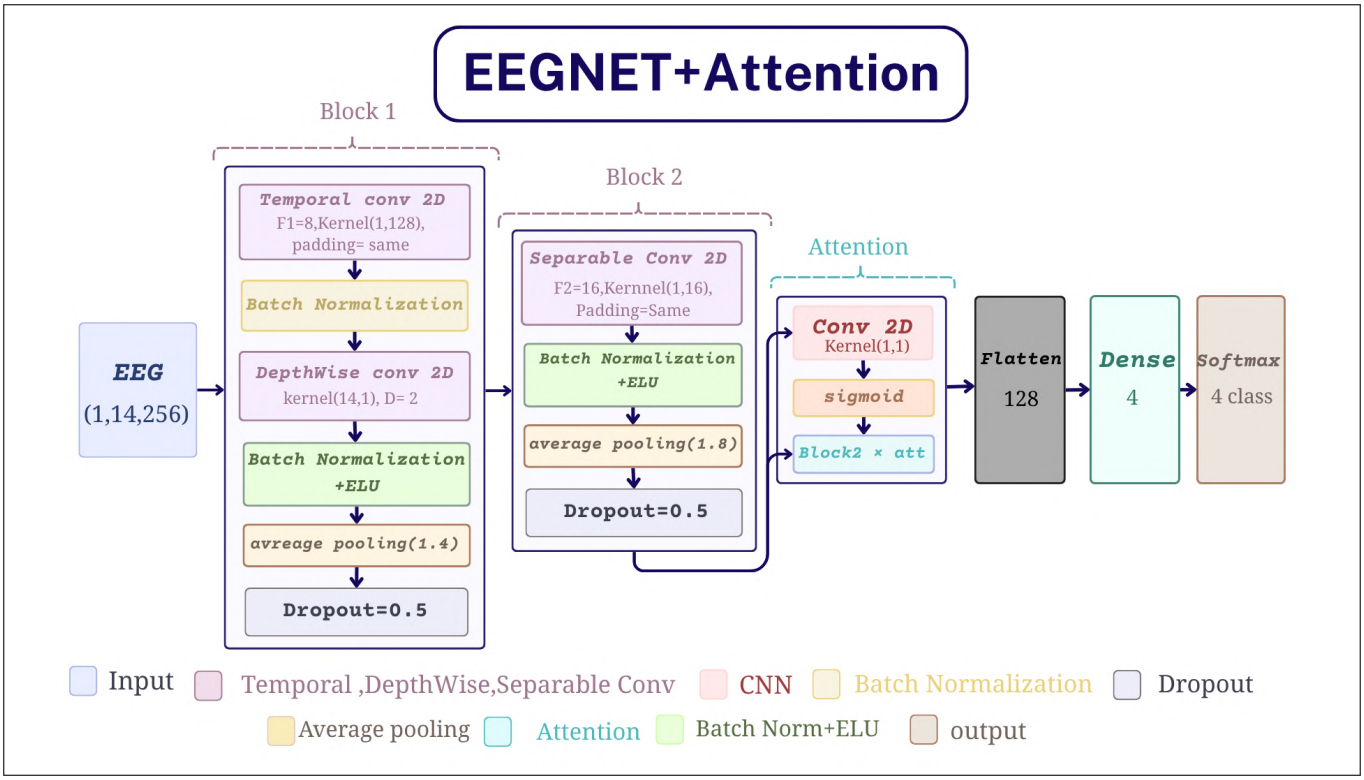


Figure 3.17: Illustrated Architecture of the Proposed EEGNet + Attention Model.

h. DAN

The DAN (Domain-Adversarial Neural Network) model relies on the HV-Flow extracted from EEG signals and represented as a spatial–spectral feature map, which passes through two Convolution2D layers followed by Batch Normalization, ReLU, and MaxPooling to extract a deep representation. This representation is then transformed into a one-dimensional vector by Flatten to be used as a shared representation that feeds into two parallel branches. The first branch specializes in emotion classification and produces the probability distribution over emotional categories, while the second branch is dedicated to domain classification, where it determines the identity of the person from the signal. This branch is integrated within an adversarial learning framework to reduce the dependence of the shared representation on individual-specific information. To achieve this, a Gradient Reversal Layer (GRL) is inserted before the gradients from the domain branch are backpropagated, where it reverses the direction of the gradient during the backpropagation phase with a coefficient α, encouraging the shared representation to ignore identity-related information and focus only on emotion-related features.

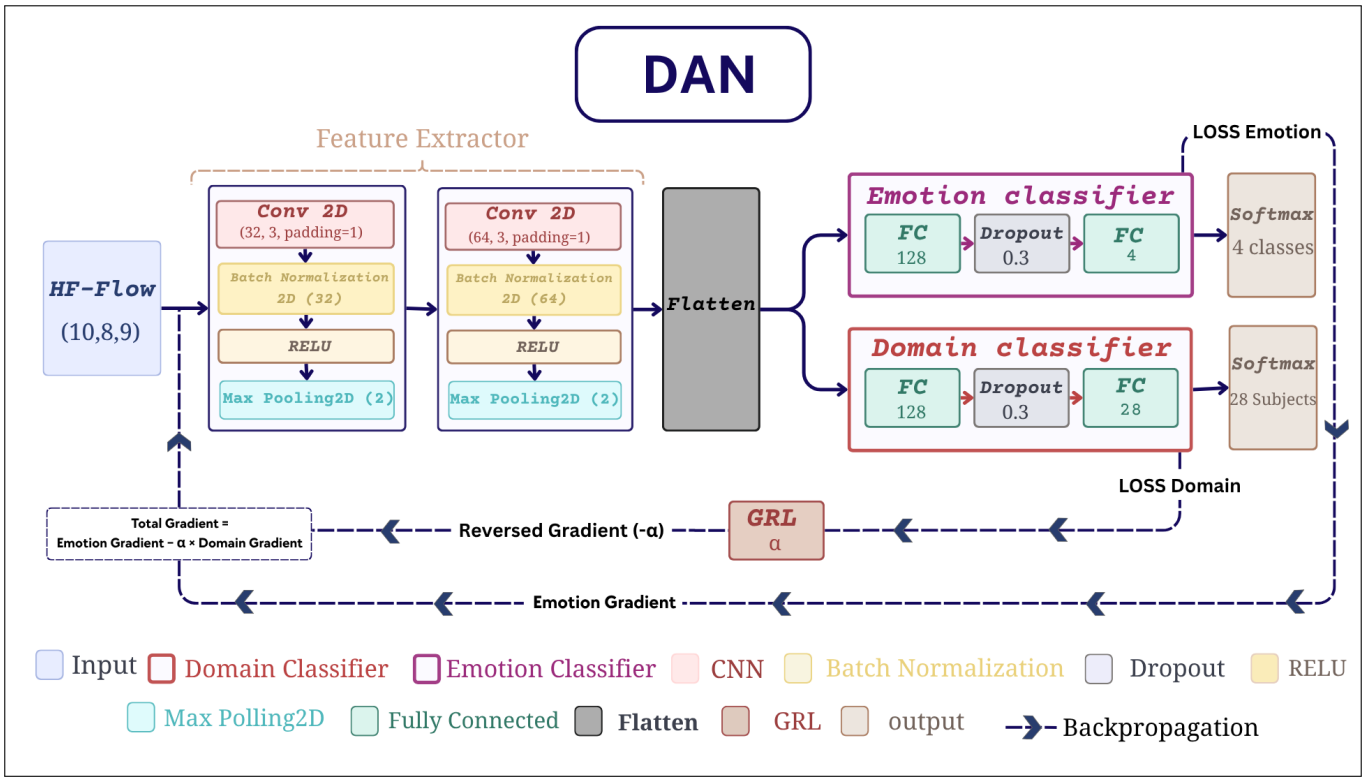


Figure 3.18: Illustrated Architecture of the Proposed DAN Model.

i. DAMGCNN + Attention

The DAMGCNN model relies on extracting Differential Entropy (DE) features from EEG signals across four frequency bands, resulting in a feature matrix x with dimensions (14 × 4), which is used to construct a Fully Connected Graph, leading to the creation of an edge_index containing 182 edges (14 × 13). Both x and edge_index are used as inputs to a GCN architecture composed of two successive layers; the first layer increases each node's representation from 4 to 64 features, while the second layer preserves the same dimension to enhance neighborhood information integration and extract a deeper and more discriminative representation. Channel-wise Attention is then applied by compressing node representations from 64 to 16 and then expanding them back to 64 using a Sigmoid activation to generate attention weights, which are multiplied element-wise with the GCN outputs to highlight the most important features. Finally, global mean pooling is used to aggregate all node representations into a single vector representing the overall brain state, which is passed to an MLP classifier to produce the probabilistic distribution of emotional categories, with the category of highest probability selected as the final classification.

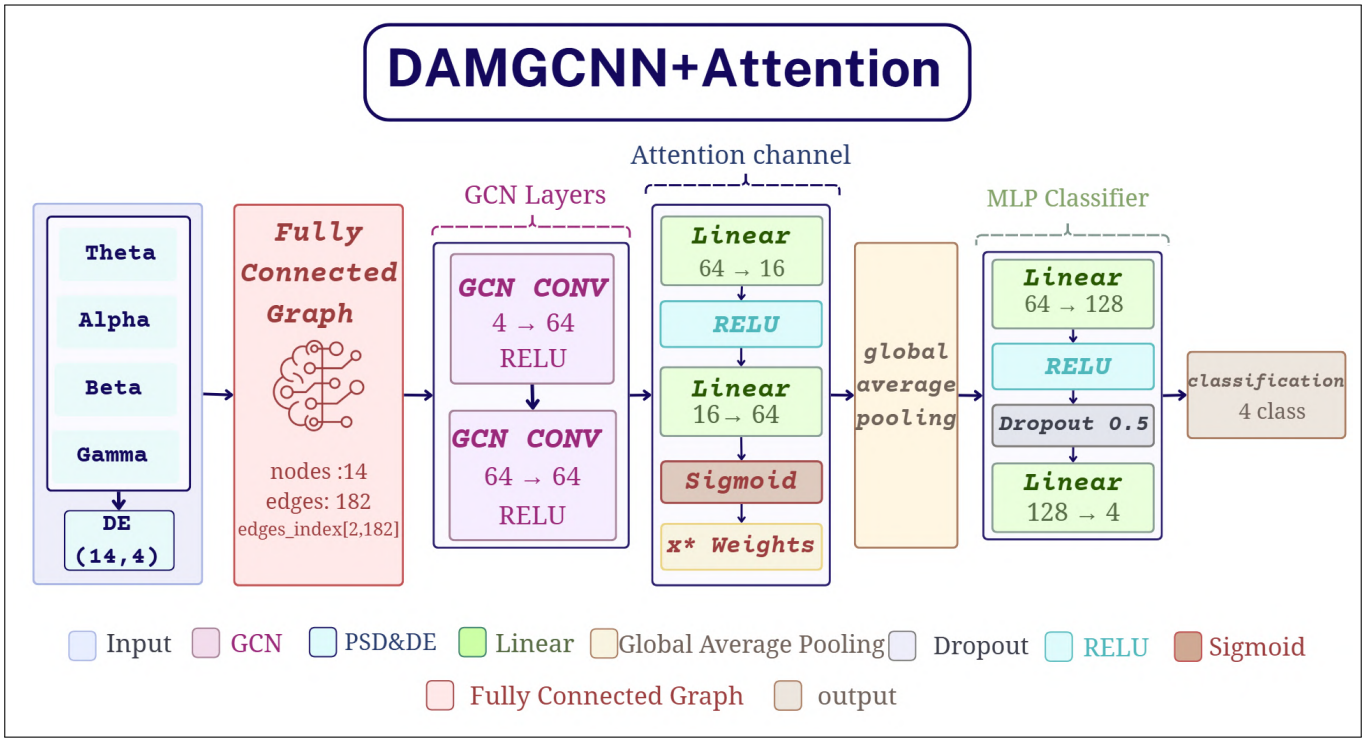


Figure 3.19: Illustrated Architecture of the Proposed DAMGCNN + Attention Model.

j. LaBraM

The LaBraM (Large Brain Model) is used, which is a foundation model based on a Transformer architecture customized for EEG signals. It is pre-trained on approximately 2,500 hours of EEG data from several diverse databases, using a self-supervised learning approach that enables the model to learn general representations of brain signals by reconstructing hidden parts of the signal. In this context, the model is adopted within a transfer learning framework for emotion classification from EEG signals, using an input of shape (14 channels × 512 samples). In the first stage, all model layers are frozen, and only the classification layer is trained without modifying the original weights, with the aim of adapting the general representations to the emotion classification task. Subsequently, the resulting weights are used as a starting point for the second stage, in which full fine-tuning is performed by unfreezing all model layers and updating all parameters, allowing the representations to be optimized to better fit the task data and enhance final performance.

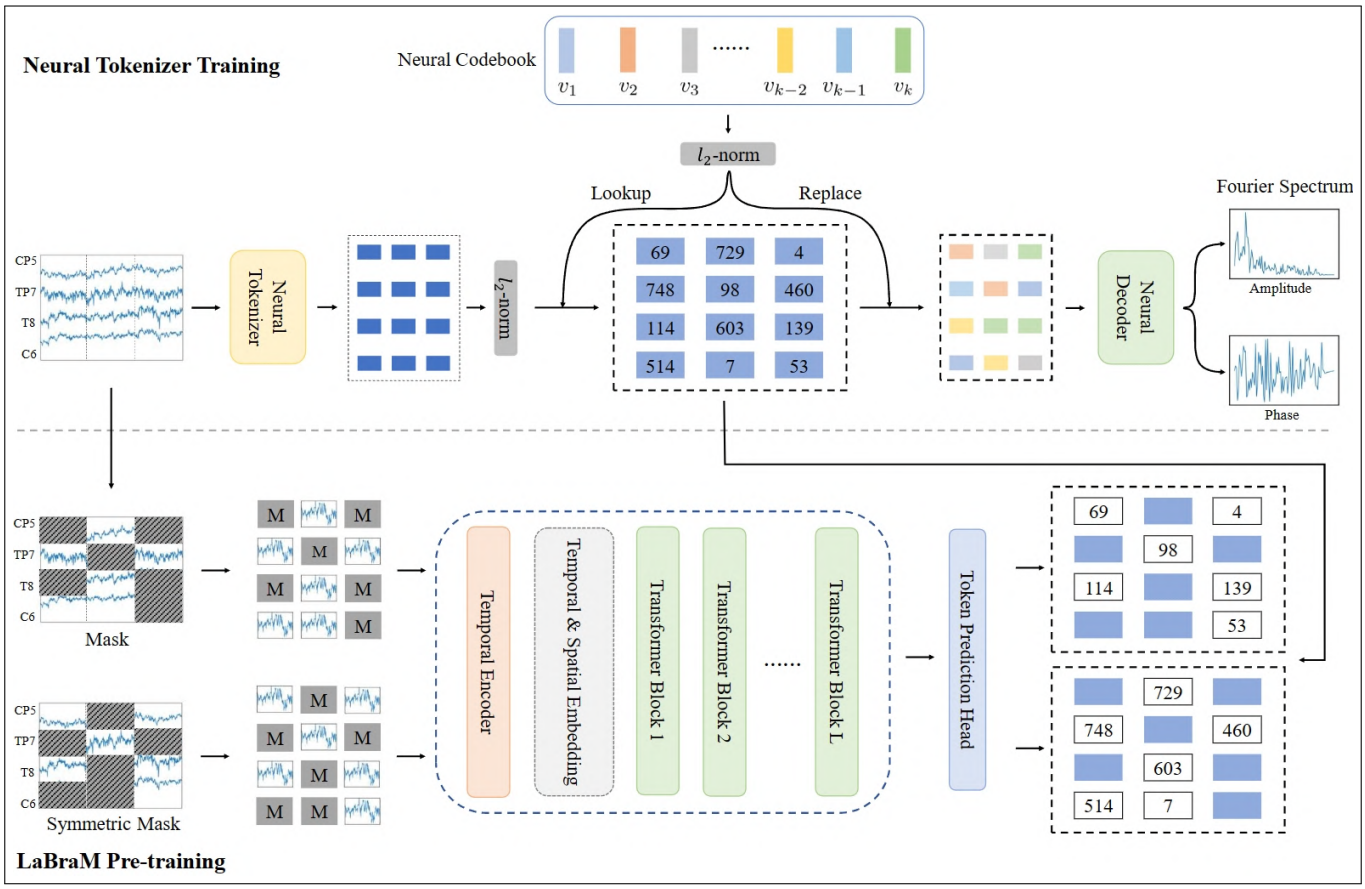


Figure 3.20: Architecture of the LaBraM Framework [14].

3.6.8 Self-Training Strategy with Hybrid Deep Learning

Due to the use of a single dataset and the high variability between subjects, an Intra-Subject Fine-Tuning strategy is applied to improve model adaptability. The model is trained separately on each subject's data in order to learn subject-specific patterns. Then, knowledge is transferred across subjects by reusing the weights learned from selected subjects as an initialization point for further training. This improves weight initialization, accelerates convergence, and allows better adaptation to individual EEG characteristics. In the subsequent stage, a full fine-tuning procedure is performed, where all parameters of the CNN–MHA–BiLSTM hybrid model are updated during training without freezing any layers. For evaluation consistency and to ensure an unbiased assessment of performance, subjects whose learned weights are used for initialization are excluded from the final testing phase and are not reintroduced in subsequent training or evaluation iterations.

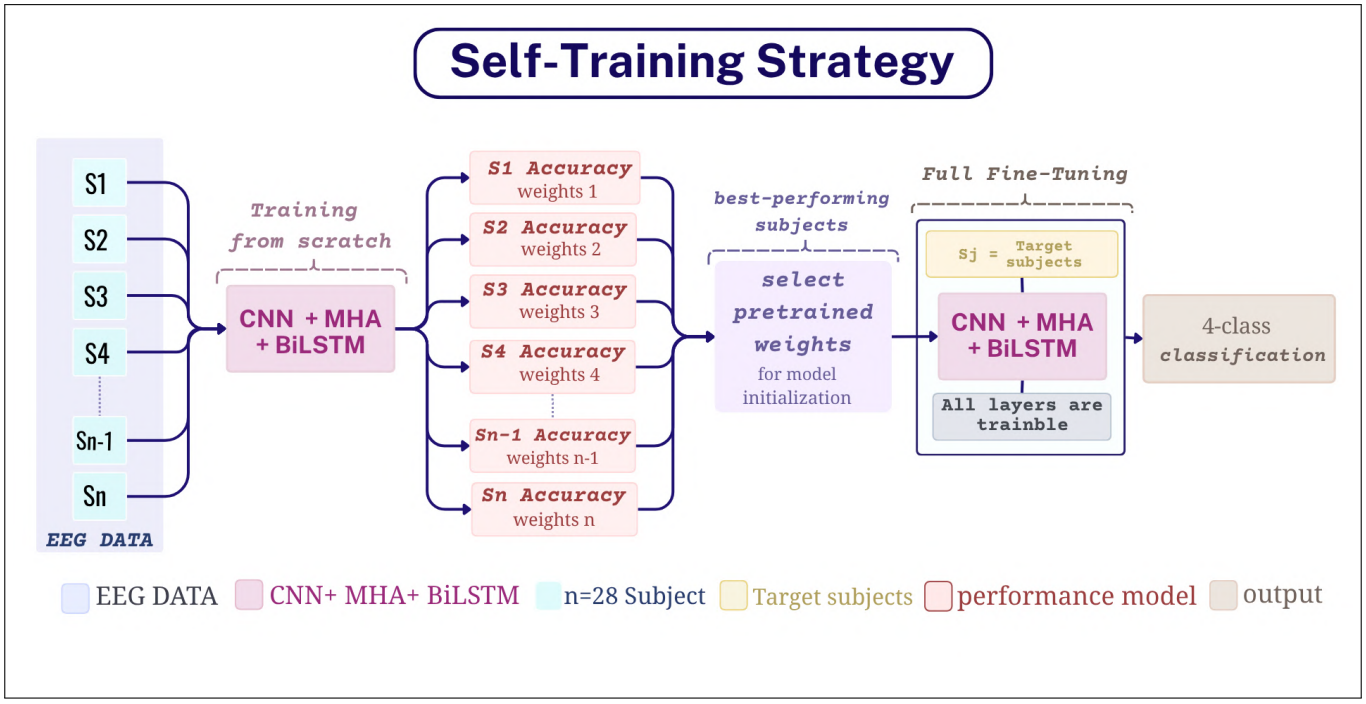


Figure 3.21: Illustrated Architecture of the Proposed Self-Training Strategy.

3.7 Conclusion

This chapter established the scientific and technical foundations of the study by defining the different stages and procedures adopted for building an EEG-based emotion recognition system. The proposed methodology enabled the development of an integrated framework that starts from raw data processing and ends with the design of suitable classification models, while taking into account the complexity and specificity of EEG signals. This framework provides a solid basis for the practical part of the study, which begins with the experimental setup in the next chapter, followed by the presentation and analysis of the obtained results in the subsequent chapters.

Experimental Setup

4.1 Introduction

This chapter presents the experimental framework used to evaluate the proposed EEG-based emotion recognition models. It describes the evaluation strategy adopted in this study, including both Subject-Dependent and Subject-Independent experimental settings, which are designed to assess the models' ability to learn participant-specific patterns and generalize to unseen individuals. The chapter also presents the performance metrics and software environment used throughout the experiments, providing the necessary framework for evaluating and comparing the proposed models.

4.2 Evaluation Strategy and Performance Metrics

The proposed models are evaluated using two main experimental setups: the Subject-Dependent setting and the Subject-Independent setting. The former is adopted to study the models' ability to learn individual EEG signal patterns using data from the same participant, while the latter is used to test the models' ability to generalize to new participants, given the substantial variability in brain signals between individuals.

4.2.1 Subject-Dependent Setting

In this setting, the model is built and evaluated independently for each participant, where data from the same individual are used for both training and testing.

a. 80/20 Split

80% of each participant's data are used for training and 20% for testing. This setup is primarily used with machine learning and ensemble learning models.

b. 10-Fold Cross-Validation

Each participant's data is divided into 10 equal folds. In each iteration, one fold is used for testing and one fold for validation, while the remaining folds are used for training. This process is repeated 10 times to ensure that all folds are evaluated. This approach is used with deep learning and hybrid models to obtain a more stable and reliable assessment.

4.2.2 Subject-Independent Setting

In this setting, the model's ability to generalize to unseen participants whose data are not used during the training phase is evaluated, making the task more challenging due to the significant inter-individual variability in EEG signals.

a. Leave-One-Subject-Out (LOSO) Approach

This approach is adopted for all models, including machine learning and deep learning models. In each iteration, one participant's data are used as the test set, while the data from the remaining

participants are used for training. This process is repeated until each participant has been used once as the test subject.

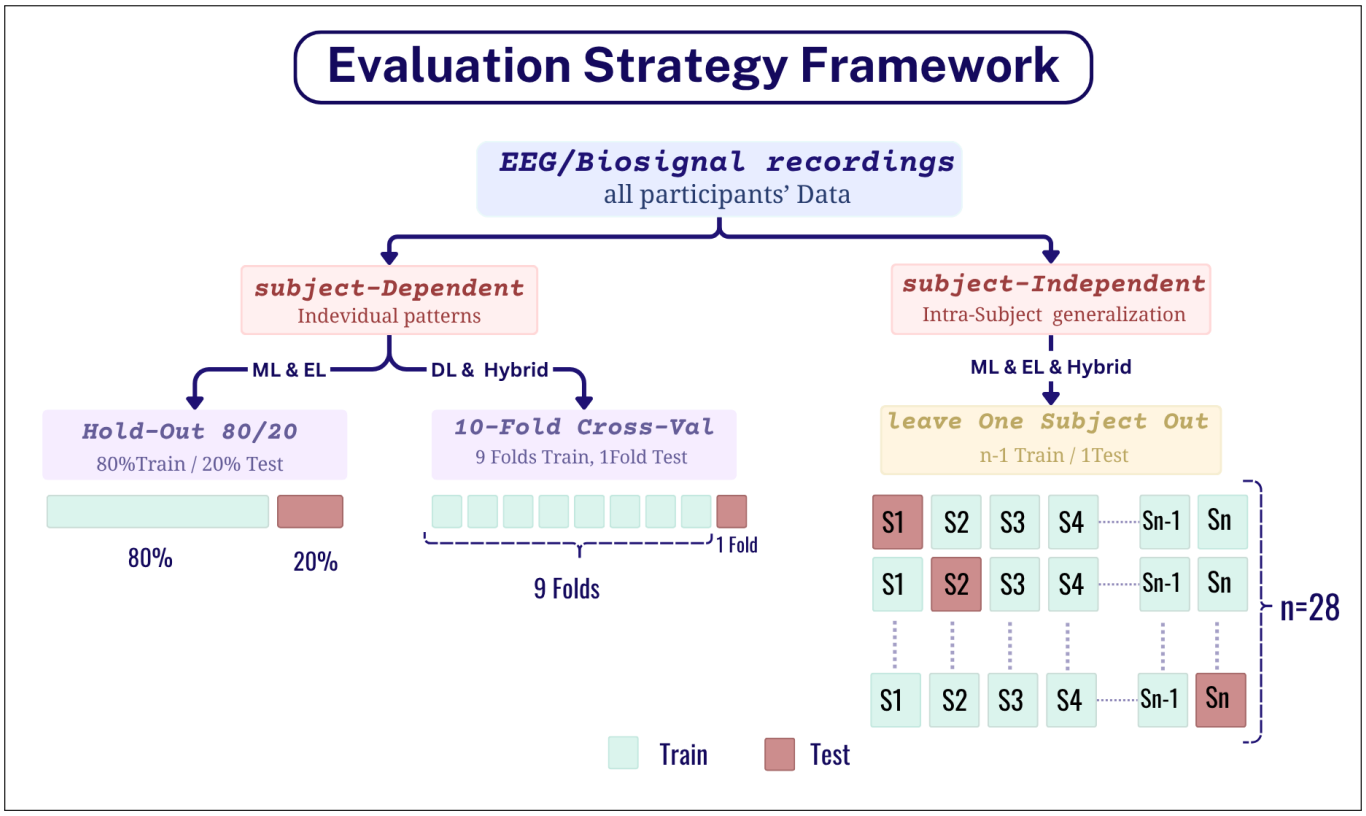


Figure 4.1: Subject-Dependent vs. Subject-Independent Validation Strategies.

4.2.3 Evaluation Metrics

Due to the non-stationary nature of EEG signals, their sensitivity to noise, and the similarity between emotional classes, a single evaluation metric may not sufficiently reflect the model's performance. Therefore, multiple complementary metrics are used to provide a reliable assessment.

In addition, since the task involves multi-class classification, performance is evaluated for each class individually and then summarized using Macro Average and Weighted Average metrics to provide an overall evaluation.

1. **Accuracy** measures the proportion of correctly classified samples relative to the total number of samples.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.1}$$

Where:

- **TP** (True Positive): The number of EEG signals that belong to a specific emotional category and are correctly classified.
- **TN** (True Negative): The number of signals that do not belong to that category and are correctly classified.
- **FP** (False Positive): The number of signals that do not belong to the category but are incorrectly classified as positive.

- **FN** (False Negative): The number of positive signals that are incorrectly classified as negative.

2. **Precision** measures the proportion of correctly predicted positive samples among all predicted positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.2}$$

3. **Recall** measures the model's ability to correctly identify all actual positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.3}$$

4. **F1-score** represents the harmonic mean between Precision and Recall.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.4}$$

5. **Confusion Matrix** illustrates the distribution of correct and incorrect predictions across classes.

$$\text{Confusion Matrix} = \begin{bmatrix} \text{TP} & \text{FN} \\ \text{FP} & \text{TN} \end{bmatrix} \tag{4.5}$$

6. **Macro Average** calculates the model's performance across all classes equally, without considering the number of samples in each class.

$$\text{Macro Average} = \frac{1}{K} \sum_{i=1}^{K} M_i \tag{4.6}$$

7. **Weighted Average** takes into account the number of samples in each class when computing the overall performance.

$$\text{Weighted Average} = \frac{\sum_{i=1}^{K} w_i M_i}{\sum_{i=1}^{K} w_i} \tag{4.7}$$

4.3 Tools and Work Environment

The experimental setup is implemented using both Google Colaboratory and Kaggle environments. The former provides a Tesla T4 GPU to accelerate computations, along with 12.7 GB of RAM and 15 GB of GPU memory, while the Kaggle platform offers a cloud computing environment supported by CPU and GPU resources with approximately 13 GB of RAM. The deep learning models are developed using Python 3, relying on a wide range of libraries to support the workflow as follows:

- **NumPy:** performing numerical operations, matrix manipulation, normalization, and statistical computations.
- **Pandas:** reading, cleaning, and preprocessing structured data (CSV files).
- **Matplotlib and Seaborn:** visualizing and analyzing EEG signals and results.
- **SciPy:** processing signals and extracting features, including filtering and spectral analysis.
- **Scikit-learn:** preprocessing data, splitting datasets, and evaluating models.
- **TensorFlow and Keras:** building and training deep learning models.
- **h5py:** efficiently storing and managing large-scale datasets in HDF5 format.

4.4 Conclusion

In conclusion, this chapter presented the experimental framework for evaluating EEG-based emotion recognition models using a structured protocol and appropriate performance metrics. Building on this foundation, the following chapters will analyze the results and discuss the impact of different experimental settings on model performance.

Results and Discussion for the Subject-Dependent Approach

5.1 Introduction

This chapter presents the experimental results and discussion of the Subject-Dependent approach. It provides a detailed overview of the experimental setup and the hyperparameter configurations used across machine learning, deep learning, and hybrid models. The chapter further evaluates ensemble learning strategies and Intra-Subject Fine-Tuning techniques to assess their effectiveness in EEG signal classification for individual participants.

5.2 Experimental Setup

5.2.1 Loss Function and Output Activation

Softmax was used to convert the network outputs into probability distributions for multi-class classification according to Equation **5.1**:

$$p_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{5.1}$$

Categorical Cross-Entropy was adopted as the loss function to quantify the discrepancy between the true labels and the predicted probabilities, as defined in Equation **5.2**:

$$L = -\sum_i y_i \log(p_i) \tag{5.2}$$

5.2.2 Hyperparameters for ML and EL Models

Table **5.1** summarizes the hyperparameter configurations used for the ML and EL models.

Model / Technique	**Hyperparameters**
KNN	n_neighbors = 5, weights = 'distance', metric = 'minkowski', p = 2
LDA	solver = 'svd', tol = 0.0001 (Default Parameters)
SVM	max_iter = 5000
RF	n_estimators = 100, n_jobs = -1
XGBoost	n_estimators = 150, learning_rate = 0.1, max_depth = 6, subsample = 0.8, colsample_bytree = 0.8, n_jobs = -1
ExtraTrees	n_estimators = 100, bootstrap = True, n_jobs = -1
LightGBM	n_estimators = 100, learning_rate = 0.1, num_leaves = 31, importance_type = 'gain', n_jobs = -1
Local Stacking	Base models: n_estimators = 100 Meta-classifier (LR): max_iter = 1000
Majority Voting	Base models: n_estimators = 100 Majority voting rule (no additional training)
Global Stacking (Fusion)	Base models: n_estimators=100 Fusion classifier (LR): max_iter=2000

Table 5.1: Hyperparameter Settings for Machine Learning and Ensemble Learning Models

A sliding **window** of 128 samples with a 50% overlap ratio was employed, yielding 598 windows per signal and a total of 2392 windows per subject.

5.2.3 Hyperparameters for DL and Hybrid Models

Table **5.2** presents the hyperparameter configurations used for the DL and hybrid models.

Architecture	Ovlp.(%)	L.R.	Optim.	Epochs	Batch	Drop.	Early Stopping
MLPNN	50	adaptive (0.001)	SGD	500	auto ($\leq$ 200)	–	–
MLPNN	95	adaptive (0.001)	SGD	500	auto ($\leq$ 200)	–	–
BiLSTM (Raw)	75	0.01	SGD	100	32	0.25	patience=10
BiLSTM (Preprocessed)	75	0.01	SGD	100	32	0.25	patience=10
Deep BiLSTM (Spectral Entropy)	75	0.001	Adam	100	32	0.15	patience=10
Deep BiLSTM (statistical features)	75	0.001	Adam	250	32	0.15	patience=20
Deep BiLSTM (EMD-VMD + Stat)	75	0.001	Adam	150	32	0.15	patience=10
CNN + LSTM	75	0.001	Adam	250	32	0.3-0.4	patience=20
CNN + BiLSTM + MHA	75	0.0005	Adam	100	32	0.4	patience=10 + ReduceLR
CNN + BiGRU + MHA	75	0.0005	Adam	100	32	0.4	patience=10 + ReduceLR(p=8,f=0.5)
Transformer + DeepBiLSTM	75	0.001	Adam	150	32	0.4–0.3	ES(p=15) + RLR(p=7,f=0.5)
CNN + MHA + BiLSTM	75	0.0005	Adam	100	32	0.4–0.5	ES(p=10) + RLR(p=8,f=0.5)
CNN + MHA + BiLSTM + Ensemble ML	75	0.0005	Adam	100	32	0.4–0.5	DL: ES(p=10), RLR(p=8); ML: SVM(C=10); XGB/LGBM(N=100)

Table 5.2: Hyperparameter Settings for Deep and Hybrid Learning Models.

For all models, stratified 10-fold cross-validation was applied to ensure balanced class representation across all folds.

5.3 Results and Discussion

5.3.1 Results of ML-Based Models

Table **5.3** illustrates the accuracy scores of the machine learning models across different feature sets using StandardScaler normalization.

Classifier	Non-Linear-Mix	Time-Domain-Mix	Frequency-Mix	Wavelet-Energy	Elite-Melange	Total-AIO
KNN	35.30	32.62	35.00	34.21	33.35	34.21
LDA	42.96	40.40	39.21	37.85	41.42	41.04
SVM	81.02	86.31	84.41	85.86	87.97	89.15
RF	84.36	86.53	86.12	87.97	86.71	89.41
XGBoost	**86.48**	**90.13**	88.97	90.90	**89.10**	**92.52**
ExtraTrees	83.52	85.89	84.60	87.39	85.84	88.82
LightGBM	86.35	90.08	**89.16**	**90.61**	88.96	92.37

Table 5.3: Average Accuracy (%) for ML Models in Subject-Dependent Approach Across Different Feature Sets Using StandardScaler Normalization.

Performance Comparison across ML Models

The experimental results indicated clear variations in performance across the evaluated classifiers. XGBoost achieved the highest accuracy (92.52%), followed by LightGBM (92.37%) and Random Forest (89.41%), highlighting the strong effectiveness of ensemble-based methods. SVM also showed competitive performance with an accuracy of 89.15%, while ExtraTrees reached 88.82%. In contrast, LDA and KNN recorded considerably lower accuracies of 35.30% and 42.96%, respectively. Therefore, the aforementioned results confirmed the superiority of ensemble learning techniques over linear models and distance-based methods such as KNN. These findings were further supported by Figure **5.1**, which clearly illustrated the performance disparities between the evaluated classifiers.

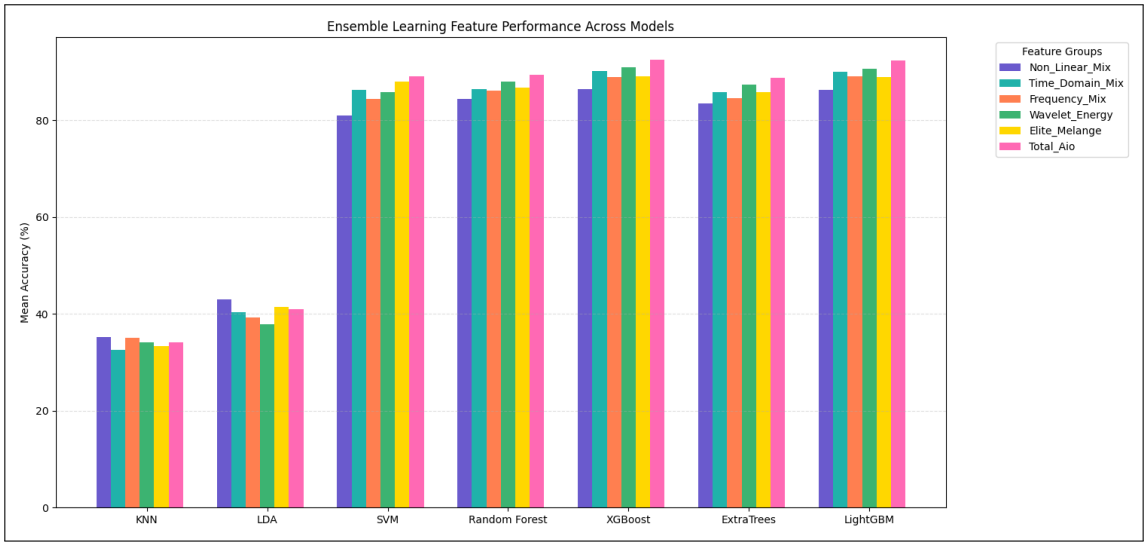


Figure 5.1: Performance Comparison of ML Models Across Distinct Feature Selection Groups (Dependent Approach).

Performance Comparison across Feature Representations

Regarding feature representations, the results indicated that the Total-AIO feature set achieved the best performance (92.52%), reflecting the effectiveness of feature fusion. It was closely followed by Wavelet-Energy (90.61%) and Time-Domain-Mix (90.13%), both of which demonstrated strong discriminative capability. Frequency-Mix and Elite-Melange exhibited similar performance levels, with accuracies of 89.16% and 89.10%, respectively, indicating only marginal differences between them. In contrast, the Non-Linear-Mix feature set obtained the lowest accuracy (86.48%). Therefore, the abovementioned results suggest that combined feature sets generally outperformed individual feature sets, as they captured more comprehensive information from EEG signals. This was further supported by Figure **5.2**, which clearly illustrates the performance variations across different feature configurations.

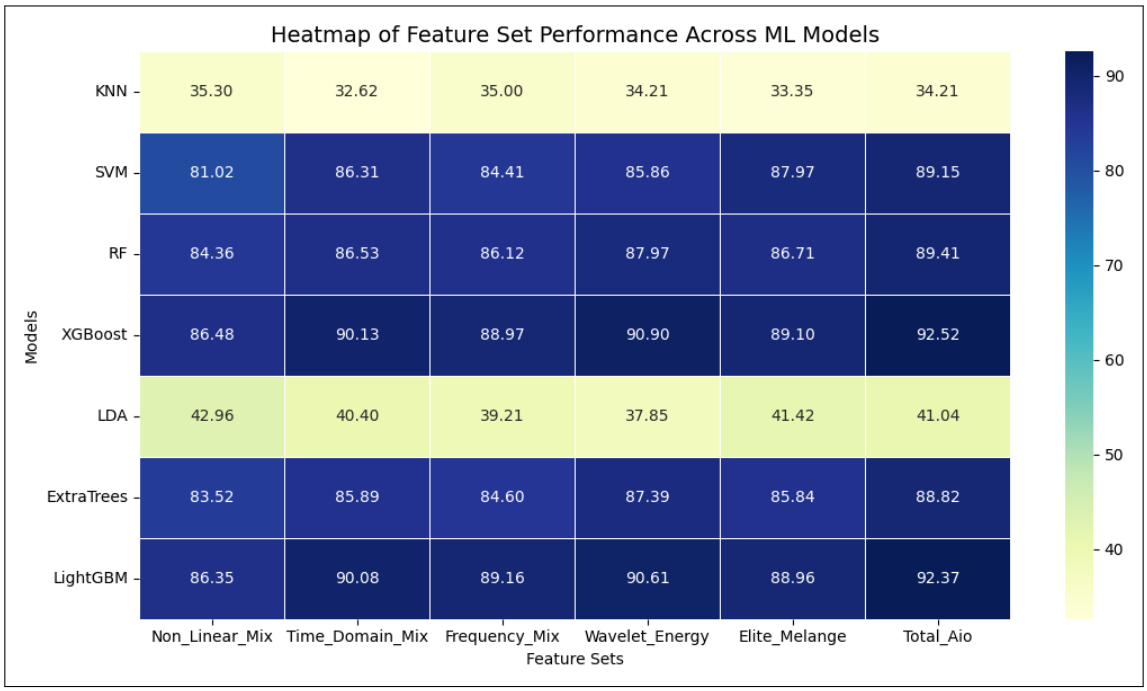


Figure 5.2: Heatmap Analysis of Classification Accuracy Across ML Models and Feature Sets (Dependent Approach).

Inter-Subject Variability

The results demonstrated noticeable variability across participants. XGBoost and LightGBM exhibited more consistent distributions, indicating better stability across subjects. However, some lower outliers were still observed even when using the same classifier and the same feature set. For example, using Total-AIO with XGBoost (see Appendix **A.5**), subject S04 achieved 99.79% accuracy, whereas S14 reached only 72.53%. Similarly, with Wavelet-Energy and SVM (see Appendix **A.3**), S04 obtained 98.74% compared to 70.23% for S28. In contrast, KNN and LDA showed a wider spread of accuracy values, as illustrated in Figure **5.4**. These differences indicated substantial variations in EEG patterns among participants, which directly influenced classification performance.

Intra-Subject Variability

The results also demonstrated clear intra-subject variability, where the classification performance of the same participant varied according to the selected feature set. For instance, under the XGBoost model (see Appendix **A.5**), subject S28 achieved 73.79% using Non-Linear-Mix, which improved to 79.03% using Total-AIO and further increased to 81.55% with Time-Domain-Mix. Similarly, subject S14 improved from 65.61% with Frequency-Mix to 77.35% with Time-Domain-Mix under the same classifier. Comparable variations were also observed for other participants, such as S09, whose accuracy increased from 90.98% with Non-Linear-Mix to 96.64% with Total-AIO. These variations in performance across feature sets are illustrated in Figure **5.3**. These findings suggest that different EEG feature sets possess different discriminative capabilities within the same participant, highlighting the importance of feature selection for classification performance.

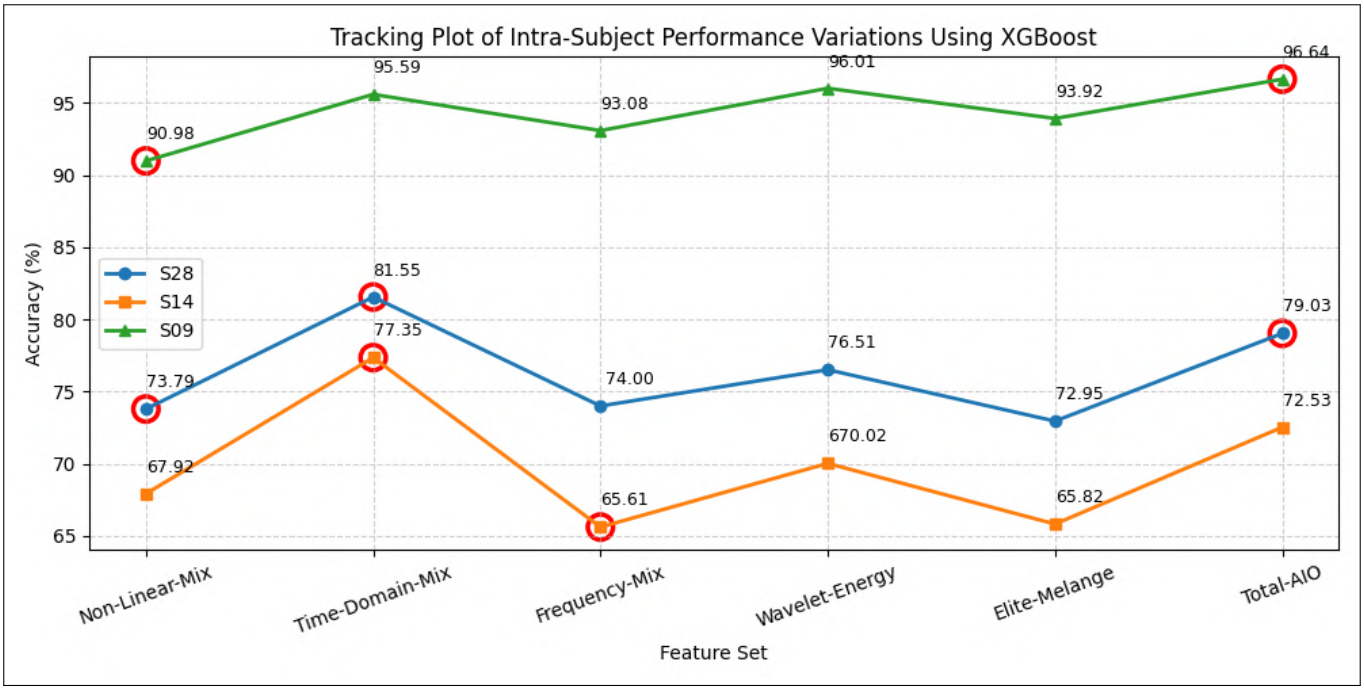


Figure 5.3: Tracking plot illustrating the performance variations of subjects S28, S14, and S09 across different feature sets using the XGBoost classifier.

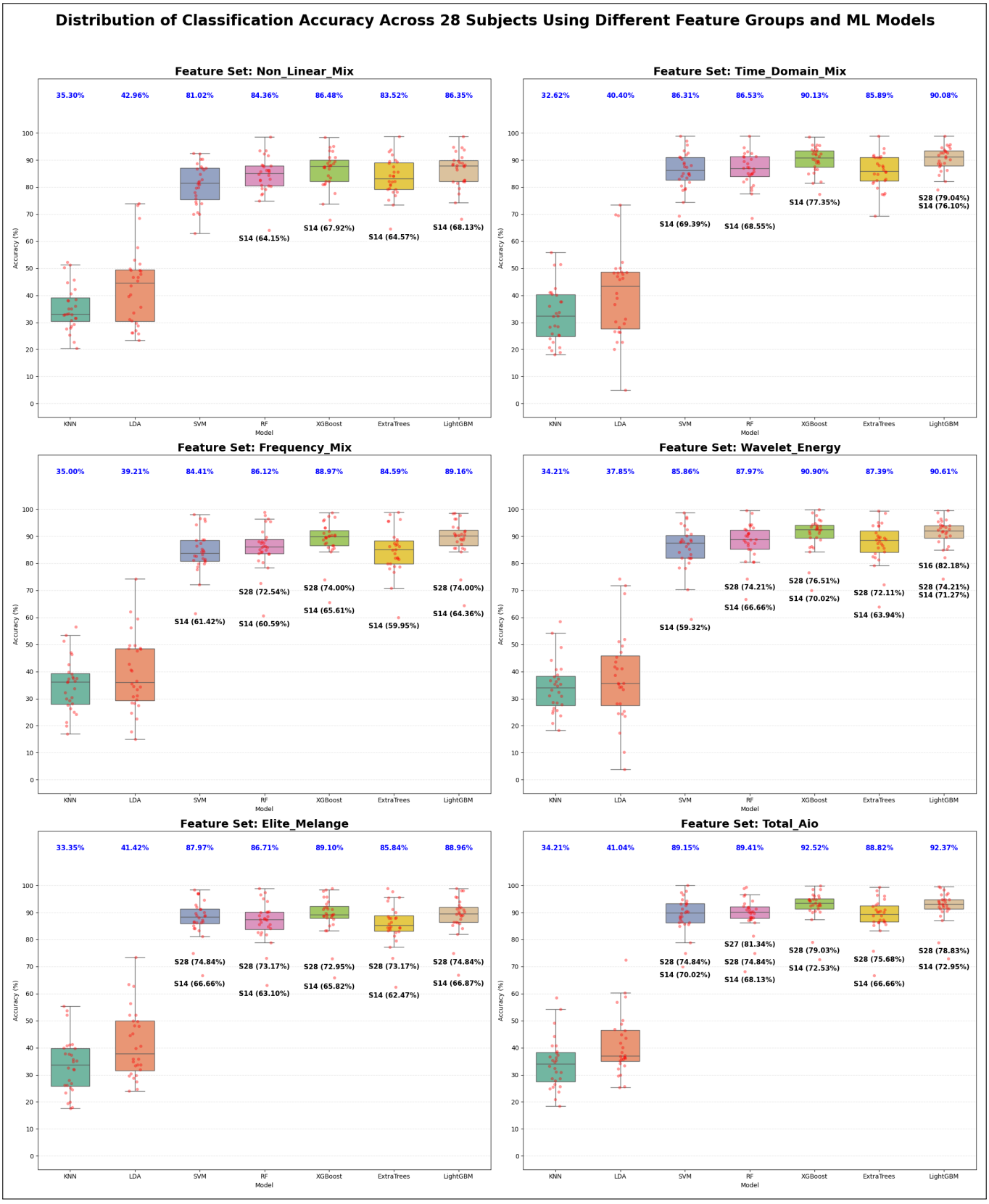


Figure 5.4: Subject-Wise Accuracy Distribution and Outlier Analysis Across ML Classifiers and Feature sets (Dependent Approach).

5.3.2 Results of DL-Based Models

Table **5.4** presents the performance of the deep learning models across different feature extraction methods and overlap percentages using StandardScaler normalization.

Model	Overlap (%)	Feature Extraction	Best Subject (%)	Worst Subject (%)	Overall Mean (%)
MLPNN	50	PSD (Pre-processed)	89.22 (S17)	22.88 (S16)	58.47
MLPNN	95	PSD (Pre-processed)	92.87 (S17)	35.61 (S16)	68.56
BiLSTM	75	Spectral Entropy (Raw)	85.23 (S04)	52.31 (S26)	67.43
BiLSTM	75	Spectral Entropy (Pre-processed)	85.66 (S18)	52.87 (S26)	69.69
Deep BiLSTM 1	75	Spectral Entropy (Pre-processed)	86.50 (S18)	52.75 (S14)	68.96
Deep BiLSTM 2	75	Basic Stats (Max, Min, Mean)	89.64 (S17)	59.60 (S14)	78.10
Deep BiLSTM 3	75	EMD + VMD + Statistical	91.12 (S17)	58.30 (S14)	78.56

Table 5.4: Comparative Average Classification Accuracy (%) of DL Models Across EEG Feature Extraction Methods with StandardScaler Normalization.

Effect of Windowing Overlap Percentage on Performance

Two overlap percentages (50% and 95%) were first evaluated using the MLPNN model to assess the effect of overlap size on classification performance. Although the 95% overlap achieved higher accuracy, it produced a large number of highly repetitive segments, resulting in increased data redundancy. Therefore, a moderate overlap ratio of 75% was selected for subsequent deep learning experiments to balance performance and data diversity.

Preprocessed EEG Signals Vs. Raw EEG Signals

The effect of using preprocessed signals versus raw signals is also assessed in this study. The BiLSTM model was evaluated using both raw and processed EEG signals. In both cases, only bandpass filtering was applied to ensure a fair comparison between the two signal types. The results showed that the processed signals achieved better classification performance than the raw signals. Consequently, the processed EEG data were adopted for subsequent deep learning experiments.

Performance Comparison between DL Models Across Feature Representations

Among the evaluated feature sets, the combined EMD, VMD, and statistical features achieved the best performance, reaching an overall mean accuracy of 78.56% using the Deep BiLSTM model, followed closely by the Basic Statistical features, which also performed strongly with the same model, achieving 78.10%. Spectral Entropy features showed similar performance with both BiLSTM and Deep BiLSTM models, reaching overall mean accuracies of 69.69% and 68.96%, respectively, indicating that increasing the model depth did not lead to a significant improvement. In contrast, PSD-based features achieved the lowest accuracies with the MLPNN model. As illustrated in Figure **5.5**, the results showed that hybrid and statistical feature representations achieved better performance compared to other feature sets in EEG signal classification using deep learning models.

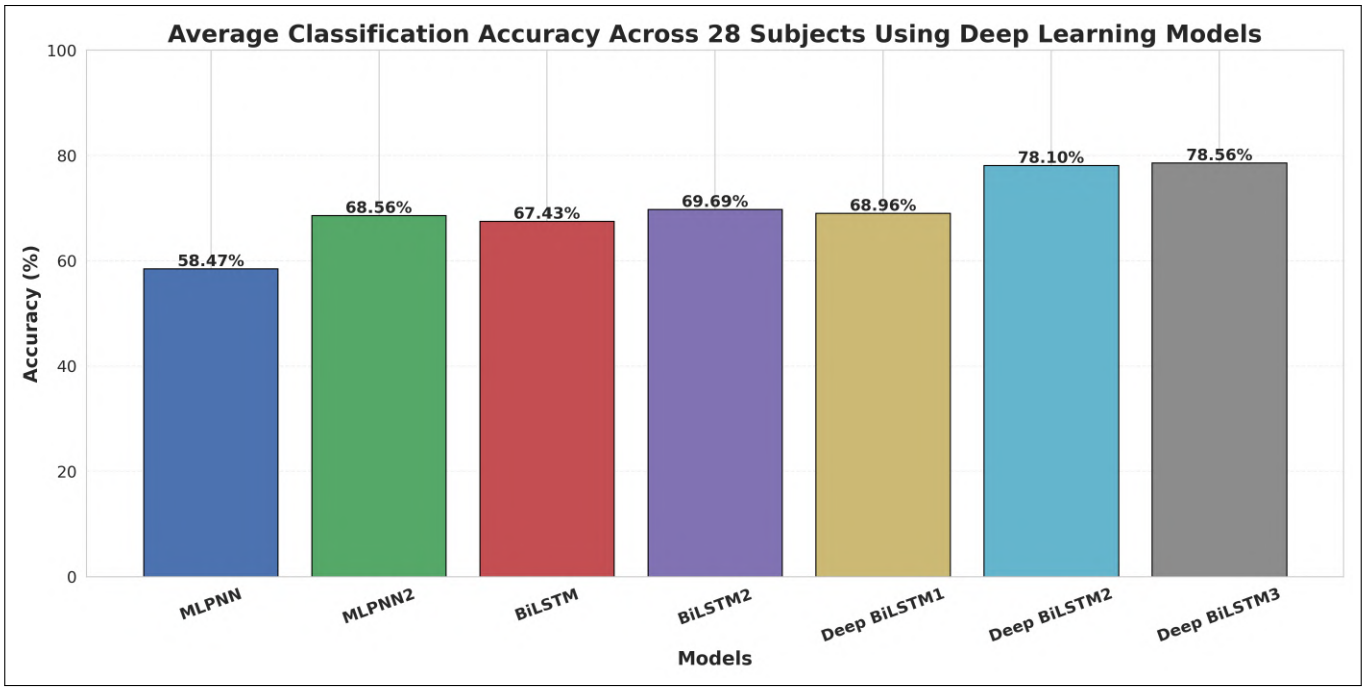


Figure 5.5: Average Classification Accuracy Across 28 Subjects Using DL Models.

Inter-Subject Variability

Figure **5.6** illustrated the distribution of classification accuracy across 28 subjects for seven combinations of deep learning architectures and feature sets, which showed an overall improvement in performance with increasing model complexity and enhanced feature representation, as well as clear differences in stability across the models. The MLPNN model (see Appendix **A.9**) exhibited the highest inter-subject variability, with a mean accuracy of 68.56% and values ranging from 35.61% to 92.87% using PSD features. In contrast, the BiLSTM model (see Appendix **A.11**) demonstrated more consistent performance, achieving a mean accuracy of 69.69% with a range from 52.87% to 85.66%, indicating reduced variability and improved consistency across subjects. The Deep BiLSTM model (see Appendix **A.14**) achieved the best overall performance and stability, with a mean accuracy of 78.56% and a range from 58.30% to 91.12%, with the strongest results obtained using the EMD, VMD, and statistical feature combination. Therefore, variability decreased progressively from MLPNN to BiLSTM and then to Deep BiLSTM, confirming that deeper architectures provided more robust EEG representations and reduced inter-subject variability.

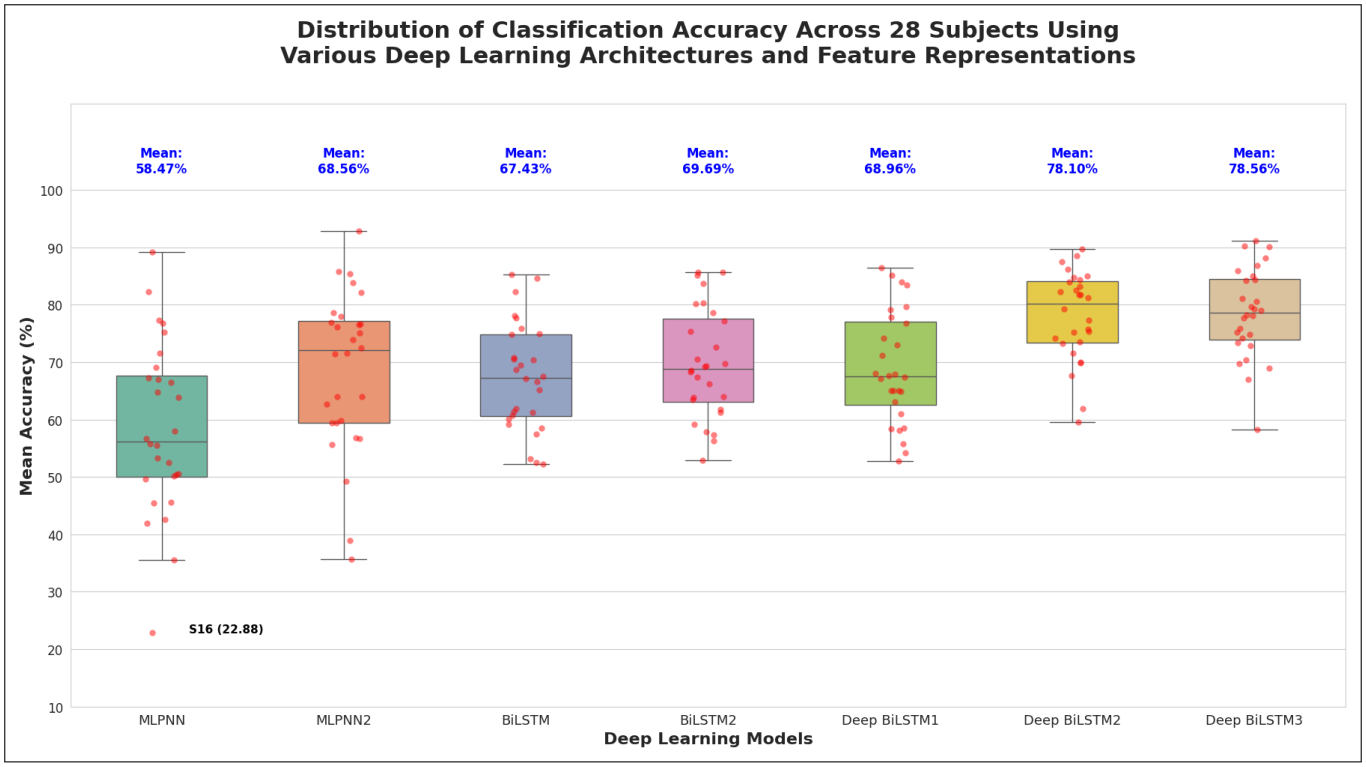


Figure 5.6: Subject-Wise Accuracy Distribution and Outlier Analysis Across DL Models.

Intra-Subject Variability

Intra-subject variability showed a clear sensitivity to the choice of architecture and feature extraction method, as the same participant's performance changed noticeably with different feature representations and model structures. This suggests that stability depends not only on the architecture but also on the quality of the extracted features. This was illustrated by participant S14, whose accuracy reached 35.51% in the MLPNN model using PSD features (see Appendix **A.8**), increased to 56.28% in the BiLSTM model using Spectral Entropy features (see Appendix **A.11**), and further improved from 52.75% in the Deep BiLSTM model using Spectral Entropy features (see Appendix **A.12**) to 58.30% when using EMD + VMD features (see Appendix **A.14**) within the same architecture. These variations in performance across architectures and feature representations are illustrated in Figure **5.7**. These observations highlight that both architectural complexity and feature selection significantly influence intra-subject performance stability.

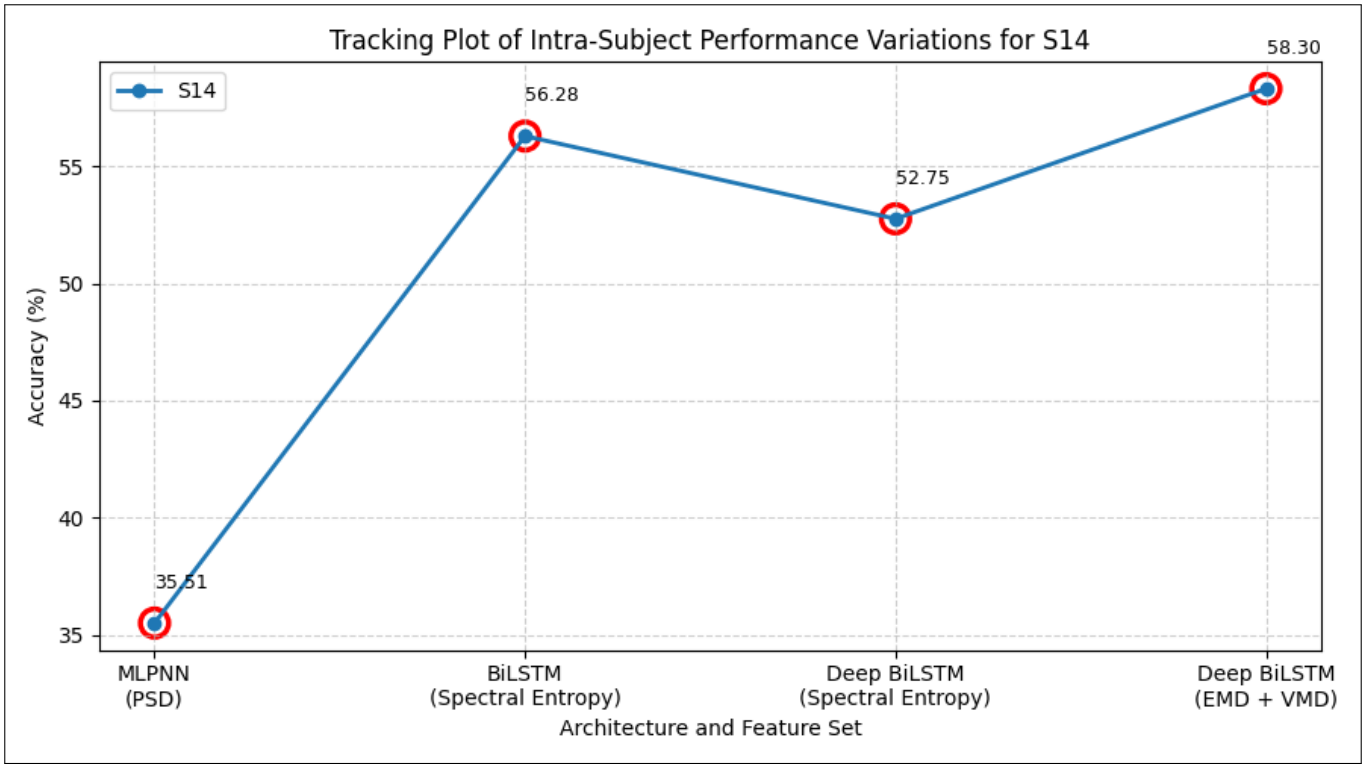


Figure 5.7: Tracking Plot of Intra-Subject Performance Variations for Subject S14 Across Different Architectures and Feature Sets.

5.3.3 Results of EL-Based Models

The ensemble learning models are composed of multiple machine learning classifiers, while Logistic Regression (LR) is employed as a meta-classifier to combine the predictions of the base models and perform the final classification. By leveraging the complementary strengths of multiple classifiers, the ensemble learning models achieved better performance than individual machine learning models, as demonstrated by the results presented in Tables **5.3** and **5.5**.

Approach	Non-Linear-Mix	Time-Domain-Mix	Frequency-Mix	Wavelet-Energy	Elite-Melange	Total-AIO
Local Stacking	87.56	90.81	89.50	91.47	89.80	**93.00**
Majority Voting	**92.45**					
Global Stacking	**95.31**					

Table 5.5: Average Accuracy (%) for Ensemble Learning Models in Subject-Dependent Approach Across Different Feature Sets Using StandardScaler Normalization.

Performance Comparison of Local Stacking Across Feature Sets

The Total-AIO set achieved the highest accuracy of 93.00% in the Local Stacking ensemble, followed by Wavelet-Energy with 91.47%, while Non-Linear-Mix recorded the lowest performance

at 87.56%, as shown in Table **5.5**. This variation showed that the type of feature representation had a clear impact on model performance, highlighting the importance of selecting informative and discriminative features to improve classification results.

Performance Comparison of Ensemble Learning Approaches

Global Stacking demonstrated superior performance compared to both the Local Stacking and Majority Voting ensembles, highlighting the advantage of integrating information from multiple feature sets. It achieved the highest accuracy among the evaluated fusion strategies, reaching 95.31%, whereas Majority Voting achieved 92.45%. This superiority reflected the ability of Global Stacking to make better use of all feature representations in a unified manner. As a result, it captured complementary information across feature spaces more effectively, leading to higher accuracy and greater stability. These observations were further confirmed by Figure **5.8**, which demonstrated that Global Stacking consistently outperformed Majority Voting across nearly all 28 subjects, maintaining superior overall accuracy.

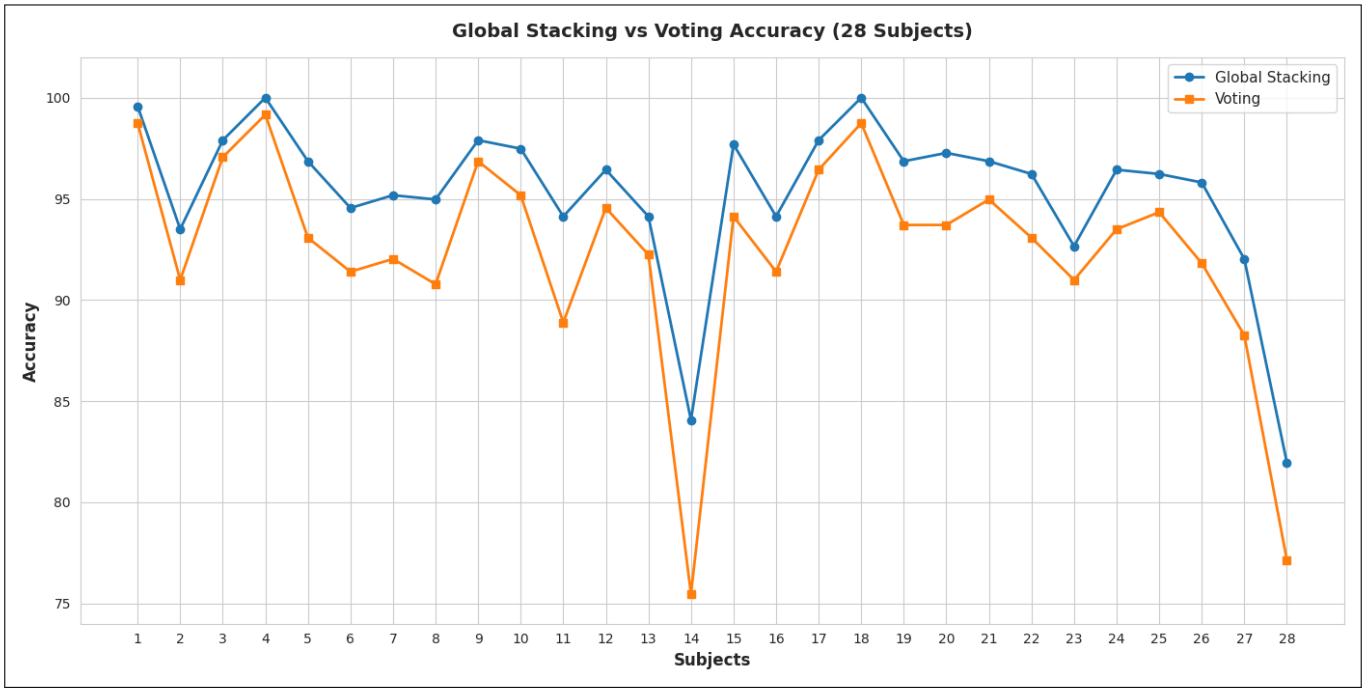


Figure 5.8: Global Stacking vs Majority Voting Accuracy (28 Subjects)(Dependent Approach).

Inter-Subject Variability

Figure **5.9** illustrated the distribution of classification accuracy across participants and demonstrated that ensemble learning models achieved more stable and consistent performance compared to individual models and deep learning approaches, with accuracy values ranging between 87.56% and 95.31%, indicating high performance and relatively low variability across participants.

The Global Stacking model delivered the best overall performance in terms of both accuracy and stability, outperforming all other approaches. It was followed by the Total-AIO model as one of the most balanced Local Stacking strategies, while Majority Voting achieved competitive results with slightly lower stability.

Despite the overall strong performance, variability was still observed across participants, with S14 and S28 identified as outliers across all models, exhibiting lower performance compared to the

remaining participants. This variability was attributed to the individual characteristics of EEG signals and the inherent differences in participants' neural responses. However, the Global Stacking model reduced this variability, maintaining stable performance even in more challenging cases and highlighting its robustness in handling high inter-subject variability.

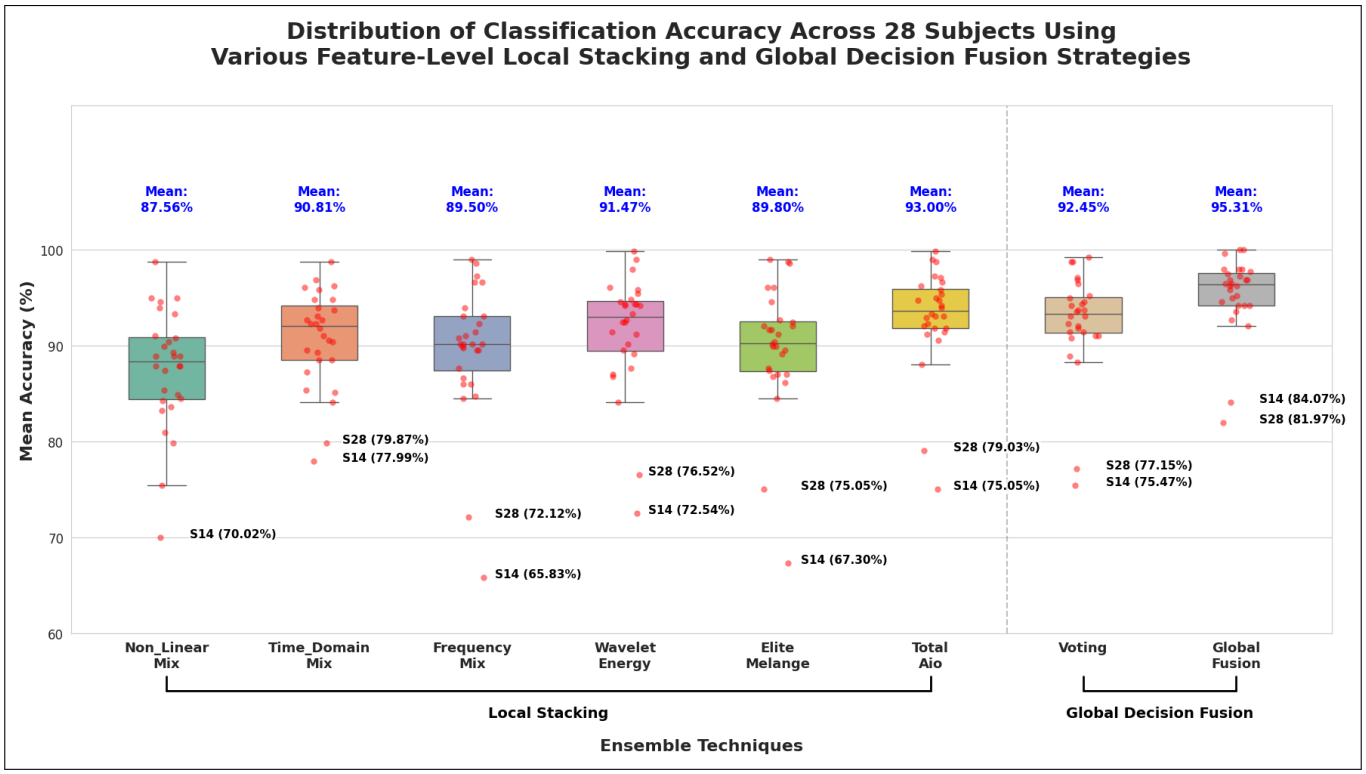


Figure 5.9: Distribution of Classification Accuracy Across 28 Subjects Using Local Stacking & Global Stacking vs Majority Voting (Dependent Approach).

5.3.4 Results of Hybrid Models

Table **5.6** summarizes the performance of different hybrid deep learning architectures, where the models are applied directly to input signal data without any feature extraction stage, using StandardScaler normalization.

Model	Best Subject (%)	Worst Subject (%)	Overall Mean (%)
CNN + LSTM	91.41 (S01/S22)	25.84 (S15)	69.30
CNN + BiLSTM + MHA	96.81 (S25)	47.80 (S14)	84.72
CNN + BiGRU + MHA	97.76 (S17)	57.53 (S14)	88.58
Transformer + DeepBiLSTM	91.46 (S22)	46.98 (S27)	73.13
CNN + MHA + BiLSTM	98.28 (S25)	56.14 (S14)	89.07
CNN + MHA + BiLSTM + EL	97.70 (S25)	59.30 (S14)	88.87

Table 5.6: Average Accuracy (%) for Hybrid Deep Learning Architectures Using StandardScaler Normalization and (75%) Overlapping Ratio.

Performance Comparison Across Hybrid DL and ML Models

The comparison of results among various hybrid deep learning architectures, illustrated in Figure **5.10**, demonstrated a gradual improvement in performance as models progressed from simpler to more complex forms. The CNN + LSTM model achieved an average performance of 69.30%, which reflected the limitations of this architecture in capturing temporal dependencies within the signal, especially when complex and non-linear patterns were present.

Replacing LSTM with BiLSTM and adding the MHA mechanism led to a clear improvement in performance, as observed in the CNN + BiLSTM + MHA model, which achieved 84.72%. This improvement can be attributed to BiLSTM's ability to process signal sequences in both directions, providing a better understanding of temporal dependencies, while the attention mechanism enhanced feature representation by focusing on the most informative parts of the signal and reducing the impact of noise.

A further improvement was observed when BiGRU was used instead of BiLSTM in the CNN + BiGRU + MHA model, increasing the performance to 88.58%. This improvement was explained by the lower computational complexity of GRU compared to LSTM, which reduced the number of parameters and training cost, leading to faster convergence and improved stability while preserving the ability to capture essential temporal dependencies in the signal.

The Transformer + DeepBiLSTM model achieved a relatively lower performance (73.13%) compared to the other models. This behavior was attributed to the limited suitability of the architecture for EEG signals, as well as the relatively small dataset size, which may have reduced its ability to learn robust feature representations.

The CNN + MHA + BiLSTM model achieved the best performance, with a score of 89.07%, confirming that the MHA mechanism enhanced feature representation by focusing on the most informative parts of the signal while providing an effective balance between spatial and temporal feature extraction, with performance comparable to the CNN + BiGRU + MHA model.

Regarding the combination of deep learning with traditional machine learning methods, the CNN + MHA + BiLSTM + Ensemble model achieved a performance of 88.87%, which was very close to the performance obtained without the ensemble stage (89.07%). This finding indicated that the ensemble component did not provide any meaningful improvement, as the deep learning backbone was already sufficient to effectively extract the most relevant features from the data.

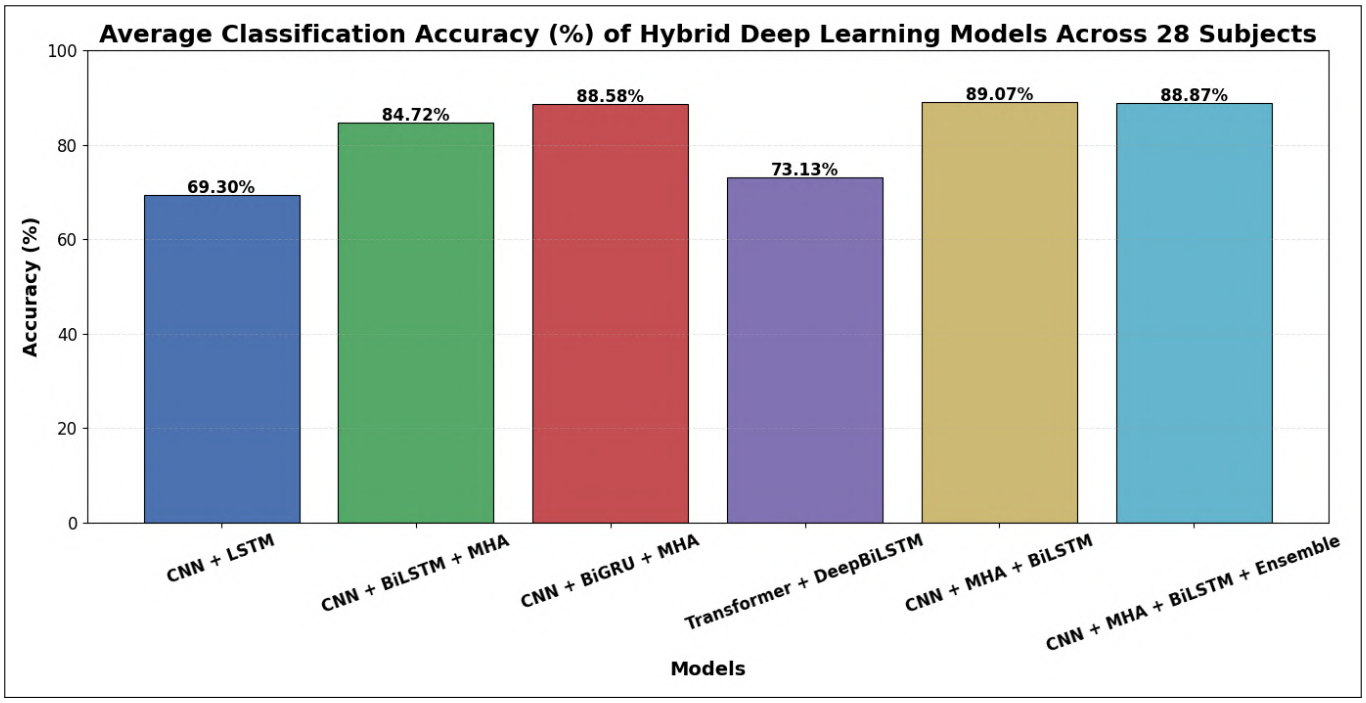


Figure 5.10: Average Classification Accuracy (%) of Hybrid Deep Learning Models Across 28 Subjects (Dependent Approach).

Inter-Subject Variability

The comparison of the proposed hybrid models based on deep learning and attention mechanisms with the baseline deep learning models highlights their outperformance, particularly for outlier cases, as illustrated in Figure **5.11**. In baseline deep learning models, a notable decrease in performance was recorded for some participants, where participant S16 reached the lowest value of 22.88% in the MLPNN model (see Appendix **A.8**), while the performance of participant S26 decreased to 52.87% in the BiLSTM model (see Appendix **A.11**), indicating that these models were more sensitive to individual signal variance and noise levels.

A relative improvement in the minimum performance level was observed in the hybrid models, particularly for low-performing cases such as S14 and S28. Whereby, the lowest accuracy score recorded for S14 gradually increased from 47.80% in the CNN + BiLSTM + MHA model (see Appendix **A.18**) to 57.53% in the CNN + BiGRU + MHA model (see Appendix **A.19**), reaching 59.30% in the CNN + MHA + BiLSTM + EL model (see Appendix **A.22**), while the performance of S28 reached 73.29% under the same model configuration. This improvement in the minimum performance level indicates that hybrid architectures contributed to reducing the severity of variability in outlier cases compared to baseline deep learning models. Such behavior can be explained by the synergy between spatial and temporal feature extraction, together with the ability of the attention mechanism to filter the most relevant features, resulting in more robust and stable signal representations.

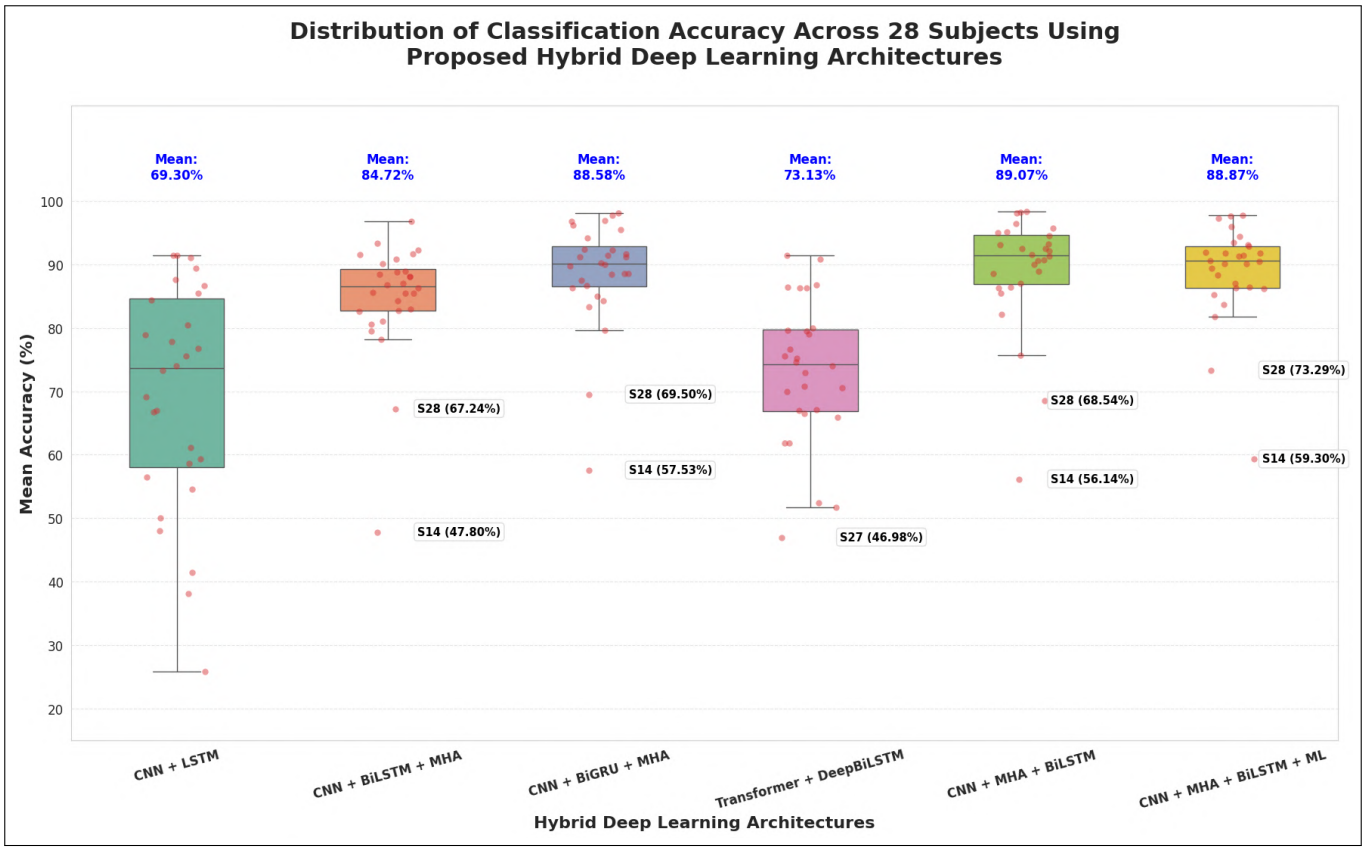


Figure 5.11: Distribution of Classification Accuracy Across 28 Subjects Using Hybrid Deep Learning Architectures

Intra-Subject Variability

To gain deeper insight into intra-subject variability, the evolution of classification accuracy across the evaluated hybrid architectures was analyzed for selected representative participants. Figure **5.12** illustrates the corresponding accuracy trends, highlighting distinct performance profiles, including highly variable, low-performing, and relatively stable participants.

For participant S15, a substantial variation in performance was observed across the evaluated hybrid models. The accuracy increased considerably from 25.84% in the CNN + LSTM (see Appendix **A.17**) model to 91.62% in the CNN + BiGRU + MHA model (see Appendix **A.19**), indicating that the transition from a unidirectional recurrent model to a bidirectional architecture enhanced by attention mechanisms led to a substantial improvement in classification performance.

For participant S14, a gradual improvement in performance was observed across the hybrid models. The accuracy increased from 47.80% in the CNN + BiLSTM + MHA (see Appendix **A.18**) model to 59.30% in the CNN + MHA + BiLSTM + EL model (see Appendix **A.22**). Despite these improvements, the relatively limited performance suggests that the underlying EEG patterns were less distinctive, making it more difficult for the evaluated architectures to extract discriminative features and achieve higher classification accuracy.

Participant S03 exhibited consistently high performance across all models, with only minor variations in accuracy. The highest accuracy of 98.06% was achieved by the CNN + MHA + BiLSTM model (see Appendix **A.21**), whereas the lowest value of 86.23% was recorded by the Transformer + DeepBiLSTM model (see Appendix **A.20**). This behavior indicates that the corresponding EEG patterns were sufficiently discriminative and could be effectively learned by different architectures.

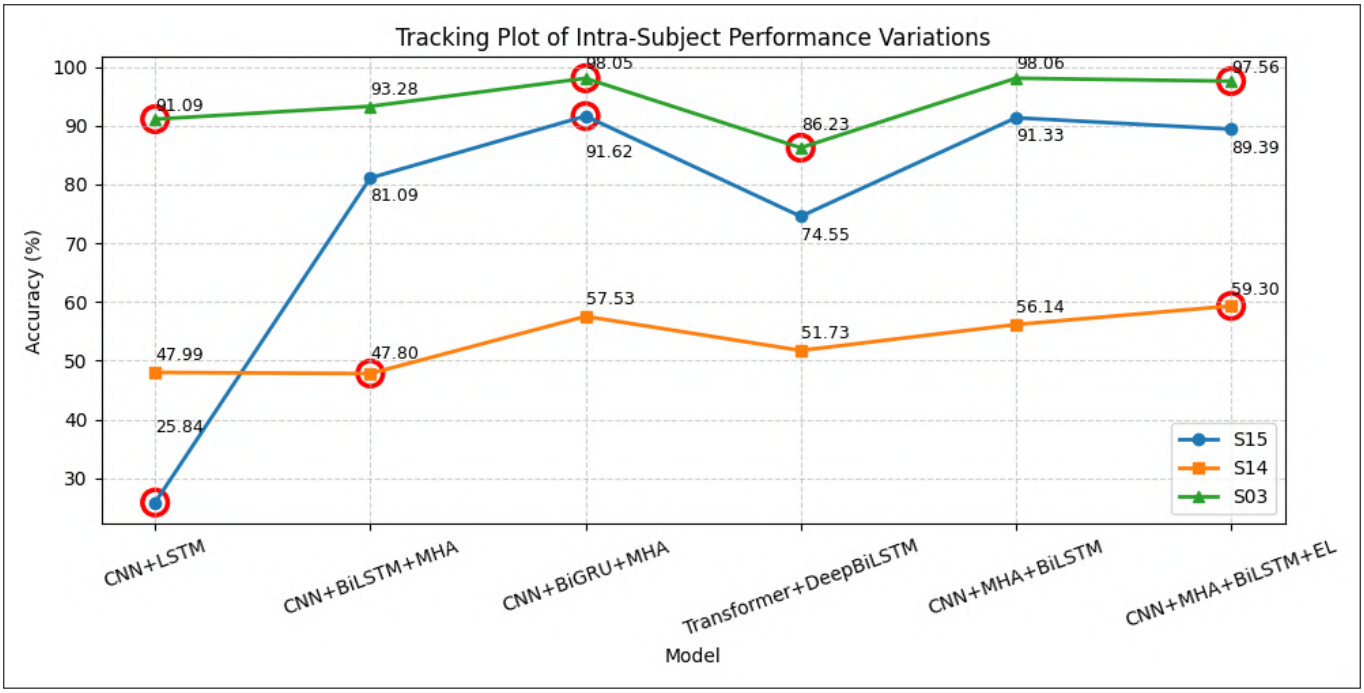


Figure 5.12: Tracking Plot of Intra-Subject Performance Variations Across Hybrid Models.

5.3.5 Results of Self-Training Strategy with Hybrid Deep Learning

In order to further improve the performance of the proposed CNN + MHA + BiLSTM hybrid architecture, an Intra-Subject Fine-Tuning strategy was adopted. The best participant, S03, who achieved the highest accuracy of 98.03%, was selected, and its pre-trained weights were used as a starting point for training the remaining subjects.

Standard Scaler Vs. Robust Scalar

Self-training strategy was evaluated under two experimental settings to study the effect of normalization methods and noise augmentation on model performance. In the first setting, StandardScaler was used without any additional modifications to the data, whereas in the second setting, RobustScaler was used with the addition of a small amount of Gaussian noise applied to EEG signal windows during training in order to improve the robustness of the learned representations against signal variability and outliers.

The results obtained under these two settings are presented in Table **5.7**.

Model	Normalization Technique	Best Subject (%)	Worst Subject (%)	Overall Mean (%)
CNN + MHA + BiLSTM	StandardScaler	99.79 (S02)	81.51 (S08)	94.44
	RobustScaler	99.75 (S17)	91.79 (S14)	97.78

Table 5.7: Impact of Normalization Techniques on the Average Classification Accuracy of Self-Training Strategy.

The results presented in Tables (see Appendix **A.23** and Appendix **A.24**), along with Figure **5.13**, highlight clear differences between the StandardScaler and RobustScaler methods in terms of dispersion, stability, and sensitivity to outliers.

After applying the Intra-Subject Fine-Tuning strategy with the StandardScaler, a clear improvement was observed compared to the baseline model (see Appendix **A.21**), with the overall average accuracy increasing from 89.07% to 94.44%. This improvement was particularly notable in low-performing cases, as S14's accuracy increased from 56.14% to 82.81% and S28's from 68.54% to 88.37% (see Appendix **A.21** and Appendix **A.23**), with a marked reduction in sharp decreases and an increase in minimum performance to 81.51% (S08) (see Appendix **A.23**). This strategy contributed to boosting the general accuracy, improving performance in weaker cases, and reducing gaps between subjects, leading to a more stable and homogeneous distribution of results compared to the baseline model.

Based on the results recorded using the RobustScaler, performance improved relative to the StandardScaler, as the overall average increased from 94.44% to 97.78% (see Appendix **A.23** and Appendix **A.24**). Moreover, the results for most participants became more stable and concentrated within a high range between 96% and 99%, instead of the dispersion observed in the initial results, with a clear improvement even in weak cases where S14's accuracy increased to 91.79%.

This improvement is due to the nature of the RobustScaler, which relies on the median and interquartile range instead of the mean and standard deviation, making it less sensitive to outliers and more capable of adjusting the data distribution and reducing the impact of sharp differences between subjects.

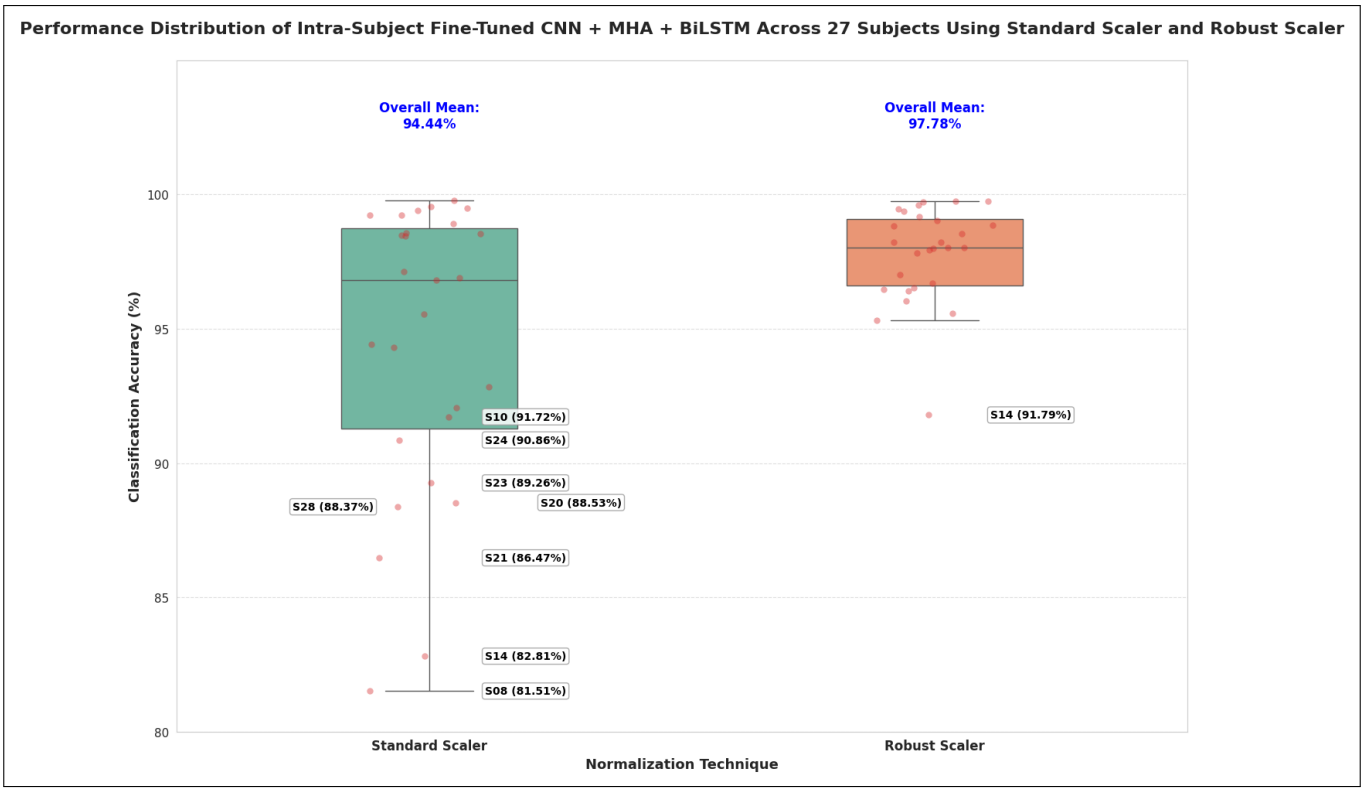


Figure 5.13: Performance Distribution of Intra-Subject Fine-Tuning Using StandardScaler and RobustScaler.

Subject-Wise Emotion Recognition Performance

To provide a more detailed evaluation, Table **5.8** presents the subject-wise performance of the best model configuration using Intra-Subject Fine-Tuning with RobustScaler and Gaussian noise. It reports Accuracy along with Macro and Weighted Recall and F1-scores across all 27 subjects, providing a detailed view of performance at the individual level.

Subject	Macro Recall (%)	Macro F1-score (%)	Weighted Recall (%)	Weighted F1-score (%)	Accuracy (%)
S01	99	99	99	99	99.37
S02	100	100	100	100	99.62
S04	100	100	100	100	99.75
S05	97	97	97	97	97.01
S06	98	98	98	98	98.02
S07	97	96	96	96	96.04
S08	97	97	97	97	96.51
S09	99	99	99	99	99.03
S10	99	99	99	99	98.86
S11	98	98	98	98	97.98
S12	98	97	98	98	98.02
S13	99	99	99	99	98.53
S14	91	91	92	92	91.79
S15	98	98	98	98	97.93
S16	99	99	99	99	99.16
S17	100	100	100	100	99.75
S18	100	100	100	100	99.83
S19	98	98	98	98	98.23
S20	96	95	95	95	95.32
S21	98	97	96	96	96.46
S22	98	98	98	98	98.23
S23	97	97	98	98	97.81
S24	99	99	99	99	98.82
S25	99	99	99	99	99.45
S26	95	95	96	96	95.58
S27	97	96	96	96	96.42
S28	97	97	97	97	96.71
MEAN	**97.8**	**97.6**	**97.8**	**97.8**	**97.78**

Table 5.8: Subject-Wise Detailed Classification Performance (Accuracy, Macro and Weighted Metrics) for RobustScaler Across 27 Subjects.

The subject-wise results in Table **5.8** demonstrated consistently high performance across all participants. Most accuracy values exceeded 95%, with the overall mean reaching 97.78%, reflecting

a generally stable behavior of the proposed hybrid model.

At the individual level, several participants achieved near-optimal performance, particularly S02, S04, S17, and S18, with accuracy values ranging from 99.62% to 99.83%. In contrast, a few subjects, such as S14 (91.79%) and S20 (95.32%), showed relatively lower performance; however, these cases remained limited and did not compromise the overall robustness of the model.

Complementing these accuracy results, the Macro and Weighted metrics remained closely aligned, with averages ranging between 97.6% and 97.8%. This indicates that the model maintained a balanced classification performance across all classes without bias toward specific categories.

Classification Error Analysis of Best-, Average-, Worst- Performing Subjects

To further assess the performance of the best performing model, a detailed error analysis is presented by showcasing the confusion matrices of the CNN-MHA-BiLSTM with self-training for the best-performing subject (S18), the average-performing subject (S20), and the lowest-performing subject (S14).

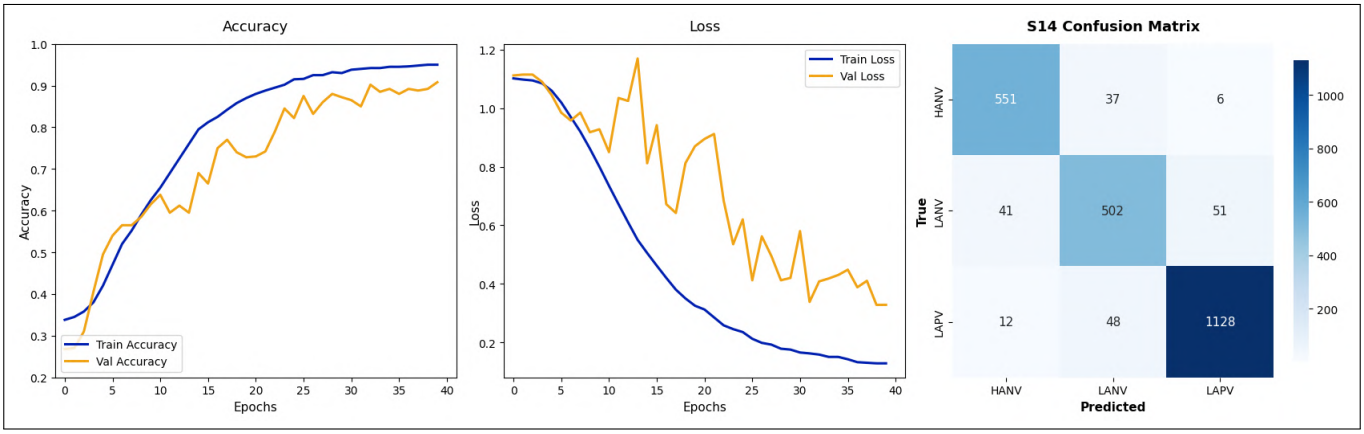


Figure 5.14: Performance Evaluation and Confusion Matrix for Subject 14 (S14).

Subject S14, corresponding to the lowest-performing case, includes three emotional classes: HANV, LANV, and LAPV. As illustrated by the confusion matrix (see Figure **5.14**), most misclassifications occur between LANV and LAPV, suggesting a partial overlap in their feature representations. Among the three classes, LANV appears to be the most challenging to classify correctly, whereas LAPV achieves the highest classification accuracy.

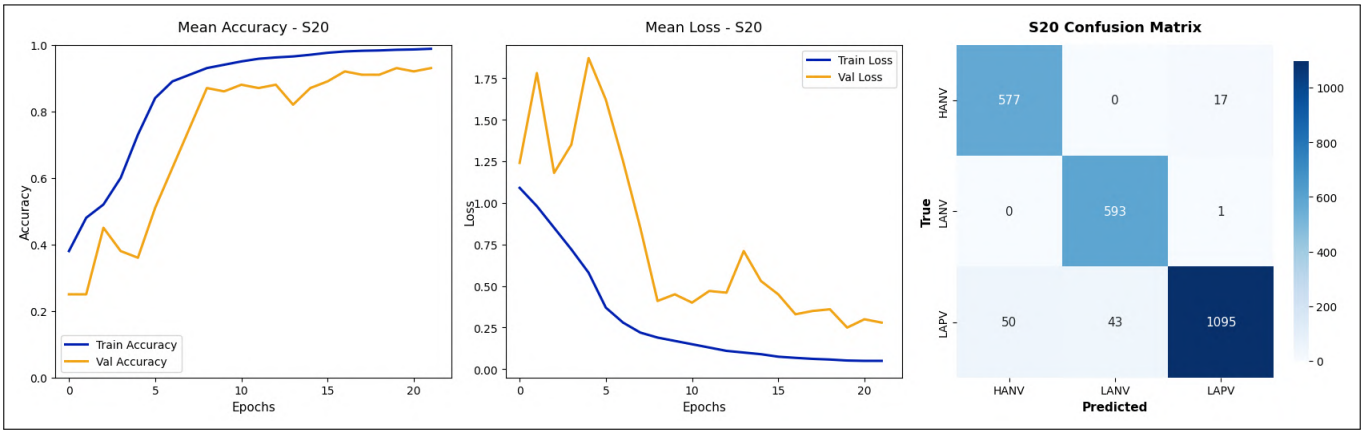


Figure 5.15: Performance Evaluation and Confusion Matrix for Subject 20 (S20).

Subject S20, representing the average-performing case, includes three emotional classes: LANV, LAPV, and HANV. As shown in the confusion matrix (see Figure **5.15**), most misclassifications involve the LAPV class, which is frequently confused with both HANV and LANV, indicating less distinct class boundaries. Despite this, LANV is classified with very high accuracy, while HANV exhibits only limited confusion with LAPV, reflecting relatively good separability.

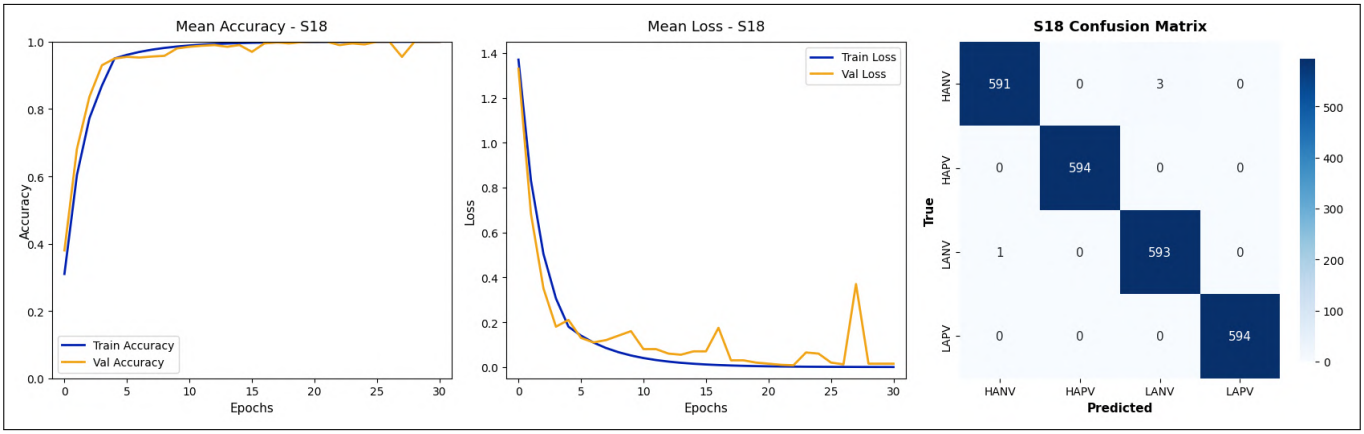


Figure 5.16: Performance Evaluation and Confusion Matrix for Subject 18 (S18).

Subject S18, corresponding to the best-performing case, includes four emotional classes: HANV, HAPV, LANV, and LAPV. As shown in the confusion matrix (see Figure **5.16**), almost all samples are correctly classified across all classes, resulting in a nearly perfectly diagonal pattern. Misclassifications are negligible, confirming excellent class separability and demonstrating highly robust model performance.

5.4 Comparison with Literature

Table **5.9** compares the proposed approach with previous studies in EEG-based emotion recognition, using datasets such as DEAP, SEED, and GAMEEMO, under Random Split and Subject-Dependent Split settings, while highlighting performance differences across methods.

Ref	Model	Dataset	Classification	Split	Results
[66]	BiLSTM	DEAP	Multi-class (3 classes)	Random Split	Valence: 50.9%; Arousal: 74.0%
[66]	BiLSTM + Attention	DEAP	Multi-class (3 classes)	Random Split	Valence: 83.5%; Arousal: 82.6%
[48]	Stacked BiLSTM	DEAP	Binary (2 classes)	Random Split	Valence: 99.45%; Arousal: 96.88%; Liking: 99.69%
[75]	DE + CNN + BiLSTM	SEED	Multi-class (3 classes)	Dependent Split	SEED: 94.82%
[76]	CIT-EmotionNet (CNN + Transformer)	SEED, SEED-IV	Multi-class (3 / 4 classes)	Dependent Split	SEED: 98.57%; SEED-IV: 92.09%
[73]	DAMGCNN	SEED, SEED-IV	Multi-class (3 / 4 classes)	Dependent Split	SEED: 99.42%; SEED-IV: 96.86%
[47]	Deep BiLSTM	GAMEEMO	Multi-class (4 classes)	Random Split	Accuracy: 90.33%
[35]	BiLSTM	GAMEEMO	Binary (2 classes)	Random Split	Accuracy: 76.91%
[28]	RCAR; RIPPER; SVM	GAMEEMO	Multi-class (4 classes)	Dependent Split	RCAR: 87.89%; RIPPER: 79.51%; SVM: 87.29%
Proposed Model	**CNN + MHA + BiLSTM**	**GAMEEMO**	**Multi-class (4 classes)**	**Dependent Split**	**Accuracy: 97.78%**
Proposed Ensemble	**Global Stacking**	**GAMEEMO**	**Multi-class (4 classes)**	**Dependent Split**	**Accuracy: 95.31%**

Table 5.9: Comparison with Related Work for EEG-Based Emotion Recognition under Random and Subject-Dependent Split Settings.

Reviewing literature contributing to EEG-based emotion recognition revealed the use of two evaluation approaches, namely, Random Split and Subject-Dependent. In the first technique, EEG data of all participants are randomly shuffled and split into training and test sets after preprocessing and windowing. Whereas, the second evaluation technique considers each participant as a full dataset and uses cross-validation for training and testing. Based on the results indicated in Table **5.9** above, we notice that frameworks evaluated using Random Split technique showed the highest performance compared to those evaluated using Subject-Dependent technique. Whereby, models such as Stacked BiLSTM [**48**] achieved accuracies of 99.45% and 96.88%, while BiLSTM with Attention [**66**] obtained approximately 83.5% and 82.6%, and Deep BiLSTM [**47**] reached about 90.33%. This was mainly attributed to the overlap of data from the same subjects between the training and testing sets, which introduced a positive bias and led to an unrealistic increase in accuracy.

The Subject-Dependent evaluation approach used by many studies in the literature resulted in relatively limited performance while providing a more accurate and realistic performance evaluation.

Nevertheless, DL-based models, such as DAMGCNN [**73**] and CIT-EmotionNet [**76**], still achieved strong results ranging from approximately 95% to 99%, whereas traditional methods such as SVM, RIPPER, and RCAR [**28**] remained at lower levels between 79% and 88% due to their limited ability to represent the complexity of EEG signals.

A comparison across datasets showed that the best performance was generally achieved on SEED and SEED-IV, with high accuracies reaching 96%–99%, whereas the results on GAMEEMO were relatively lower, ranging from 76% to 90%, due to the increased classification complexity and the larger number of emotional states.

Compared to previous studies on emotion recognition using EEG signals, both proposed frameworks: CNN + MHA + BiLSTM with self-training and Global Stacking Ensemble Learning demonstrated higher classification performance when trained and tested on the GAMEEMO dataset using 10-fold cross-validation subject-dependent evaluation. Whereby, the former achieved an average accuracy of 97.78% and the latter obtained 95.31% by maintaining strong and stable performance across all participants. These results indicate the effectiveness of the proposed methodology in achieving high performance for EEG-based emotion classification in a high-activity context involving playing games.

5.5 Conclusion

This chapter presented the outcomes obtained from applying the proposed architectures within the Subject-Dependent framework. A comprehensive statistical analysis was conducted to examine performance variability and dispersion across participants. Furthermore, a comparative evaluation between machine learning, deep learning, ensemble learning, and hybrid models was performed to highlight their respective strengths and limitations. Among the evaluated approaches, the proposed CNN + MHA + BiLSTM model with Intra-Subject Fine-Tuning achieved the best overall performance, reaching an average accuracy of 97.78%, while the proposed Global Stacking Ensemble obtained a competitive accuracy of 95.31%, demonstrating strong stability across participants. The findings were also discussed in relation to existing literature, providing a basis for assessing general trends in model performance.

Results and Discussion for the Subject-Independent Approach

6.1 Introduction

This chapter presents and analyzes the classification results obtained under the Subject-Independent approach. It describes the experimental setup and hyperparameter settings adopted to ensure consistency and enable fair comparisons across different experiments. The chapter further evaluates the performance of deep learning models, hybrid architectures, and ensemble methods, including EEGNet, DAN, and LaBraM, and examines the effectiveness of self-training strategies in a cross-subject learning scenario.

6.2 Experimental Setup

6.2.1 Loss Function and Output Activation

The same loss function and output activation settings described in Chapter 5 were maintained throughout all experiments to ensure consistency across the evaluated models.

6.2.2 Hyperparameters for ML and EL Models

For consistency and fair comparison, the independent models were trained using the same hyperparameter configurations, training strategies, and overlap ratio adopted for the dependent models.

6.2.3 Hyperparameters for Hybrid Models

Table **6.1** summarizes the hyperparameter settings adopted for the hybrid models, with the same classification configuration maintained throughout the experiments.

Architecture	**Ovlp.(%)**	**L.R.**	**Optim.**	**Epochs**	**Batch**	**Drop.**	**Early Stopping**
EEGNet + Attention	75	0.001	Adam	100	32	0.5	patience=10
EEGNet + Fine-Tuning	75	0.001 / 0.0001 (FN)	Adam	10 + 5 FT	32	0.3	–
EEGNet + Attention + Fine-Tuning	75	0.001 / 0.0001	Adam	10 + 5 FT	32	0.3	–
DAN	0	0.001	Adam	10	64	0.3	–
DAN + Augmentation (Gaussian Noise)	0	0.001	Adam	50	64	0.4	ReduceLR (patience=5, factor=0.5)
DAMGCN + Attention	75	0.001	Adam	50	32	0.5	patience=8
LaBraM (Pretrained Transfer Learning)	50	1e-4 / 5e-6	Adam	100 + 50 FT	32	–	–
CNN + BiLSTM + MHA	75	0.0005	Adam	100	32	0.4	patience=10 + ReduceLR (p = 5)
CNN + BiGRU + MHA	75	0.0005	Adam	100	32	0.4	patience=10 + RLR(p=8,f=0.5)
CNN + MHA + BiLSTM	75	0.0005	Adam	50	32	0.4–0.5	ES(p=10) + RLR(p=8,f=0.5)

Table 6.1: Hyperparameter Settings for Deep and Hybrid Learning Models.

6.3 Results and Discussion

6.3.1 Results of ML-Based Models

The findings presented in Table **6.2** show the classification performance of the machine learning models obtained from the various feature sets after applying StandardScaler normalization.

Classifier	Non-Linear-Mix	Time-Domain-Mix	Frequency-Mix	Wavelet-Energy	Elite-Melange	Total-AIO
KNN	36.86	33.56	36.49	37.76	37.59	38.19
LDA	**42.96**	40.40	39.21	37.85	41.42	41.04
SVM	42.90	**40.92**	39.73	39.23	41.73	40.04
RF	39.21	38.99	44.19	44.29	43.71	45.19
XGBoost	39.53	39.28	**45.80**	43.44	46.00	**46.30**
ExtraTrees	40.32	39.24	45.10	**45.40**	44.38	44.24
LightGBM	39.39	38.67	45.21	42.76	**46.17**	45.66

Table 6.2: Average Accuracy (%) for Machine Learning Models in Subject-Independent Approach Across Different Feature Sets Using StandardScaler Normalization.

Performance Comparison across ML Models

The obtained results revealed relatively similar performance among the evaluated classifiers, with slight variations in accuracy. XGBoost achieved the highest accuracy (46.30%), followed closely by LightGBM (46.17%), ExtraTrees (45.40%), and Random Forest (45.19%). LDA and SVM yielded comparable results, with accuracies of 42.96% and 42.90%, respectively, while KNN recorded the lowest performance at 38.19%. Therefore, ensemble-based models demonstrated superior classification capability compared to linear and distance-based methods, as illustrated in Figure **6.1**.



Figure 6.1: Performance Comparison of Machine Learning Models Across Distinct Feature Selection Groups (Independent Approach).

Performance Comparison across Feature Representations

Among the investigated feature sets, Total-AIO achieved the highest discriminative performance with a peak accuracy of 46.30%, closely followed by Elite-Melange (46.17%). Frequency-Mix and Wavelet-Energy also delivered strong results, achieving accuracies of 45.80% and 45.40%, respectively. Meanwhile, Non-Linear-Mix showed moderate performance (42.96%), while Time-Domain-Mix recorded the lowest accuracy (40.92%). Therefore, these findings indicated that hybrid and information-rich feature representations were more effective in capturing discriminative patterns relevant to the classification task, as illustrated in Figure **6.2**.

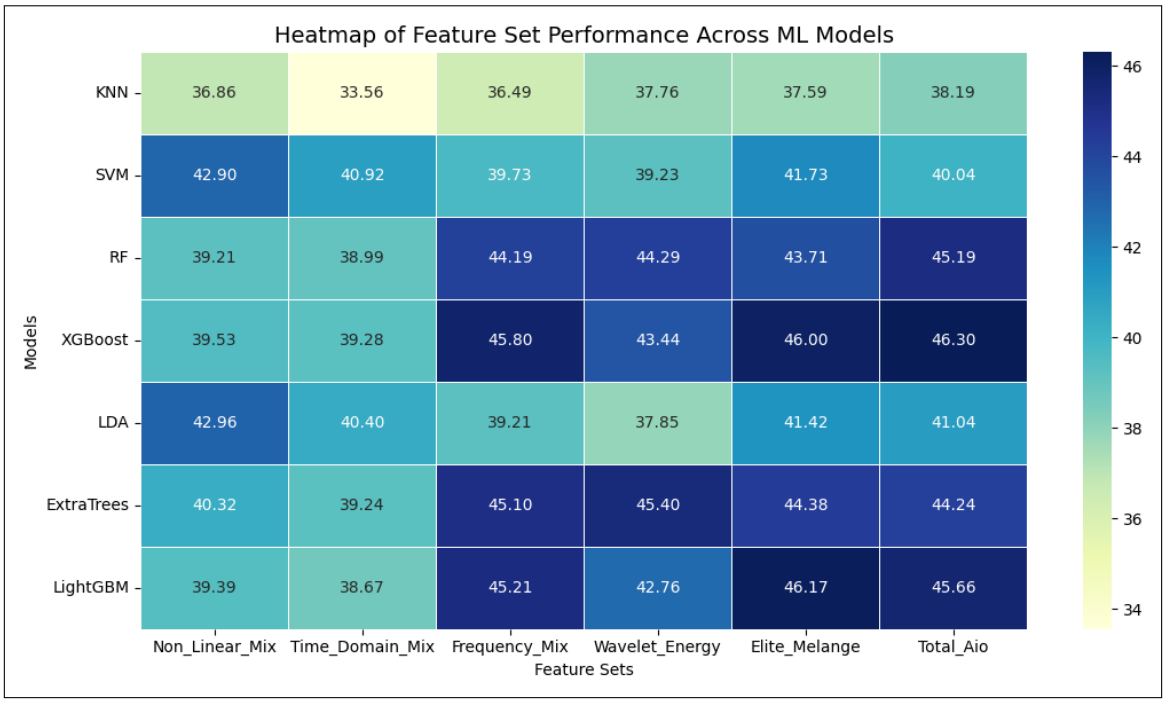


Figure 6.2: Heatmap Analysis of Classification Accuracy Across Machine Learning Models and Feature Sets (Independent Approach).

Inter-Subject Variability

The results of EEG signal classification revealed notable variations among participants, particularly in the performance of the XGBoost and LightGBM models, which achieved the highest average accuracies while exhibiting higher sensitivity to inter-subject differences. For example, the LightGBM model in the Elite-Melange set achieved 75.00% for participant S23, whereas it dropped to 13.17% for S05 (see Appendix **B.7**). Similarly, the XGBoost model in the Total-AIO set reached 75.00% for S23 compared to only 19.37% for S05 (see Appendix **B.5**), and in the Frequency-Mix set, it recorded 74.96% for S09 versus 16.11% for S20, thereby reflecting the sensitivity of boosting algorithms to variability in EEG signals across subjects.

The KNN, LDA, and SVM models exhibited lower variability compared to XGBoost and LightGBM, although their average accuracies were relatively lower. These findings indicated more consistent behavior across participants, with performance values clustering around the overall group means. For instance, the KNN model in the Non-Linear-Mix set achieved 37.42% for S10 and 34.90% for S11 (see Appendix **B.1**). Similarly, the SVM model in the Time-Domain-Mix set produced close results for S1 and S2, with 47.27% and 48.02%, respectively (see Appendix **B.3**), which indicated limited dispersion.

In contrast, the Random Forest and Extra Trees models demonstrated a better balance between accuracy and stability, exhibiting more consistent performance across participants and reduced dispersion, thereby indicating convergence around the overall mean performance. For instance, the Random Forest model in the Frequency-Mix set achieved 50.13% for S24 and 50.96% for S28 (see Appendix **B.4**). Likewise, the Extra Trees model in the Total-AIO set yielded 48.99% and 49.20%, respectively (see Appendix **B.6**), reflecting stable performance across the sample.

Further analysis of Figure **6.3** indicated that model stability was influenced by the feature set. KNN appeared more stable in the Non-Linear-Mix set, while LDA and SVM showed improved consistency within the Elite-Melange set. Random Forest demonstrated the highest stability in the Frequency-Mix set, whereas Extra Trees exhibited balanced behavior across the Total-AIO and Frequency-Mix sets. In contrast, XGBoost and LightGBM achieved relatively improved stability within the Elite-Melange set, although notable inter-subject variability persisted.

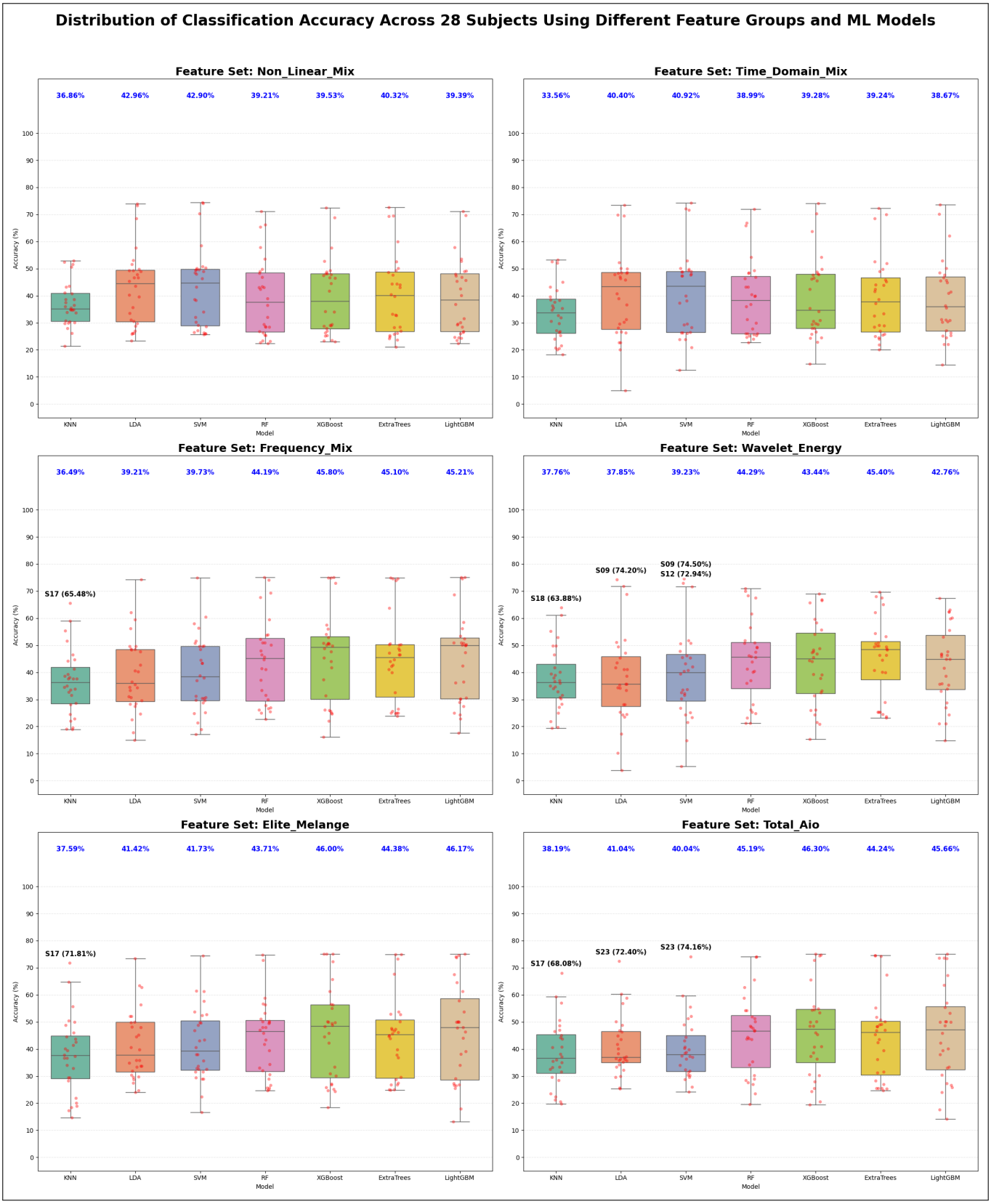


Figure 6.3: Subject-Wise Accuracy Distribution and Outlier Analysis Across Machine Learning Classifiers and Feature Subsets (Independent Approach).

Intra-Subject Variability

The results showed significant intra-subject variability, as model performance for the same participant differed depending on the type of feature set used (see Figure **6.4(a)**). This reflected the sensitivity of classification performance to the method of EEG signal representation. For instance, in the KNN model for participant S17, accuracy reached 71.81% using Elite-Melange, while it decreased to 20.13% using Time-Domain-Mix (see Appendix **B.1**). Similarly, the Random Forest model for participant S01 achieved an accuracy of 59.43% using Frequency-Mix, compared to 22.73% using Non-Linear-Mix (see Appendix **B.4**).

In contrast, some participants exhibited more stable performance across different feature sets, with limited variability in classification accuracy within specific models (see Figure **6.4(b)**). For instance, participant S10 maintained consistent performance with the LDA model, with accuracy ranging approximately between 45% and 48% (see Appendix **B.2**), while participant S11 displayed stable behavior with the KNN model within a range of 35% to 39% (see Appendix **B.1**).

Likewise, participant S13 showed relative stability in the Extra Trees model, with accuracy values remaining within approximately 46% to 50% (see Appendix **B.6**). The aforementioned results indicated that some participants were less sensitive to variations in feature representation, suggesting the presence of more robust and stable EEG patterns across different feature extraction methods.

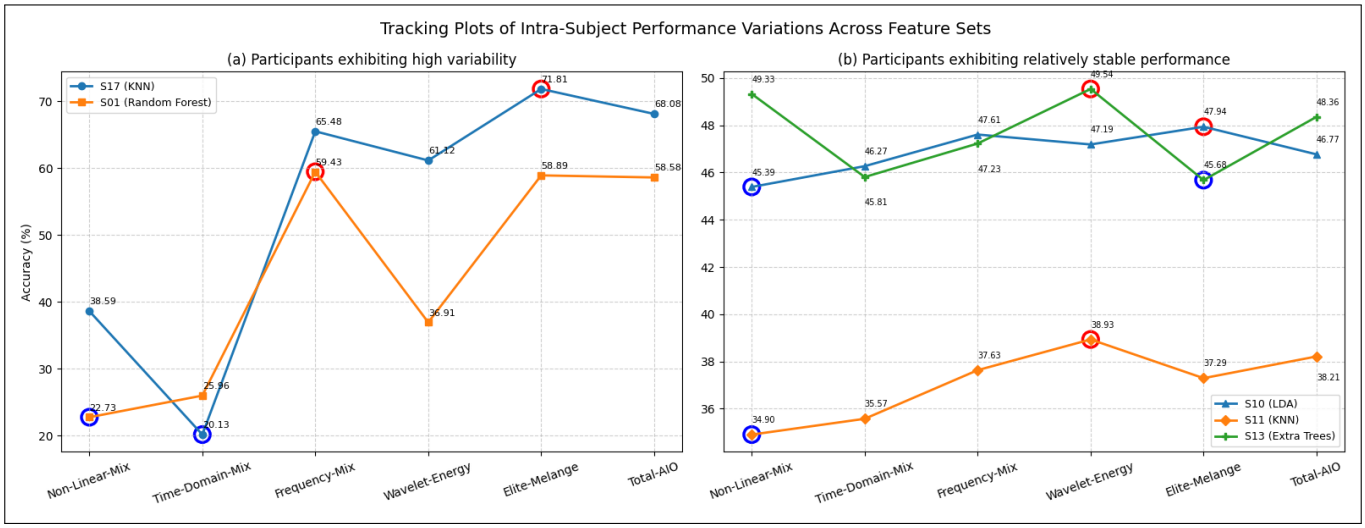


Figure 6.4: Tracking plots of intra-subject performance variations across feature sets for participants exhibiting high variability and relatively stable performance.

6.3.2 Results of EL-Based Models

Ensemble Learning-Based Models demonstrated clear superiority over individual models in terms of accuracy, as shown in Tables **6.2** and **6.3**. Individual models did not exceed 46%, whereas ensemble methods achieved substantially higher performance, confirming the effectiveness of model aggregation in improving both accuracy and generalizability.

Approach	Non-Linear-Mix	Time-Domain-Mix	Frequency-Mix	Wavelet-Energy	Elite-Melange	Total-AIO
Local Stacking	62.47	62.65	69.25	68.53	67.90	**69.25**
Majority Voting	**70.79**					
Global Stacking	**81.70**					

Table 6.3: Average Accuracy (%) for Ensemble Learning Models in Subject-Independent Approach Across Different Feature Sets Using StandardScaler Normalization

Performance Comparison of Local Stacking Across Feature Sets

The Local Stacking models showed differences in performance across feature sets. Frequency-Mix and Total-AIO achieved the highest accuracy (69.25%), followed by Wavelet-Energy and Elite-Melange. In contrast, Non-Linear-Mix and Time-Domain-Mix achieved lower but similar accuracies of 62.47% and 62.65%, respectively, highlighting the impact of feature extraction on classification quality.

Performance Comparison of Ensemble Learning Approaches

The results showed that the Majority Voting model outperformed the Local Stacking models in terms of accuracy, while Global Stacking delivered the best performance among all ensemble learning methods, outperforming both Local Stacking and Majority Voting and confirming its effectiveness in achieving the highest levels of accuracy and stability. This was further validated by Figure **6.5**, which clearly demonstrated the consistent superiority of Global Stacking over Majority Voting across all 28 subjects.

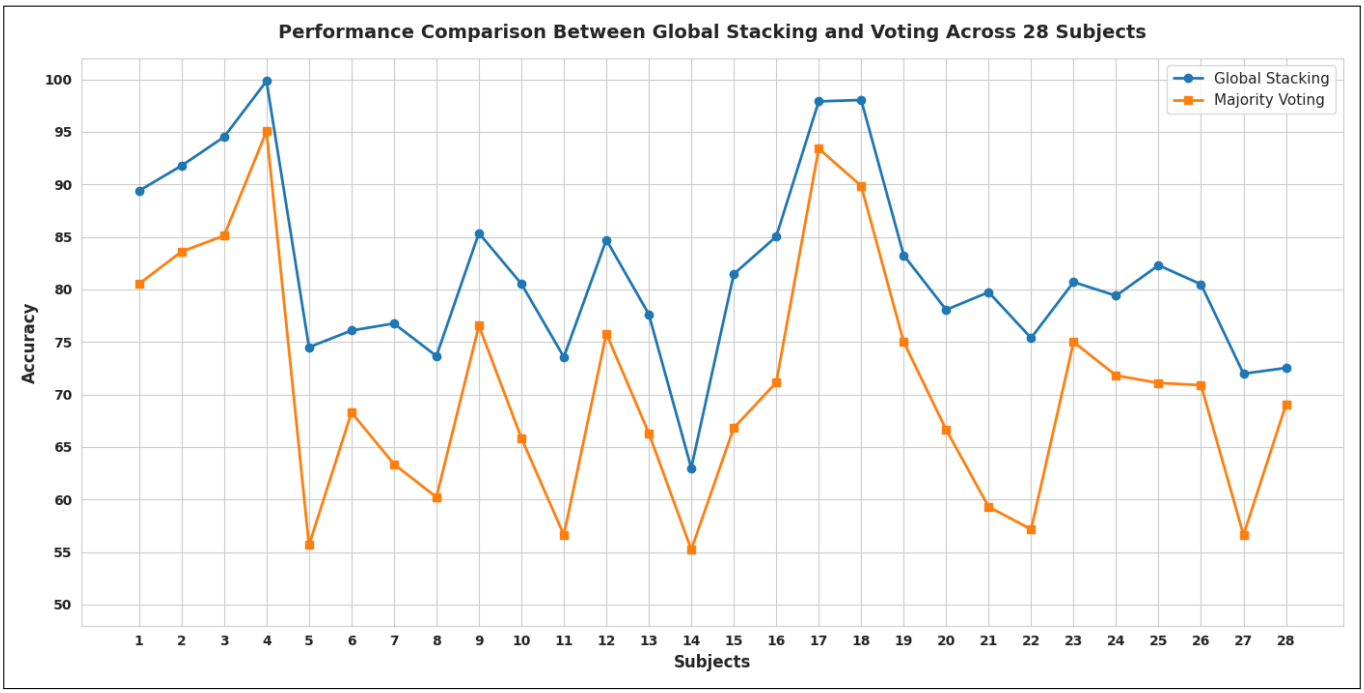


Figure 6.5: Global Stacking vs Majority Voting Accuracy (28 Subjects)(Independent Approach).

Inter-Subject Variability

Ensemble learning methods demonstrated clear superiority over individual models in terms of accuracy and generalization across participants. Figure **6.6** revealed variations in dispersion across ensemble approaches, where Global Stacking achieved the lowest variability and the highest stability, followed by Majority Voting, which exhibited moderate dispersion, reflecting a balance between consistency and variability. Similarly, within the Local Stacking models, Frequency-Mix and Total-AIO exhibited the highest levels of inter-participant variability, whereas Time-Domain-Mix showed lower dispersion, indicating more stable and consistent performance within this type of feature extraction. The aforementioned observations suggested that dispersion reflected differences in performance across participants and did not necessarily indicate model superiority in terms of accuracy.

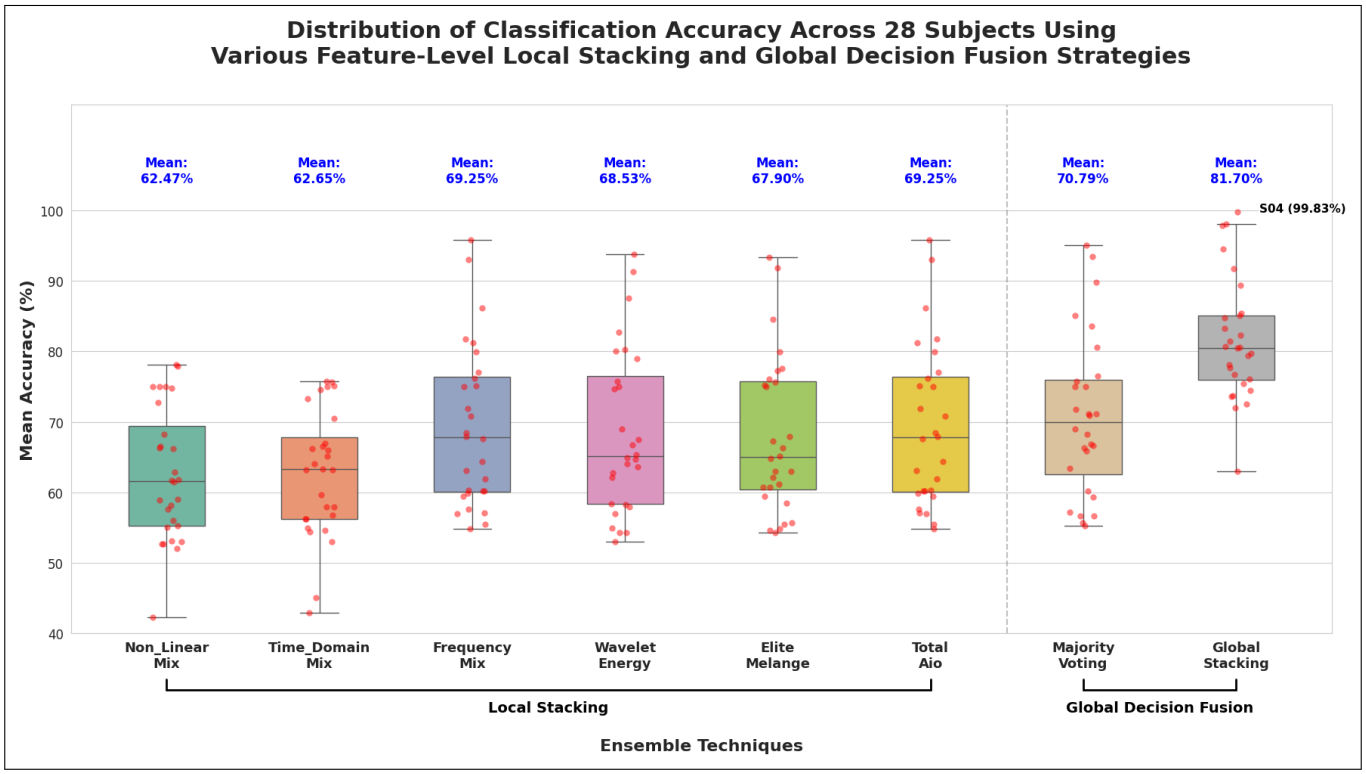


Figure 6.6: Distribution of Classification Accuracy Across 28 Subjects Using Local Stacking & Global Stacking vs Majority Voting (Independent Approach).

6.3.3 Results of Hybrid Models

Table **6.4** presents the subject-wise classification accuracy for various hybrid architectures. All models were trained using StandardScaler normalization and evaluated under different overlap ratios.

Model	Overlap (%)	Best Subject (%)	Worst Subject (%)	Overall Mean (%)
EEGNet + Attention	50	75.33 (S02)	18.33 (S20/S23)	39.83
EEGNet + Fine-Tuning	50	72.90 (S03)	17.08 (S27)	39.04
EEGNet + Attention + Fine-Tuning	75	73.32 (S17)	16.28 (S26)	38.65
DAN	0	73.74 (S09)	21.14 (S27)	40.04
DAN + Augmentation (Gaussian Noise)	0	50.00 (S02)	05.03 (S12)	31.42
DAMGCN + Attention	75	72.31 (S12)	15.28 (S18)	40.56
LaBraM	50	74.33 (S09)	23.00 (S20)	43.68
CNN + BiLSTM + MHA	75	71.38 (S09)	08.41 (S21)	40.91
CNN + BiGRU + MHA	75	73.65 (S03)	15.15 (S26)	46.81
CNN + MHA + BiLSTM	75	86.03 (S17)	23.86 (S26)	51.54

Table 6.4: Average Accuracy (%) for Hybrid Deep Learning Architectures with Corresponding Overlapping Ratios.

Performance Comparison Across Hybrid Deep Learning Models

The results presented in Table **6.4** showed that the overall performance of the models varied according to the architecture and integrated techniques, with values generally remaining within an average range.

Regarding the EEGNet family, whether with the addition of Attention, Fine-Tuning, or a combination of both, the results remained close, with averages of 39.83%, 39.04%, and 38.65%, respectively. This similarity indicates that these modifications did not achieve a significant improvement in performance; rather, they maintained relative stability with slight fluctuation.

The DAN model achieved an average of 40.04%, showing a slight improvement compared to EEGNet. However, applying data augmentation using Gaussian noise led to a clear performance decrease to 31.42%, suggesting that this type of augmentation is not suitable for EEG signals and

negatively affects the model's generalization ability.

The DAMGCN + Attention model also recorded an average of 40.56%, reflecting a modest yet stable improvement. This suggests that integrating graph-based spatial relationships with the attention mechanism contributes partially to performance enhancement.

The LaBraM model achieved better performance, reaching 43.68%, indicating its ability to capture more informative EEG representations. This performance may be attributed to its pretrained architecture, which benefits from strong prior representations and facilitates more effective feature extraction.

Hybrid models based on CNN with temporal modeling units and attention mechanisms showed superior performance under the same Subject-Dependent settings. The CNN + BiLSTM + MHA model achieved an average of 40.91%, while CNN + BiGRU + MHA recorded 46.81%, and the CNN + MHA + BiLSTM model attained the best performance at 51.54%.

The results indicate that improving EEG model performance depends primarily on the strength of the integrated architecture rather than on partial modifications or data augmentation techniques. This is reflected in the superior performance achieved by models combining local and temporal representations with attention mechanisms. Figure **6.7** illustrated these differences in performance among the evaluated models.

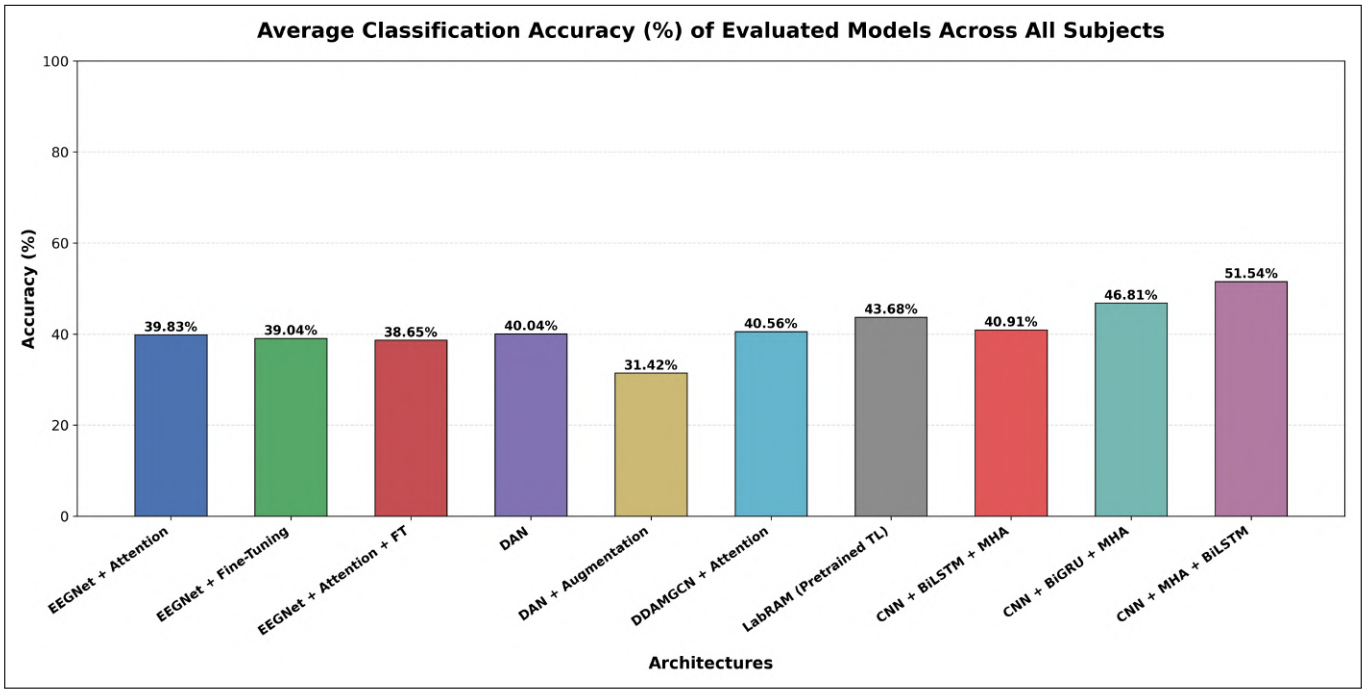


Figure 6.7: Average Classification Accuracy (%) of Hybrid Deep Learning Models Across 28 Subjects (Independent Approach).

Inter-Subject Variability

Inter-Subject variability analysis showed that all evaluated models exhibited clear performance differences across subjects, albeit at varying degrees. Highly complex hybrid architectures (see Appendix **B.10**), such as CNN + MHA + BiLSTM, CNN + BiGRU + MHA, and CNN + BiLSTM + MHA, demonstrated the highest level of variability, where performance ranged from very low values such as 8.41% (CNN + BiLSTM + MHA, S21) and 15.15% (CNN + BiGRU + MHA, S26) to very high values exceeding 70% and reaching a maximum of 86.03% (CNN + MHA + BiLSTM, S17),

indicating strong instability across subjects despite varying overall performance levels.

LaBraM, DAN, and DAMGCN + Attention showed moderate variability, with subject-wise accuracy ranging from approximately 15% to 74%, reflecting noticeable inter-subject differences but less extreme than those observed in more complex hybrid models.

Similarly, EEGNet variants (Attention, Fine-Tuning, and their combination) exhibited comparatively lower variability, with a narrower performance range and more stable behavior across subjects.

The results confirmed that all models were affected by inter-subject variability, which tended to increase with model complexity. Figure **6.8** further illustrated the dispersion and variability across subjects for all models.

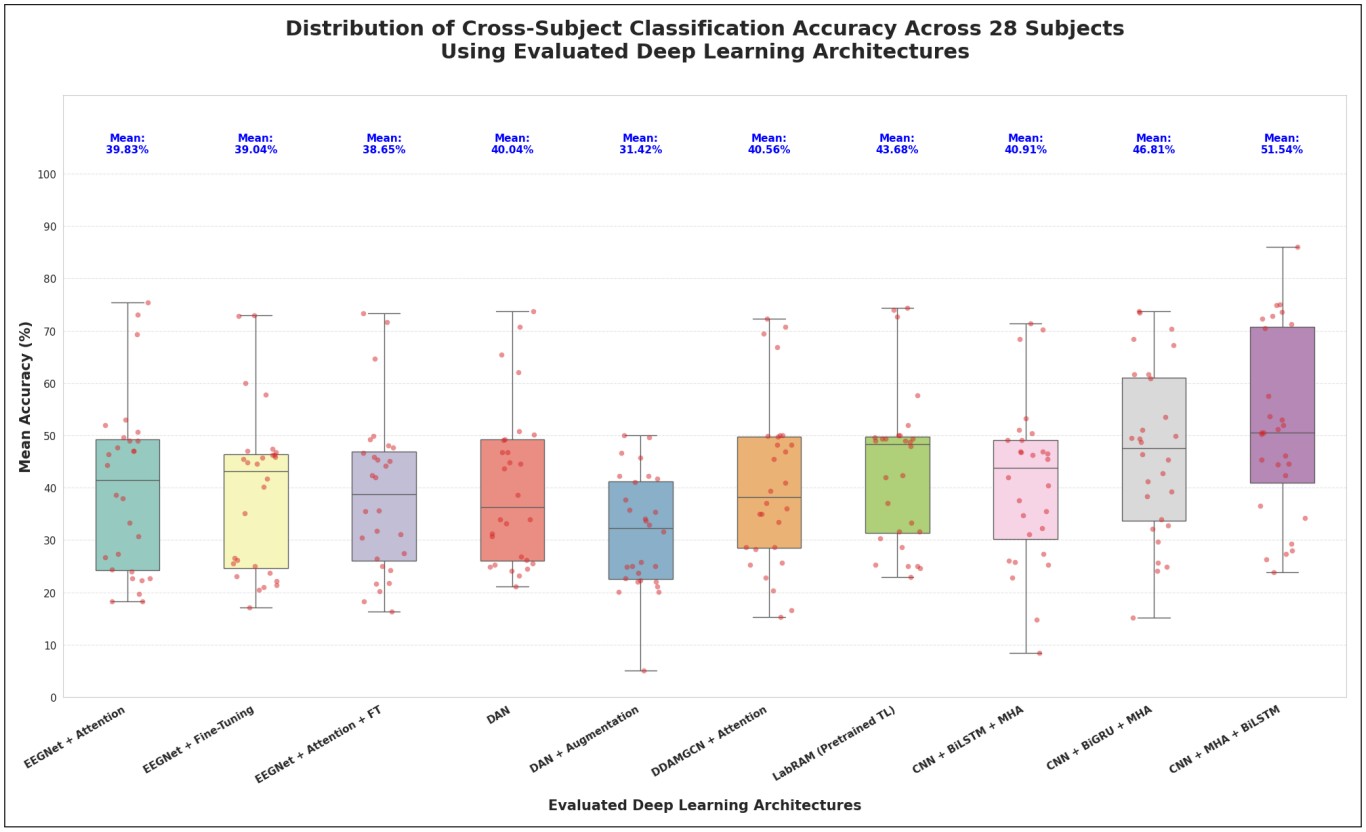


Figure 6.8: Subject-Wise Accuracy Distribution and Outlier Analysis Across Hybrid Deep Learning Models (Independent Approach).

Intra-Subject Variability

The obtained results showed that changing the model architecture could lead to substantial differences in the performance of the same subject. To illustrate this behavior, Subjects S14 and S28 were considered as representative examples. Subject S14 exhibited pronounced fluctuations, with accuracy ranging from 19.67% to 49.54%, indicating high sensitivity to model changes. In contrast, Subject S28 showed relatively more stable behavior, although noticeable variations were still observed, with accuracy ranging from 21.84% to 53.66%. These findings indicate that model architecture plays a significant role in intra-subject performance, resulting in different performance levels across models. Figure **6.9** further illustrates these intra-subject performance variations across the hybrid models.

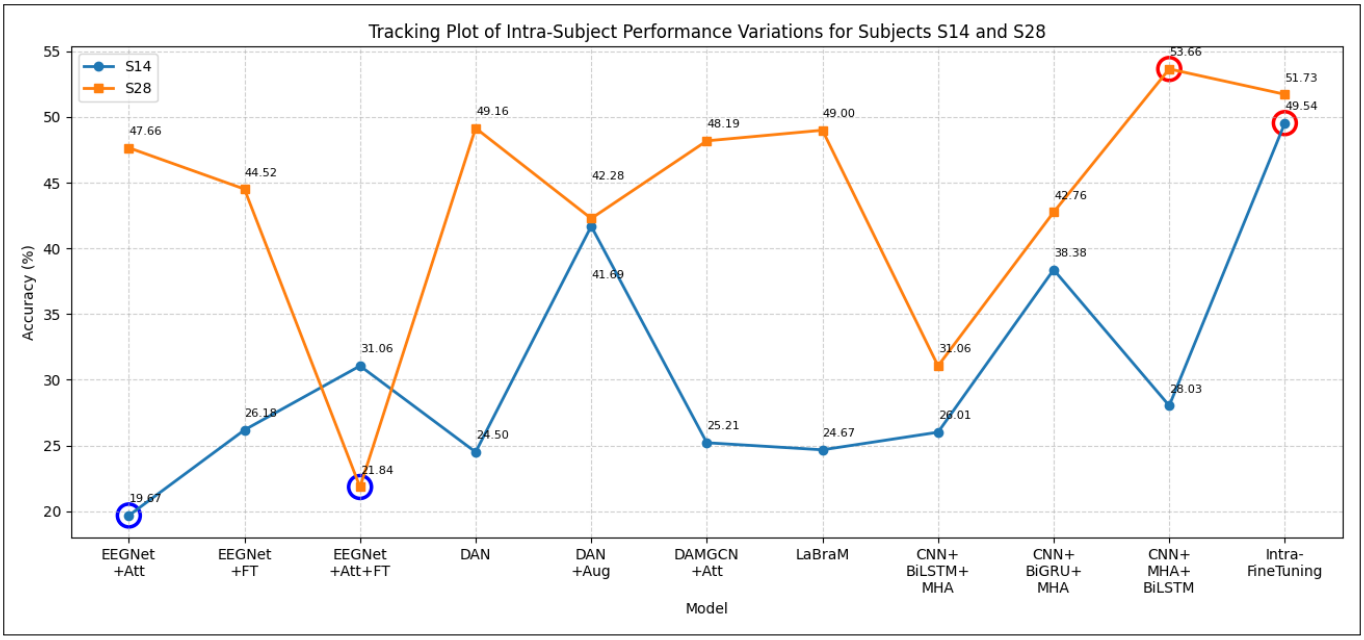


Figure 6.9: Tracking Plot Illustrating Intra-Subject Performance Variations for Subjects S14 and S28 Across Different Hybrid Models.

6.3.4 Results of Self-Training Strategy with Hybrid Deep Learning

The CNN + MHA + BiLSTM model, using RobustScaler preprocessing, achieved the best overall performance (51.54% accuracy) and was selected for further improvement. Following the initial training phase, subjects with accuracies above 70% were identified (S01, S03, S04, S07, S09, S12, S17, S23), and their learned weights were used as a starting point for model initialization. Using this initialized model, a leave-one-subject-out (LOSO) strategy was applied to the remaining 20 subjects, where the model was fine-tuned on the training subjects and evaluated on the held-out subject in each iteration. The fine-tuning strategy resulted in a substantial improvement in performance, increasing the overall mean accuracy from 51.54% to 60.31% (Tables **6.4** and **6.5**).

Subject	Macro Recall (%)	Macro F1-score (%)	Weighted Recall (%)	Weighted F1-score (%)	Accuracy (%)
S02	38.00	42.00	52.00	60.00	52.36
S05	29.00	28.00	50.00	44.00	50.42
S06	31.00	33.00	61.00	67.00	61.36
S08	45.00	47.00	63.00	65.00	63.38
S10	54.00	52.00	69.00	69.00	69.11
S11	49.00	48.00	59.00	61.00	58.92
S13	22.00	24.00	44.00	48.00	44.32
S14	27.00	24.00	50.00	40.00	49.54
S15	77.00	77.00	77.00	77.00	76.77
S16	49.00	43.00	49.00	43.00	49.24
S18	83.00	83.00	83.00	83.00	82.95
S19	32.00	32.00	56.00	64.00	56.48
S20	61.00	61.00	83.00	83.00	83.00
S21	26.00	24.00	46.00	40.00	46.34
S22	47.00	48.00	63.00	67.00	62.92
S24	28.00	31.00	56.00	61.00	55.85
S25	60.00	59.00	80.00	81.00	80.47
S26	41.00	42.00	59.00	60.00	58.96
S27	40.00	37.00	52.00	51.00	52.06
S28	26.00	24.00	52.00	49.00	51.73
MEAN	**44.70**	**44.93**	**59.22**	**58.37**	**60.31**

Table 6.5: Subject-Wise Detailed Classification Performance (Accuracy, Macro and Weighted Metrics) for RobustScaler Across 20 Subjects.

Subject-Wise Emotion Recognition Performance

Table **6.5** shows that the weighted evaluation metrics achieved higher values compared to their macro-averaged counterparts. Specifically, the mean Weighted Recall and Weighted F1-score reached 59.22% and 58.37%, respectively, indicating the model's ability to correctly classify a substantial proportion of samples while maintaining a reasonable balance between precision and recall. In contrast, the Macro Recall and Macro F1-score recorded lower values of 44.70% and 44.93%, respectively, reflecting performance disparities across classes. Moving to the analysis of model performance at the participant level, the results reveal a clear variation in accuracy among individuals, ranging from approximately 44% to over 83%, indicating differences in the model's generalization capability across participants. However, following the fine-tuning stage, a relative improvement in performance stability and a reduction in variability were observed compared to the previous phase. Figure **6.10** provides a visual overview of the performance distribution across subjects.

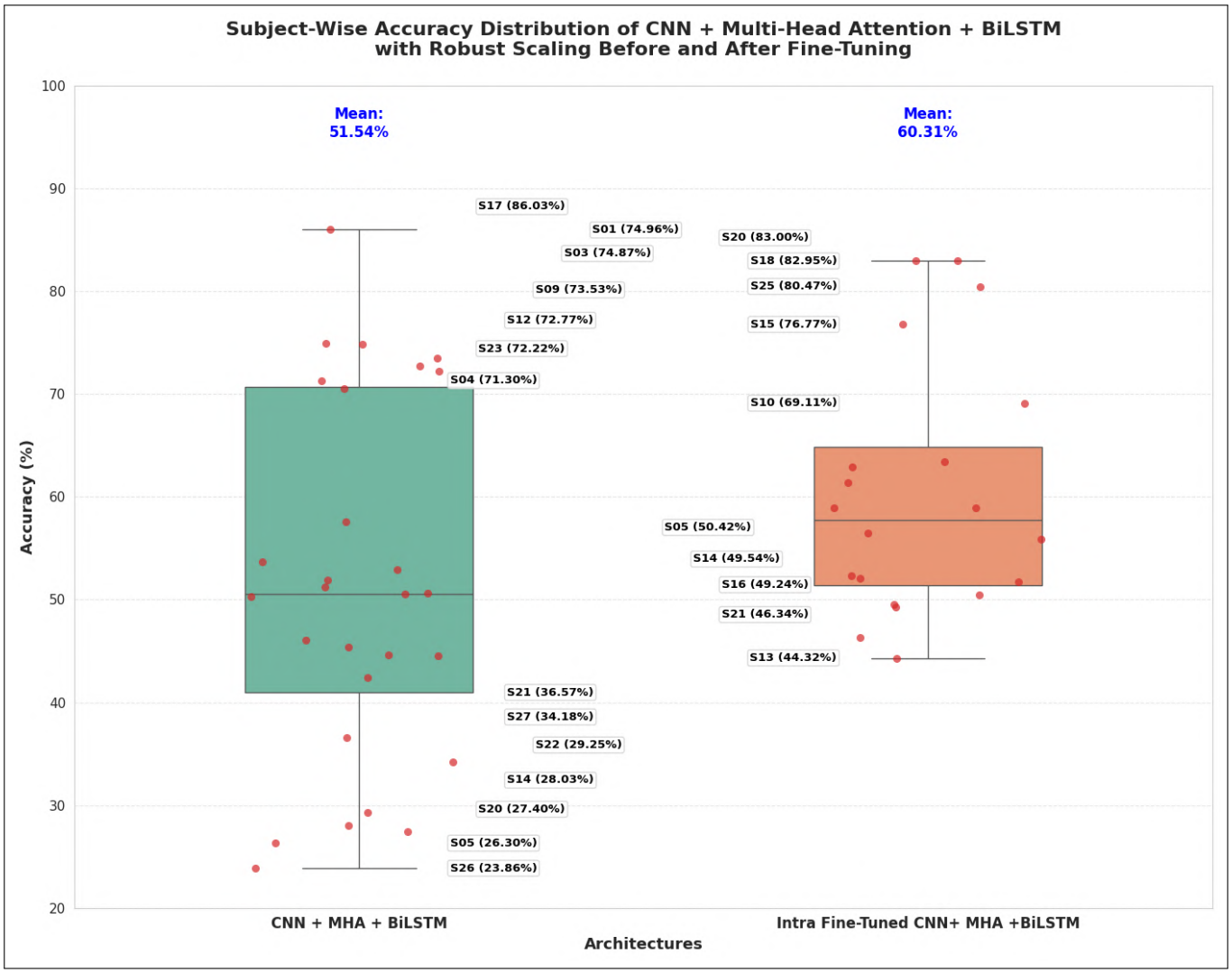


Figure 6.10: Performance Distribution of Intra-Subject Fine-Tuning (Before and After) Using RobustScaler.

Performance Stability After Intra Fine-Tuning

Through visual inspection of Figure **6.10**, it was evident that the Intra Fine-Tuning process improved the distribution of performance across participants, as the lowest accuracy increased compared to the baseline model. In addition, previously observed very low-performance cases, such as S26 and S05, which exhibited accuracies below 30%, were no longer present, and the lowest recorded accuracy remained above 44% (participant S13). These observations suggest reduced variability and greater consistency in performance across participants.

Confusion Matrix-Based Error Analysis

Building on the improved performance stability achieved after Intra-Subject Fine-Tuning, a more detailed class-level analysis was conducted using the confusion matrices of the best-performing model. This analysis aimed to identify misclassification patterns and further assess the model's ability to discriminate between different emotional classes.

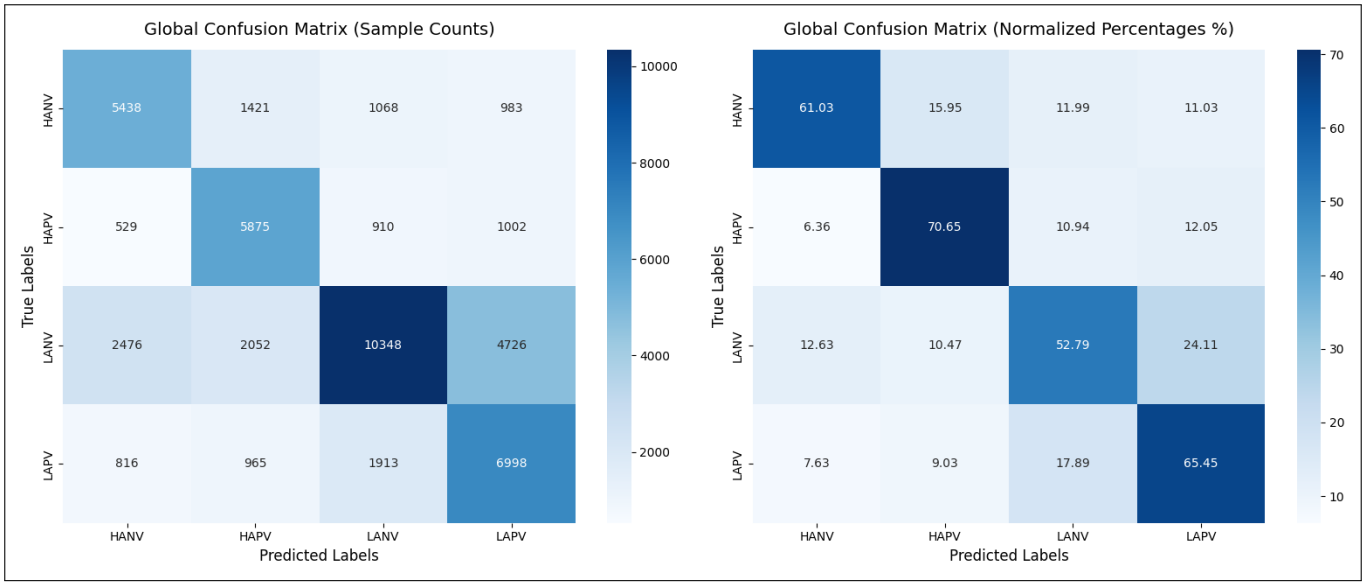


Figure 6.11: Global Classification Performance and Confusion Matrices.

- **Class-Wise Recognition Performance**

  Regarding the recognition performance of individual classes, Figure **6.11** illustrates the normalized confusion matrix and reveals noticeable differences among the classes. The HAPV class achieved the highest classification rate (70.65%), indicating that its EEG patterns were the most distinguishable and consistently recognized by the model. Similarly, the LAPV class exhibited strong performance, with 65.45% of its samples correctly classified. In contrast, HANV achieved a moderate recognition rate of 61.03%, while LANV obtained the lowest classification accuracy (52.79%), making it the most challenging class for the model.

- **Misclassification Error Analysis**

  Beyond the overall class-wise performance, an examination of the misclassification patterns showed that the largest source of error occurred between LANV and LAPV. Specifically, 24.11% of LANV samples were incorrectly classified as LAPV, while 17.89% of LAPV samples were misclassified as LANV. This mutual confusion suggests that these two emotional states share similar EEG characteristics, making them difficult to separate using the extracted features. In comparison, the confusion between HANV and HAPV was lower but still noticeable, with 15.95% of HANV samples being misclassified as HAPV (see Figure **6.11**).

- **Effect of Class Distribution on Recognition Performance**

  To further investigate these results, the relationship between class distribution and recognition performance was examined. Although LANV was the most represented class in the dataset (41.25%, 19,602 samples), it achieved the lowest recognition rate (52.79%). In contrast, HAPV obtained the highest performance (70.65%) despite having a smaller proportion of samples (17.50%, 8,316 samples). Meanwhile, HANV and LAPV accounted for 18.75% and 22.50% of the dataset, respectively, and achieved intermediate recognition rates of 61.03% and 65.45%. These results suggest that class distribution does not necessarily guarantee better classification performance, and that model performance is more strongly influenced by the discriminability of EEG features than by the number of samples in each class.

- **Misclassification Analysis in the Valence–Arousal Space**

  To better understand these misclassifications, the Valence–Arousal scatter plot shown in Figure (Figure **6.12**) was analyzed to examine the spatial distribution of emotional classes. The results revealed a clear overlap between LANV and LAPV in the central region of the affective space, particularly for Valence values between 4 and 6. This pattern was consistently observed across several subjects (see Appendix **C.1**). For example, Subject 20 reported a Valence value of 6, Subject 19 reported a value of 5, and Subject 24 reported a value of 4. Accordingly, LANV and LAPV appear to occupy adjacent regions along the Valence dimension, which likely contributed to the confusion observed between these two classes in the confusion matrix (see Figure **6.11**).

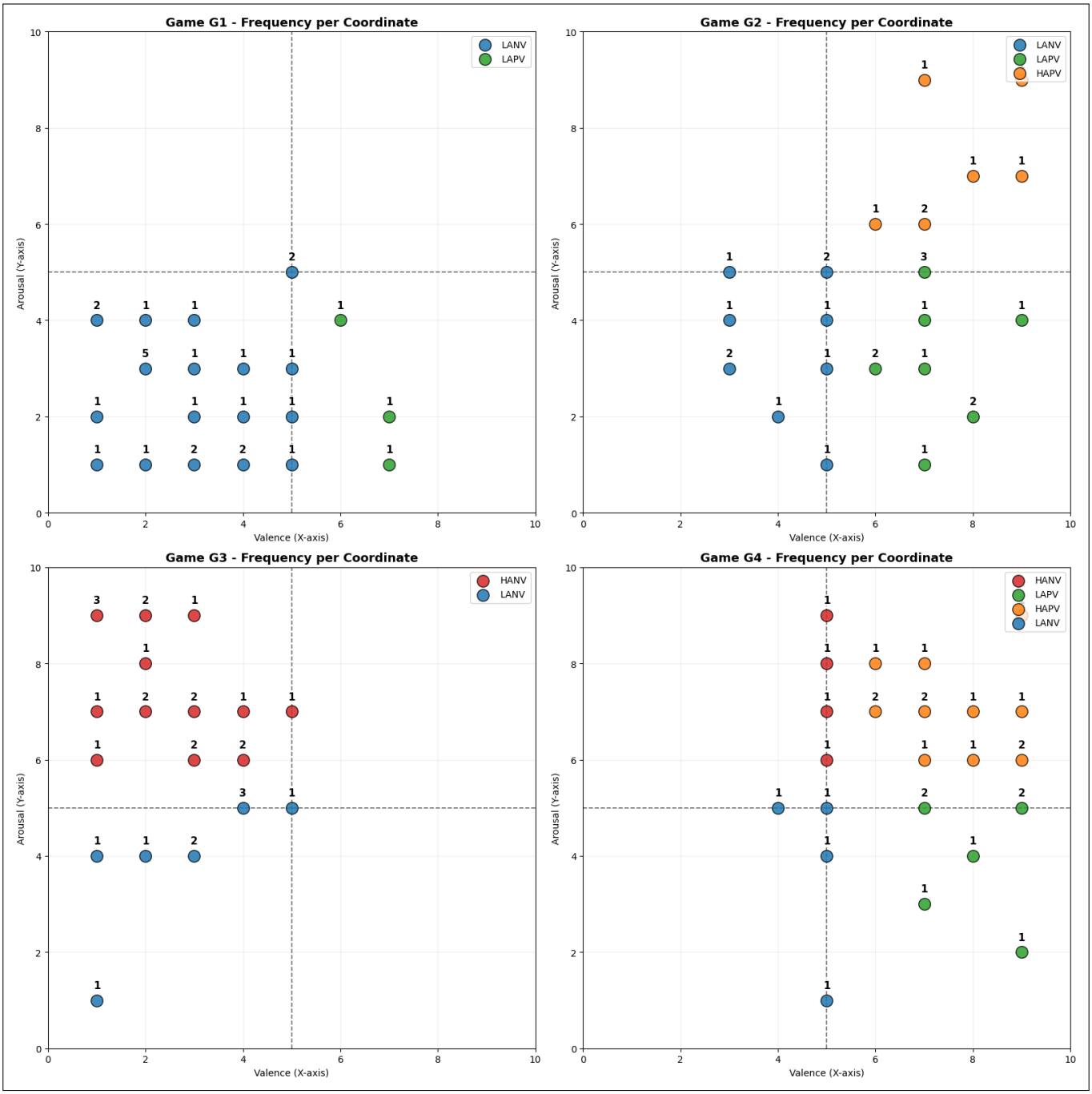


Figure 6.12: Valence–Arousal Coordinate Frequency Distribution and Class Mapping Across the Four Experimental Games.

6.4 Comparison with Literature

Table **6.6** presents a comparison between the proposed method and existing approaches for EEG-based emotion recognition. The evaluation is conducted using commonly used datasets such as DEAP, SEED, and GAMEEMO under Subject-Independent settings, particularly LOSO, in order to analyze and compare the performance of different methods.

Ref	Model	Dataset	Classification	Split	Results
[78]	WST + LDA + KNN + Voting	DEAP	Multi-class (4 classes)	LOSO	Without Voting: 65.67%; With Voting: 97.58%
[73]	DAMGCNN	SEED/SEED-IV	Multi-class (3 / 4 classes)	LOSO	SEED: 73.21%; SEED-IV: 68.22%
[74]	DSSA-Net	SEED/SEED-IV	Multi-class (3 / 4 classes)	LOSO	SEED: 87.03%; SEED-IV: 75.86%
[77]	RF / LSTM-GRU	GAMEEMO	Multi-class (4 classes)	(22train/6test)	RF: 79.6%; LSTM-GRU: 94.5%
[78]	WST + LDA + KNN + Voting	GAMEEMO	Multi-class (4 classes)	LOSO	Without Voting: 79.02%; With Voting: 100%
Proposed Model	**CNN + MHA + BiLSTM**	**GAMEEMO**	**Multi-class (4 classes)**	**LOSO**	**Accuracy: 60.31%**
Proposed Ensemble	**Global Stacking**	**GAMEEMO**	**Multi-class (4 classes)**	**LOSO**	**Accuracy: 81.70%**

Table 6.6: Comparison with Related Work for Subject-Independent EEG-Based Emotion Recognition.

A review of previous studies on EEG-based emotion recognition revealed that different evaluation protocols and model architectures lead to considerable variations in performance. Studies based on non-strict Subject-Independent settings generally reported higher accuracies, whereas stricter LOSO evaluation provided more realistic estimates of model generalization performance.

On the GAMEEMO dataset, RF and LSTM-GRU models [**77**] were evaluated using a non-strict Subject-Independent split involving 22 training subjects and 6 testing subjects. The RF model achieved an accuracy of 79.6%, while the LSTM-GRU model reached 94.5%. Such high performance may be partly explained by the relatively small test set, which may be less representative and include comparatively easier samples. As a result, the reported accuracies may not fully reflect the true generalization capability of these models.

The WST-based hybrid model [**78**] also reported very high accuracies, reaching 100% on GAMEEMO and 97.58% on DEAP after applying voting. Such high performance may be associated with the ensemble strategy, the independent processing of EEG channels, and the use of complete EEG signals without window segmentation. Nevertheless, the perfect accuracy reported on GAMEEMO appears optimistic considering the high variability and noise characteristics of EEG signals.

Under the stricter LOSO protocol, the DAMGCN and DSSA-Net models [**74**] showed more realistic performance, with accuracy ranging from 68% to 87%. Despite the use of advanced techniques such as graph learning and attention mechanisms, the performance decrease observed on the SEED-IV dataset suggests that generalization becomes more difficult in more complex emotion classification scenarios.

Compared to previous studies on emotion recognition using EEG signals, the proposed CNN + MHA + BiLSTM model with self-training and the Global Stacking Ensemble framework demonstrated competitive performance on the GAMEEMO dataset under the strict LOSO evaluation protocol. The former achieved an average accuracy of 60.31%, while the latter reached 81.70%, showing a clear improvement in classification performance and reducing inter-subject variability. These results highlight the effectiveness of the proposed methodologies and emphasize the importance of model fusion in improving Subject-Independent EEG-based emotion recognition. Nevertheless, achieving robust generalization across subjects remains a major challenge due to the variability of EEG signals and the influence of factors such as dataset characteristics, model architecture, and evaluation protocol.

6.5 Conclusion

This chapter provided a comprehensive evaluation of model performance under the Subject-Independent setting, where testing was conducted on unseen participants excluded from the training process. It examined the generalization capability of deep learning, hybrid, and ensemble-based architectures, as well as their effectiveness in mitigating inter-subject variability. The chapter also assessed the contribution of self-training strategies in cross-subject learning. Among the evaluated approaches, the Global Stacking Ensemble achieved the best performance, reaching an average accuracy of 81.70%, demonstrating strong robustness and generalization capability across unseen subjects. In contrast, the CNN + MHA + BiLSTM model with Intra-Subject Fine-Tuning obtained a lower cross-subject performance of approximately 60.31%, highlighting the difficulty of transferring deep learning models to unseen participants without sufficient adaptation. Furthermore, a detailed comparison with related studies was conducted to evaluate the robustness and practical applicability of the proposed approaches for Subject-Independent EEG-based emotion recognition.

General Conclusion

The main problem addressed in this thesis is the limited ability of EEG-based emotion recognition systems to achieve reliable performance across different users in computer gaming environments. This limitation is mainly caused by the high variability of EEG signals between individuals, as well as their noisy and non-stationary nature, which makes it difficult for models to learn stable and consistent patterns of emotional states. As a result, models that perform well on certain subjects often fail to maintain the same performance when applied to unseen users.

To address this problem, this study utilized the GAMEEMO dataset to develop and compare several artificial intelligence approaches, including ML, EL, DL, and hybrid models, while also employing fine-tuning techniques to enhance the performance of certain models. Experiments were conducted under both Subject-Dependent and Subject-Independent settings. The Subject-Dependent setting focuses on learning individual-specific characteristics or "brain signatures," whereas the Subject-Independent setting evaluates the models' ability to generalize to unseen subjects.

The experimental results demonstrated that deep learning models achieved the best performance in the Subject-Dependent setting, with the CNN–MHA–BiLSTM architecture attaining an accuracy of 97.78% after applying Intra-Subject Fine-Tuning. In contrast, the Global Stacking Ensemble framework achieved an accuracy of 81.70% in the Subject-Independent setting, demonstrating superior generalization capability across unseen subjects. Moreover, the proposed approach achieves an improvement of 11.26% and 3.39% over existing state-of-the-art methods in the Subject-Dependent and Subject-Independent settings, respectively.

Overall, the findings indicate that the optimal model selection strongly depends on the evaluation scenario. Deep learning models are more effective in Subject-Dependent settings, whereas ensemble learning methods provide better robustness and generalization in cross-subject scenarios. The study also highlights the importance of preprocessing techniques and feature extraction methods in enhancing the performance of EEG-based emotion recognition systems.

The main contributions of this thesis include a comprehensive evaluation of machine learning, ensemble learning, deep learning, and hybrid approaches for EEG-based emotion recognition. They also include the development of a CNN–MHA–BiLSTM framework with Intra-Subject Fine-Tuning, the development and evaluation of a Global Stacking Ensemble model, and the analysis of Subject-Dependent and Subject-Independent settings.

Despite the encouraging results obtained, certain limitations remain, mainly related to the reliance on a single and relatively small dataset, as well as the inherent variability of EEG signals. Therefore, future work may focus on using larger and more diverse datasets with a greater number of participants and a wider range of experimental conditions, which would allow better learning of inter-subject differences and improve generalization capability. In addition, clustering-based strategies could be explored to group individuals according to their neural characteristics. Furthermore, transfer learning and domain adaptation techniques, as well as the integration of multimodal data,

may be investigated to further enhance performance and improve generalization in real-world applications.

Bibliography

[1] Y. Li, Y. Gu, J. Teng, S. Zheng, Y. Pang, X. Lu, *et al.*, "Advancing eeg-based brain-computer interface technology via pedot: Pss electrodes," *Matter*, vol. 7, no. 9, pp. 2859–2895, 2024.

[2] H. A. Hamzah and K. K. Abdalla, "Eeg-based emotion recognition systems; comprehensive study," *Heliyon*, vol. 10, no. 10, 2024.

[3] C. Cannard, H. Wahbeh, and A. Delorme, "Brainbeats as an open-source eeglab plugin to jointly analyze eeg and cardiovascular signals," *Journal of Visualized Experiments: JoVE*, vol. 206, p. 45, 2024.

[4] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from eeg-based brain–computer interface using machine learning: a comprehensive review," *Neural Computing and Applications*, vol. 34, no. 15, pp. 12527–12557, 2022.

[5] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO," *Biomedical Signal Processing and Control*, vol. 60, p. 101951, 2020.

[6] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[7] H. Liu, Y. Zhang, Y. Li, and X. Kong, "Review on emotion recognition based on electroencephalography," *Frontiers in Computational Neuroscience*, vol. 15, p. 758212, 2021.

[8] X. Chen, F. Kopsaftopoulos, Q. Wu, H. Ren, and F.-K. Chang, "A self-adaptive 1d convolutional neural network for flight-state identification," *Sensors*, vol. 19, no. 2, p. 275, 2019.

[9] Z. Wan, W. Cheng, M. Li, R. Zhu, and W. Duan, "Gdnet-eeg: An attention-aware deep neural network based on group depth-wise convolution for ssvep stimulation frequency recognition," *Frontiers in Neuroscience*, vol. 17, p. 1160040, 2023.

[10] S. Joshi and F. Joshi, "Human emotion classification based on eeg signals using recurrent neural network and knn," *arXiv preprint arXiv:2205.08419*, 2022.

[11] P. Sreehari, U. Raghavendra, and A. Gudigar, "A review of deep learning techniques for eeg-based emotion recognition: Models, methods, and datasets," *F1000Research*, vol. 14, p. 1276, 2025.

[12] A. W. Yimam, M. Vafaeipour, M. Messagie, K. A. Fante, E. M. Molla, T. M. Azerefegn, and T. Coosemans, "State-of-the-art of machine learning methods for fault detection and health monitoring of wind turbine system components: A comprehensive review," *Engineering Applications of Artificial Intelligence*, vol. 162, p. 112645, 2025.

[13] M. Karthiga, E. Suganya, S. Sountharrajan, B. Balusamy, and S. Selvarajan, "Eeg based smart emotion recognition using meta heuristic optimization and hybrid deep learning techniques," *Scientific Reports*, vol. 14, no. 1, p. 30251, 2024.

[14] W.-B. Jiang, L. Zhao, and B.-L. Lu, "Large brain model for learning generic representations with tremendous eeg data in bci," in *International Conference on Learning Representations*, vol. 2024, pp. 16405–16426, 2024.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[17] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: A compact convolutional neural network," in *in 2018 International Conference on Acoustics*, 2018.

[18] H. Zhang, Q. Q. Zhou, H. Chen, X. Q. Hu, W. G. Li, Y. Bai, and X. L. Li, "The applied principles of eeg analysis methods in neuroscience and clinical neurology," *Military Medical Research*, vol. 10, no. 1, p. 67, 2023.

[19] Y. Qin, Y. Zhang, Y. Zhang, S. Liu, and X. Guo, "Application and development of eeg acquisition and feedback technology: A review," *Biosensors*, vol. 13, no. 10, p. 930, 2023.

[20] J. Safai *et al.*, "A comprehensive review of eeg electrode technologies: Advancements, applications, and future directions," *ScienceDirect*, 2025.

[21] W. Kong and X. Jin, "Basics of eeg signals," in *Brain Informatics and Health*, pp. 15–31, Springer Nature, 2025.

[22] A. Chaddad, C. Desrosiers, and M. Toews, "Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques," *Sensors*, vol. 23, no. 14, p. 6434, 2023.

[23] E. T. Attar, "Review of electroencephalography signals approaches for mental stress assessment," *Neurosciences Journal*, vol. 27, no. 4, pp. 209–215, 2022.

[24] M. Rashid, N. Sulaiman, A. PP Abdul Majeed, R. M. Musa, A. F. Ab. Nasir, B. S. Bari, and S. Khatun, "Current status, challenges, and possible solutions of eeg-based brain-computer interface: a comprehensive review," *Frontiers in neurorobotics*, vol. 14, p. 25, 2020.

[25] R. F. Lyu, "Deep learning approaches for eeg-based healthcare applications: A comprehensive review," *Frontiers in Human Neuroscience*, vol. 19, p. 1689073, 2026.

[26] R. Guo, H. Guo, L. Wang, M. Chen, D. Yang, and B. Li, "Development and application of emotion recognition technology — a systematic literature review," *BMC Psychology*, vol. 12, p. 95, 2024.

[27] S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations," *Information Fusion*, vol. 102, p. 102019, 2024.

[28] M. X. Lim and J. Teo, "Predicting game-induced emotions using EEG, data mining and machine learning," *Bulletin of the National Research Centre*, vol. 48, no. 1, p. 57, 2024.

[29] L. A. Martínez-Tejada, A. Puertas-González, N. Yoshimura, and Y. Koike, "Exploring eeg characteristics to identify emotional reactions under videogame scenarios," *Brain Sciences*, vol. 11, no. 3, p. 378, 2021.

[30] C. A. U. Hassan *et al.*, "EmoTrans: Attention-based emotion recognition," *Scientific Reports*, 2025.

[31] J. V. M. R. Fernandes *et al.*, "Emotion detection from EEG signals using machine deep learning models," *Bioengineering*, vol. 11, no. 8, p. 782, 2024.

[32] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE transactions on affective computing*, vol. 12, no. 2, pp. 479–493, 2018.

[33] C. Y. Park, N. Cha, S. Kang, A. Kim, A. H. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, "K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations," *Scientific Data*, vol. 7, no. 1, p. 293, 2020.

[34] V. Jadhav, N. Tiwari, and M. Chawla, "EEG-based emotion recognition using transfer learning based feature extraction and convolutional neural network," in *Proceedings of the International Conference on Signal and Information Processing*, 2023.

[35] T. B. Alakus and I. Turkoglu, "Emotion recognition with deep learning using GAMEEMO data set," *Electronics Letters*, vol. 56, no. 25, 2020.

[36] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, vol. 16, no. 1-6, p. 1, 2008.

[37] J. Lichtenauer and M. Soleymani, "Mahnob-hci-tagging database," 2011.

[38] J. Chen, X. Wang, C. Huang, X. Hu, X. Shen, and D. Zhang, "A large finer-grained affective computing eeg dataset," *Scientific Data*, vol. 10, no. 1, p. 740, 2023.

[39] Q. Yang, Y. Zhang, S. Chen, Y. Xu, J. Sun, and Z. Ma, "Mped: Quantifying point cloud distortion based on multiscale potential energy discrepancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6037–6054, 2022.

[40] S. Pant, H.-J. Yang, E. Lim, S.-H. Kim, and S.-B. Yoo, "Phymer: physiological dataset for multimodal emotion recognition with personality as a context," *IEEE Access*, vol. 11, pp. 107638–107656, 2023.

[41] Z. Lan, Y. Liu, O. Sourina, L. Wang, R. Scherer, and G. Müller-Putz, "Safe: An eeg dataset for stable affective feature selection," *Advanced Engineering Informatics*, vol. 44, p. 101047, 2020.

[42] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "Ascertain: Emotion and personality recognition using commercial sensors," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 147–160, 2016.

[43] Y. Ma, Y. Liu, and W. Zhang, "A comprehensive review of deep learning in EEG-based emotion recognition: Classifications, trends, and practical implications," *PeerJ Computer Science*, vol. 10, p. e1821, 2024.

[44] Y. Liu, W. Xue, L. Yang, and M. Li, "Deep learning-based EEG emotion recognition: A review," *Brain Sciences*, vol. 16, no. 1, p. 41, 2025.

[45] N. R. Babu and V. Viswanathan, "Mfenet: A multi-feature extraction network for enhanced emotion detection using eeg and stft," *IEEE Access*, 2025.

[46] Z. Fan, F. Chen, X. Xia, and Y. Liu, "EEG emotion classification based on graph convolutional network," *Applied Sciences*, vol. 14, no. 2, p. 726, 2024.

[47] A. Abdulrahman, M. Baykara, and T. B. Alakus, "A novel approach for emotion recognition based on EEG signal using deep learning," *Applied Sciences*, 2022.

[48] M. Algarni, S. Alshahrani, *et al.*, "Deep learning-based approach for emotion recognition using electroencephalography (eeg) signals using bi-directional long short-term memory (bi-lstm)," *Applied Sciences*, vol. 12, no. 3, p. 1234, 2022.

[49] M. Jafari *et al.*, "Emotion recognition in EEG signals using deep learning methods: A review," *Computers in Biology and Medicine*, vol. 165, p. 107450, 2023.

[50] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.

[51] V. Doma and M. Pirouz, "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals," *Journal of Big Data*, vol. 7, no. 1, p. 18, 2020.

[52] T. Bayes, "An essay towards solving a problem in the doctrine of chances," *Biometrika*, vol. 45, no. 3-4, pp. 296–315, 1958.

[53] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[54] M. J. Margaret, N. M. Banu, S. Madhumithaa, and A.-S. K. Pathan, "On the prediction models for brain signal-based emotion recognition," *Scientific Reports*, 2026.

[55] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[56] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[57] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Chapman and Hall/CRC, 2017.

[58] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[59] M. A. Blanco-Ríos, M. O. Candela-Leal, C. Orozco-Romo, P. Remis-Serna, C. S. Vélez-Saboyá, J. d. J. Lozoya-Santos, M. Cebral-Loureda, and M. A. Ramírez-Moreno, "Real-time eeg-based emotion recognition for neurohumanities: perspectives from principal component analysis and tree-based algorithms," *Frontiers in Human Neuroscience*, vol. 18, p. 1319574, 2024.

[60] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[61] R. Hasan and S. M. R. Islam, "A comparative analysis of emotion recognition from eeg signals using temporal features and hyperparameter-tuned machine learning techniques," *MethodsX*, vol. 15, p. 103468, 2025.

[62] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[63] F. Zhou, T. Chen, and B. Lei, "Cross-subject emotion recognition with sparsely-labeled peripheral physiological data using shap-explained tree ensembles," *arXiv preprint arXiv:2211.02923*, 2022.

[64] R. Alhalaseh and S. Alasasfeh, "Machine-learning-based emotion recognition system using EEG signals," *Computers*, vol. 9, no. 4, p. 95, 2020.

[65] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[66] Y. Kim and A. Choi, "Eeg-based emotion classification using long short-term memory network with attention mechanism," *Sensors*, vol. 20, no. 23, p. 6727, 2020.

[67] K. Cho, B. Van Merriënboer, Ç. Gulçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1724–1734, 2014.

[68] G. A. V. M. Giri and M. L. Radhitya, "Electroencephalogram-based emotion classification using machine learning and deep learning techniques," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 18, no. 3, 2024.

[69] X. Du, Y. Meng, S. Qiu, Y. Lv, and Q. Liu, "Eeg emotion recognition by fusion of multi-scale features," *Brain Sciences*, vol. 13, no. 9, p. 1293, 2023.

[70] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[71] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*, pp. 1180–1189, PMLR, 2015.

[72] W. Lu, H. Liu, H. Ma, T.-P. Tan, and L. Xia, "Hybrid transfer learning strategy for cross-subject eeg emotion recognition," *Frontiers in Human Neuroscience*, vol. 17, p. 1280241, 2023.

[73] W. Chen, Y. Liao, R. Dai, Y. Dong, and L. Huang, "Eeg-based emotion recognition using graph convolutional neural network with dual attention mechanism," *Frontiers in Computational Neuroscience*, vol. 18, p. 1416494, 2024.

[74] J. Liu, L. He, H. Chen, and D. Jiang, "Directional spatial and spectral attention network (DSSA-Net) for EEG-based emotion recognition," *Frontiers in Neurorobotics*, vol. 18, p. 1481746, 2025.

[75] F. Cui *et al.*, "A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition," *Mathematics*, 2022.

[76] W. Lu, H. Ma, and T.-P. Tan, "CIT-EmotionNet: CNN interactive transformer network for EEG emotion recognition," *arXiv preprint arXiv:2305.05548*, 2023.

[77] A. Rehman, I. Heldal, and J. C.-W. Lin, "Spatiotemporal eeg-based emotion recognition using sam ratings from serious games with hybrid deep learning," *arXiv preprint arXiv:2508.21103*, 2025.

[78] A. Elrefaiy, N. Tawfik, N. Zayed, and I. Elhenawy, "Subject-independent multi-channel voting for eeg-based emotion recognition using wavelet scattering deep network and advanced signal metrics," *Pattern Analysis and Applications*, vol. 28, no. 3, p. 130, 2025.

Subject-Dependent Classification Results

Machine Learning-Based Models Results

- **KNN**

Table A.1: KNN-Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	22.78	32.93	45.72	34.94	20.43	32.72	31.59	29.19	50.25	37.92	30.79	52.26	40.65	33.18	27.68	36.03	33.05	31.59	38.09	25.34	28.69	27.98	51.22	35.02	44.76	32.84	38.51	42.24	**35.30**
Time-Domain-Mix	18.08	40.14	28.73	37.63	19.63	22.69	25.80	25.29	51.47	36.03	33.64	51.26	40.44	32.42	24.03	28.36	20.68	25.17	28.52	18.96	33.35	20.72	55.91	37.58	42.53	32.21	41.07	41.02	**32.62**
Frequency-Mix	36.49	25.00	27.56	46.27	19.97	29.87	26.26	24.25	56.50	37.29	36.03	51.30	37.50	32.26	28.06	37.50	30.41	46.94	36.37	17.03	37.96	21.27	53.44	33.68	39.09	29.24	39.76	42.58	**35.00**
Wavelet-Energy	24.75	35.32	25.46	33.26	20.93	25.67	27.73	26.47	54.15	36.70	37.33	49.04	40.86	35.32	28.52	32.42	28.61	31.04	30.91	18.29	40.81	23.74	58.47	36.33	38.72	34.69	38.09	44.25	**34.21**
Elite-Melange	17.58	26.13	27.89	39.97	19.84	24.58	25.21	23.41	53.73	37.71	31.84	52.01	41.11	32.47	26.09	34.90	26.88	37.37	35.61	17.95	35.11	19.38	55.41	37.79	41.19	32.05	40.77	39.77	**33.35**
Total-AIO	24.79	35.32	25.46	33.22	20.97	25.71	27.73	26.43	54.15	36.66	37.29	49.16	40.81	35.36	28.57	32.42	28.61	31.04	30.91	18.37	40.81	23.74	58.47	36.33	38.67	34.69	38.13	44.13	**34.21**

- **LDA**

Table A.2: LDA-Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	57.59	51.55	26.89	39.56	28.73	43.54	40.23	46.64	73.99	45.39	49.29	68.54	49.79	29.78	26.17	30.54	26.13	31.04	33.52	23.32	35.65	25.84	73.24	47.86	48.95	49.33	46.73	53.10	**42.96**
Time-Domain-Mix	47.02	48.70	40.73	30.24	22.78	36.70	31.29	48.53	73.41	46.27	50.08	69.76	47.82	28.06	26.64	26.38	29.57	26.43	39.01	5.03	20.05	22.73	69.42	49.92	48.28	45.93	48.24	52.27	**40.40**
Frequency-Mix	15.06	33.35	48.45	40.27	27.43	42.70	24.75	48.28	62.12	47.61	48.70	59.52	56.12	34.40	29.61	31.12	49.66	28.31	17.70	28.52	34.52	22.53	74.25	36.49	49.66	35.57	40.52	30.75	**39.21**
Wavelet-Energy	41.78	28.15	51.05	34.35	24.45	41.02	25.34	43.62	74.20	47.19	51.89	71.85	41.02	28.19	33.39	35.40	45.39	34.19	3.86	10.19	23.57	17.32	68.79	35.70	35.65	38.67	24.29	49.50	**37.85**
Elite-Melange	31.96	29.74	49.87	45.18	28.82	44.46	39.77	48.20	62.79	47.94	52.06	63.34	56.29	35.82	30.41	33.64	52.01	27.47	34.86	33.43	29.61	23.95	73.45	35.91	49.92	33.72	40.60	24.66	**41.42**
Total-AIO	36.54	34.27	58.89	38.34	25.67	44.88	41.78	50.04	60.28	46.77	43.62	36.79	46.39	35.61	35.19	36.33	56.92	33.39	29.91	35.78	36.12	25.38	72.40	37.21	48.78	29.66	32.30	40.10	**41.04**

- **SVM**

Table A.3: SVM-Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	87.00	76.93	92.24	90.35	84.69	77.98	79.87	74.63	90.35	87.00	75.68	86.79	80.92	62.89	70.02	70.02	92.45	88.67	82.80	81.13	81.76	87.63	79.66	82.18	86.37	74.00	73.79	70.64	**81.02**
Time-Domain-Mix	92.03	81.76	95.59	98.95	88.46	84.06	89.09	84.48	92.66	91.19	83.01	87.63	83.43	69.39	79.45	79.24	93.50	97.06	85.11	85.11	90.98	88.05	84.90	87.42	90.35	78.82	80.50	74.42	**86.31**
Frequency-Mix	95.60	81.55	96.44	98.11	81.76	79.87	80.71	80.92	88.46	87.42	77.77	84.27	82.59	61.42	88.67	79.87	94.33	96.64	84.69	85.32	83.22	88.67	82.80	84.27	86.37	81.13	78.61	72.11	**84.41**
Wavelet-Energy	97.06	81.97	93.92	98.74	88.88	81.76	80.08	84.06	92.45	90.14	81.97	87.21	81.97	59.32	88.26	78.40	94.75	96.64	85.32	88.67	83.22	87.84	86.16	88.05	90.77	88.05	78.19	70.23	**85.86**
Elite-Melange	96.86	83.22	97.06	98.32	89.93	86.16	88.88	85.11	92.03	91.19	86.58	88.05	86.16	66.66	88.67	84.07	94.54	96.85	88.88	86.58	91.19	89.72	87.21	86.37	92.87	84.07	81.13	74.84	**87.97**
Total-AIO	97.90	88.46	96.64	100.00	90.14	86.16	87.42	86.16	94.33	91.40	84.90	89.72	86.58	70.02	93.29	85.53	94.75	97.48	88.05	92.66	90.56	93.08	85.74	88.46	93.29	89.93	78.82	74.84	**89.15**

- **RandomForestClassifier**

Table A.4: RandomForestClassifier-Classification accuracy (%) per subject (S01–S28).

features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	91.61	86.16	93.50	98.53	85.53	80.71	83.01	80.29	87.63	88.05	82.38	86.16	88.26	64.15	77.14	79.24	92.24	93.50	84.69	84.48	82.59	85.74	79.45	87.84	86.16	80.50	77.56	74.84	**84.36**
Time-Domain-Mix	93.08	85.74	94.54	98.95	87.00	83.01	84.90	80.71	92.45	91.19	84.69	90.35	89.51	68.55	79.87	78.82	91.40	92.45	85.11	86.79	87.42	88.67	85.11	91.40	87.21	84.27	81.97	77.56	**86.53**
Frequency-Mix	96.44	87.63	95.60	98.95	85.95	83.65	83.44	83.44	89.73	88.68	80.92	86.37	88.05	60.59	91.61	84.07	95.39	97.69	84.91	86.16	85.12	87.42	80.29	88.05	85.95	84.28	78.41	72.54	**86.12**
Wavelet-Energy	94.34	85.32	93.08	99.58	88.68	87.42	80.71	86.58	93.71	92.66	85.12	89.94	91.19	66.66	90.77	80.50	94.33	98.53	92.24	87.42	83.43	88.88	88.46	90.98	90.14	87.84	80.50	74.21	**87.97**
Elite-Melange	96.65	86.37	95.18	98.95	88.47	84.70	85.53	81.76	90.36	90.15	82.60	87.42	90.15	63.10	88.89	81.76	94.13	97.48	87.42	85.53	87.21	87.84	84.07	89.94	87.63	82.60	78.83	73.17	**86.71**
Total-AIO	96.65	88.05	96.02	99.37	91.40	88.05	90.15	88.26	93.29	91.82	86.16	92.03	88.26	68.13	90.36	87.00	94.34	98.95	89.73	90.99	90.15	91.20	87.42	92.24	89.73	87.42	81.34	74.84	**89.41**

- **XGBoost**

Table A.5: XGBoost – Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	93.50	88.26	95.17	98.32	87.84	84.27	87.21	83.43	90.98	89.51	82.18	88.67	90.98	67.92	77.77	80.92	93.29	94.75	87.63	87.42	86.79	87.84	82.18	88.88	89.72	81.34	80.92	73.79	**86.48**
Time-Domain-Mix	95.38	89.09	95.59	98.53	92.24	89.51	88.88	86.58	95.59	93.29	86.58	93.71	92.24	77.35	85.32	81.97	93.92	94.75	89.93	92.24	90.35	92.03	87.63	91.40	93.29	89.72	84.90	81.55	**90.13**
Frequency-Mix	97.06	91.19	95.81	98.74	89.51	89.30	87.21	85.95	93.08	91.82	85.53	90.14	88.67	65.61	93.08	86.58	96.01	97.48	90.35	90.56	89.93	90.14	86.58	89.72	87.63	85.11	84.27	74.00	**88.97**
Wavelet-Energy	97.69	85.95	93.50	99.79	94.13	91.19	86.16	89.51	96.01	95.80	89.51	92.66	92.45	70.02	93.50	84.27	94.75	98.74	93.71	92.66	88.67	91.82	92.03	94.33	93.08	90.98	85.74	76.51	**90.90**
Elite-Melange	98.32	88.67	95.81	98.95	91.19	88.26	88.67	88.05	93.92	93.08	85.11	91.40	89.30	65.82	92.03	83.22	95.80	97.90	91.82	89.09	89.09	89.09	87.63	89.72	91.19	85.53	83.22	72.95	**89.10**
Total-AIO	98.53	92.66	95.38	99.79	93.50	92.45	92.66	91.40	96.64	96.01	90.56	94.75	92.45	72.53	94.33	89.93	96.22	98.53	94.96	94.33	92.87	93.29	90.77	94.75	94.75	90.14	87.42	79.03	**92.52**

- **ExtraTreesClassifier**

Table A.6: ExtraTreesClassifier – Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	89.93	84.06	93.71	98.74	82.18	80.71	80.92	78.19	85.53	89.30	82.18	85.53	88.88	64.57	75.26	78.19	92.03	93.08	84.06	79.45	81.97	87.63	79.24	89.51	84.27	79.24	76.93	73.37	**83.52**
Time-Domain-Mix	92.66	84.70	94.34	98.95	86.37	84.90	83.01	81.97	91.82	90.98	82.38	87.84	90.77	69.39	77.77	77.14	91.40	91.40	85.53	87.42	85.53	88.67	81.55	90.98	88.05	82.38	79.66	77.35	**85.89**
Frequency-Mix	95.60	86.16	96.22	98.95	81.76	86.79	78.61	81.97	88.88	87.00	78.82	85.11	88.26	59.95	89.73	82.18	95.60	98.11	84.90	86.16	79.87	81.55	79.87	87.42	83.64	77.98	76.72	70.85	**84.60**
Wavelet-Energy	95.18	84.28	93.71	99.37	88.26	87.21	79.24	86.58	91.40	93.92	82.18	89.30	89.73	63.94	90.14	82.38	94.96	98.53	89.30	87.42	83.64	85.74	85.53	93.92	88.67	88.88	81.34	72.11	**87.39**
Elite-Melange	95.39	84.28	95.60	98.95	86.37	84.28	83.65	83.23	91.19	89.94	79.45	85.74	87.84	62.47	88.05	83.02	94.34	97.69	84.48	84.69	84.48	87.84	82.59	88.47	87.84	81.34	77.14	73.17	**85.84**
Total-AIO	96.44	87.21	96.02	99.37	88.88	88.88	86.37	86.16	93.29	92.87	85.32	90.98	90.98	66.66	89.94	86.58	94.96	98.11	86.58	90.56	87.21	90.14	85.53	92.24	90.14	86.58	83.22	75.68	**88.82**

- **LightGBM**

Table A.7: LightGBM – Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	93.29	89.10	94.75	98.74	88.05	81.97	86.37	81.97	90.98	89.94	81.55	88.46	88.88	68.13	77.56	82.59	93.71	94.54	87.21	87.84	87.63	88.26	82.18	88.26	89.72	82.38	79.45	74.21	**86.35**
Time-Domain-Mix	94.33	90.14	95.81	98.95	92.66	88.67	88.67	84.48	95.59	92.66	88.05	93.08	93.29	76.10	86.16	82.18	93.71	95.17	89.72	91.82	91.40	92.66	87.63	90.77	94.54	88.88	85.95	79.04	**90.08**
Frequency-Mix	97.69	91.61	96.44	98.53	88.68	89.10	88.05	85.53	93.08	92.03	85.32	90.77	88.67	64.36	93.50	87.00	96.44	98.32	89.30	90.56	90.98	90.56	85.53	90.35	90.14	85.53	84.27	74.00	**89.16**
Wavelet-Energy	96.44	86.16	93.29	99.58	93.92	90.14	85.32	89.73	96.01	95.38	90.14	91.61	91.82	71.27	93.71	82.18	95.59	98.74	94.34	92.24	88.05	92.66	91.61	93.92	93.08	90.98	84.90	74.21	**90.61**
Elite-Melange	98.11	89.10	95.81	98.95	90.99	85.95	87.63	86.37	93.92	91.40	86.16	91.82	89.73	66.87	92.66	84.07	95.59	98.11	90.15	90.35	88.67	89.30	86.37	89.10	91.61	85.32	81.97	74.84	**88.96**
Total-AIO	98.32	92.66	96.65	99.58	93.71	91.40	91.40	90.56	96.64	94.75	91.61	93.08	92.03	72.95	93.71	88.67	97.07	99.16	94.34	94.54	93.08	92.66	90.35	94.75	94.75	92.03	87.00	78.83	**92.37**

Deep Learning-Based Models Results

- **MLPNN:**

Table A.8: MLPNN (overlap 50%) - Classification accuracy (%) per subject (S01–S28) & fold.

Subject	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	54.17	29.17	0.00	70.83	54.17	4.17	0.00	0.00	95.83	4.17	70.83	0.00	0.00	0.00	25.00	16.67	87.50	0.00	0.00	0.00	4.17	16.67	0.00	41.67	95.83	0.00	20.83	33.33
FOLD 2	79.17	87.50	100.0	95.83	83.33	66.67	66.67	45.83	100.0	91.67	95.83	37.50	66.67	33.33	66.67	20.83	91.67	29.17	33.33	83.33	50.00	54.17	12.50	66.67	95.83	45.83	70.83	75.00
FOLD 3	73.91	56.52	78.26	60.87	52.17	43.48	52.17	69.57	86.96	78.26	69.57	95.65	39.13	17.39	60.87	17.39	73.91	52.17	82.61	30.43	65.22	47.83	60.87	47.83	56.52	86.96	43.48	39.13
FOLD 4	100.0	0.00	17.39	100.0	4.35	21.74	73.91	73.91	100.0	65.22	82.61	82.61	0.00	0.00	30.43	30.43	52.17	0.00	95.65	52.17	4.35	26.09	91.30	82.61	30.43	95.65	0.00	73.91
FOLD 5	95.65	78.26	95.65	91.30	73.91	52.17	34.78	56.52	100.0	73.91	91.30	91.30	95.65	47.83	91.30	8.70	100.0	86.96	100.0	78.26	86.96	52.17	100.0	100.0	91.30	100.0	47.83	73.91
FOLD 6	39.13	4.35	26.09	100.0	13.04	65.22	52.17	13.04	69.57	34.78	0.00	86.96	47.83	13.04	0.00	8.70	91.30	52.17	95.65	65.22	43.48	47.83	86.96	39.13	65.22	43.48	43.48	21.74
FOLD 7	0.00	65.22	65.22	100.0	78.26	91.30	30.43	69.57	43.48	73.91	17.39	82.61	26.09	8.70	82.61	30.43	95.65	100.0	100.0	91.30	95.65	100.0	91.30	65.22	95.65	91.30	86.96	17.39
FOLD 8	73.91	43.48	91.30	60.87	69.57	56.52	60.87	60.87	47.83	56.52	47.83	95.65	52.17	43.48	56.52	47.83	100.0	47.83	91.30	60.87	82.61	86.96	78.26	73.91	91.30	65.22	30.43	73.91
FOLD 9	60.87	39.13	78.26	47.83	43.48	8.70	0.00	0.00	0.00	82.61	8.70	100.0	78.26	95.65	4.35	8.70	100.0	0.00	82.61	0.00	43.48	52.17	69.57	52.17	52.17	86.96	82.61	39.13
FOLD 10	95.65	21.74	86.96	95.65	82.61	91.30	47.83	65.22	26.09	86.96	95.65	95.65	100.0	95.65	86.96	39.13	100.0	86.96	91.30	95.65	56.52	82.61	100.0	95.65	78.26	100.0	69.57	78.26
Mean	**67.24**	**42.54**	**63.91**	**82.32**	**55.49**	**50.13**	**41.88**	**45.45**	**66.97**	**64.80**	**57.97**	**76.79**	**50.58**	**35.51**	**50.47**	**22.88**	**89.22**	**45.53**	**77.25**	**55.72**	**53.24**	**56.65**	**69.08**	**66.49**	**75.25**	**71.54**	**49.60**	**52.57**

Overall Mean Accuracy	**58.47%**

Table A.9: MLPNN (overlap 90%) - Classification accuracy (%) per subject (S01–S28) & fold.

Subject	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	74.14	33.19	4.74	70.26	75.00	38.79	74.57	9.91	92.24	1.72	64.22	17.67	0.00	4.31	38.79	17.67	89.22	7.33	34.48	61.21	86.21	40.09	0.00	48.71	95.26	43.53	23.28	35.34
FOLD 2	85.34	87.07	100.0	93.97	77.16	78.02	59.91	74.57	100.0	90.52	95.69	54.31	73.71	31.90	78.45	30.60	91.81	34.05	75.00	87.07	66.81	66.81	46.12	71.55	96.55	58.19	100.0	97.41
FOLD 3	69.70	45.45	76.19	61.04	76.19	74.89	50.22	76.62	86.15	88.31	73.16	98.27	42.86	23.38	62.77	24.68	71.00	96.54	86.15	61.04	77.92	57.58	95.24	67.97	60.17	79.65	47.62	54.98
FOLD 4	100.0	15.58	83.12	100.0	25.97	44.59	76.62	80.52	98.70	72.29	81.39	81.39	1.30	2.16	55.41	48.05	90.04	66.67	99.13	70.56	59.31	35.06	98.70	86.15	32.47	97.84	4.33	75.32
FOLD 5	98.70	91.77	99.13	91.77	95.24	88.31	42.86	83.12	100.0	78.79	75.32	99.57	100.0	45.89	99.13	25.97	100.0	97.40	100.0	93.94	91.77	55.84	100.0	100.0	83.12	100.0	69.70	70.56
FOLD 6	69.26	19.91	42.86	94.81	41.13	83.98	44.16	6.06	88.31	51.95	3.46	89.61	59.74	19.05	0.00	29.87	87.45	68.83	94.81	67.10	30.74	64.07	87.88	51.95	66.23	64.50	51.08	25.11
FOLD 7	82.68	85.28	97.84	100.0	100.0	97.40	50.65	90.04	49.78	81.39	38.53	98.27	36.36	26.84	72.29	52.38	99.13	100.0	100.0	86.15	99.57	99.57	86.58	76.62	99.13	81.82	84.85	37.23
FOLD 8	82.25	43.72	95.24	58.87	81.39	61.47	73.16	73.59	63.64	58.44	51.52	99.13	57.58	48.92	58.87	55.84	100.0	49.78	86.15	70.99	87.01	90.91	81.39	75.32	92.21	64.94	45.89	70.99
FOLD 9	95.67	52.38	90.48	73.16	53.25	71.43	25.97	63.64	38.53	100.0	27.71	100.0	85.28	93.51	8.66	19.91	100.0	28.14	81.39	64.50	47.62	50.22	74.03	90.91	62.34	95.67	78.79	46.32
FOLD 10	100.0	18.61	90.48	77.92	100.0	100.0	70.56	68.83	33.33	92.64	83.55	99.57	100.0	93.07	92.64	51.08	100.0	90.91	96.54	98.27	67.53	80.09	96.97	100.0	78.35	100.0	88.74	85.28
Mean	**85.78**	**49.30**	**78.01**	**82.18**	**72.53**	**73.89**	**56.87**	**62.69**	**75.07**	**71.61**	**59.45**	**83.78**	**55.68**	**38.90**	**56.70**	**35.61**	**92.87**	**63.96**	**85.36**	**76.08**	**71.45**	**64.02**	**76.69**	**76.92**	**76.58**	**78.61**	**59.43**	**59.86**

Overall Mean Accuracy	**68.56%**

- **BiLSTM:**

Table A.10: BiLSTM (Raw) – Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	47.06	51.26	68.91	49.16	40.76	33.61	55.88	44.12	78.15	55.04	39.08	75.21	57.14	50.00	28.57	17.65	60.92	27.73	68.07	51.26	49.16	67.65	73.95	42.86	67.23	39.50	36.97	65.13
FOLD 2	83.19	46.64	76.05	85.29	61.34	62.61	56.72	52.94	74.37	45.80	56.72	68.91	61.34	61.34	54.62	47.48	76.47	65.13	63.03	55.46	88.66	60.08	76.89	43.28	68.07	57.98	58.40	71.01
FOLD 3	90.34	69.75	92.02	86.97	67.65	68.07	49.16	39.50	73.11	57.14	64.29	71.43	60.08	47.48	59.24	73.53	92.86	73.95	72.27	58.40	74.37	58.82	75.63	49.16	74.79	62.18	62.18	63.45
FOLD 4	71.43	63.45	96.64	97.06	64.71	84.45	42.86	50.00	70.59	60.92	63.87	73.95	70.59	57.98	67.23	73.11	94.54	88.66	78.57	73.95	67.23	71.01	73.53	50.42	79.41	61.76	39.50	68.49
FOLD 5	84.45	65.97	80.25	92.44	80.25	62.18	57.98	59.66	82.35	52.10	53.36	62.61	54.62	55.04	62.18	60.92	78.15	92.02	76.47	64.29	73.95	72.69	71.43	44.54	80.67	54.62	51.26	73.53
FOLD 6	64.71	54.20	85.29	94.96	54.62	47.48	65.55	63.87	78.15	60.08	68.91	80.67	68.91	48.32	71.43	73.11	88.24	91.18	78.99	67.65	64.71	76.05	76.05	63.45	69.75	44.54	50.42	55.04
FOLD 7	85.23	83.54	89.03	97.89	75.53	75.53	86.08	68.35	75.53	75.11	78.06	78.06	65.82	54.01	59.92	75.95	84.39	89.03	85.23	81.01	46.84	77.64	70.46	68.78	66.67	58.23	56.54	70.46
FOLD 8	75.11	94.09	85.65	73.84	76.79	78.06	76.79	73.42	67.09	66.67	72.15	61.60	72.57	59.92	70.46	75.53	77.64	89.87	82.70	76.79	57.81	73.84	81.01	56.12	54.85	50.63	61.60	72.57
FOLD 9	72.15	80.59	84.81	90.72	73.84	76.37	71.31	67.09	75.11	59.92	46.84	70.04	82.70	75.11	66.24	54.01	83.54	82.70	91.14	76.37	42.19	63.29	80.59	62.03	72.15	46.84	60.34	82.28
FOLD 10	74.68	77.64	87.34	83.97	79.75	62.87	56.96	73.00	75.11	51.90	70.89	62.45	76.79	65.40	72.57	56.54	86.08	76.37	83.97	61.18	36.29	73.00	79.32	43.88	70.46	46.84	54.85	85.23
MEAN	**74.83**	**68.71**	**84.60**	**85.23**	**67.52**	**65.12**	**61.93**	**59.19**	**74.95**	**58.47**	**61.42**	**70.49**	**67.06**	**57.46**	**61.25**	**60.78**	**82.28**	**77.66**	**78.04**	**66.64**	**60.12**	**69.41**	**75.89**	**52.45**	**70.41**	**52.31**	**53.21**	**70.72**

Overall Mean Accuracy	**67.43%**

Table A.11: BiLSTM (Preprocessed) – Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	61.76	13.87	62.61	30.67	42.44	60.92	60.92	52.10	79.83	52.52	37.39	74.79	59.24	47.90	31.09	7.14	58.40	48.74	73.11	53.78	61.76	65.55	76.47	68.49	69.75	36.55	38.66	56.72
FOLD 2	78.15	48.74	74.79	90.34	65.13	54.62	89.92	65.13	81.93	53.36	55.04	71.01	55.88	58.82	41.60	54.62	81.09	76.47	74.79	59.24	84.87	64.71	79.83	73.53	79.41	51.68	58.82	70.59
FOLD 3	93.70	62.61	93.70	93.28	70.17	63.87	77.73	60.08	76.47	64.71	63.87	68.49	57.14	54.20	73.95	73.11	97.48	94.12	75.21	65.13	75.63	66.81	82.77	63.03	74.37	60.08	62.18	79.41
FOLD 4	86.13	78.99	97.90	96.64	68.91	66.81	73.95	84.45	78.99	60.08	57.98	67.65	62.18	49.16	77.73	71.85	94.12	91.18	75.63	77.31	72.27	81.09	73.95	73.11	78.99	53.36	50.42	78.57
FOLD 5	86.13	81.51	84.45	97.06	77.73	54.62	81.93	76.05	83.19	64.71	57.56	69.33	60.08	54.20	78.15	67.65	85.29	92.02	84.87	69.75	76.05	76.89	81.51	46.64	77.73	49.16	50.84	81.09
FOLD 6	71.01	77.73	94.12	65.55	57.56	39.92	68.07	48.32	77.31	53.36	76.47	59.66	48.32	42.02	79.41	76.47	81.93	93.28	92.86	50.84	67.23	79.83	75.63	46.64	76.05	47.90	54.62	61.34
FOLD 7	86.92	81.01	92.83	96.20	73.00	76.37	72.57	51.90	79.32	59.92	80.17	77.64	58.23	48.95	73.42	86.92	92.41	97.05	87.76	84.39	75.11	82.28	65.82	51.05	63.29	59.92	64.56	67.51
FOLD 8	83.97	89.45	73.84	79.32	83.12	73.84	55.70	67.09	75.11	73.42	70.46	71.31	62.87	69.62	82.28	84.81	78.90	95.78	77.22	73.84	60.76	84.81	74.68	56.54	71.31	64.14	70.04	62.03
FOLD 9	83.54	81.43	89.03	93.67	76.37	74.26	51.90	68.78	79.75	56.12	54.85	69.20	86.08	72.15	83.12	57.38	94.09	89.45	75.53	69.62	56.96	70.46	78.90	48.10	65.40	50.63	70.46	69.62
FOLD 10	70.46	59.07	88.61	94.09	78.06	69.62	49.79	66.24	74.26	53.59	59.07	75.53	67.93	65.82	75.95	58.23	92.41	78.48	86.08	57.81	54.43	81.01	82.70	46.41	69.20	55.27	57.81	66.24
MEAN	**80.18**	**67.44**	**85.19**	**83.68**	**69.25**	**63.48**	**68.25**	**64.01**	**78.62**	**59.18**	**61.29**	**70.46**	**61.80**	**56.28**	**69.67**	**63.82**	**85.61**	**85.66**	**80.31**	**66.17**	**68.51**	**75.34**	**77.23**	**57.35**	**72.55**	**52.87**	**57.84**	**69.31**

Overall Mean Accuracy	**69.69%**

- **Deep BiLSTM:**

Table A.12: Deep BiLSTM (Spectral Entropy) – Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	57.98	32.77	59.66	53.36	49.58	52.52	52.52	51.26	76.47	50.00	33.61	74.79	68.49	43.70	30.67	18.91	63.03	57.98	72.69	53.36	52.10	61.76	75.63	68.07	67.23	46.64	50.00	57.98
FOLD 2	81.93	38.66	72.69	89.92	63.45	66.81	74.79	63.45	78.15	52.52	56.30	74.79	54.20	50.00	39.50	47.06	81.51	73.53	77.31	55.88	85.71	62.18	80.25	73.11	68.07	50.84	54.20	64.29
FOLD 3	88.24	71.43	96.64	92.44	70.17	65.55	69.33	57.98	75.21	54.62	60.08	74.79	50.84	47.48	71.01	68.49	91.18	88.24	75.21	69.33	79.83	71.85	81.09	63.03	74.79	57.14	63.87	78.99
FOLD 4	84.45	73.53	88.66	94.96	68.07	73.95	76.89	82.35	76.89	51.26	60.92	74.79	60.50	50.42	76.89	70.59	92.86	92.86	75.21	73.11	72.69	81.09	80.25	71.01	81.51	63.45	59.24	78.99
FOLD 5	93.28	84.03	85.71	90.76	71.85	52.52	70.59	71.43	80.25	56.30	57.56	68.91	59.66	52.52	75.63	65.13	84.45	94.54	82.77	71.43	63.03	75.21	78.99	62.18	77.73	53.78	57.56	77.31
FOLD 6	75.63	67.23	92.02	68.91	51.68	39.50	59.66	52.52	77.31	55.88	71.85	73.11	52.10	48.32	81.51	78.99	82.35	93.28	93.70	48.32	67.23	79.41	75.21	45.80	76.05	50.84	57.14	52.10
FOLD 7	75.53	70.46	91.14	96.20	75.95	78.06	70.04	62.45	75.11	59.49	74.68	75.11	59.07	48.10	74.68	78.90	90.72	96.20	84.81	78.06	67.51	75.53	73.42	46.84	66.67	62.87	65.40	71.31
FOLD 8	78.06	87.76	75.53	81.43	80.17	75.95	58.65	65.40	75.95	62.45	62.45	75.11	66.24	62.45	81.86	81.01	78.48	96.20	77.22	70.89	70.46	75.95	78.06	45.15	66.24	54.85	68.78	61.60
FOLD 9	86.08	84.39	89.87	95.78	77.22	77.22	63.29	69.62	74.26	56.12	48.10	75.11	76.37	67.09	77.64	64.56	90.30	90.30	82.70	69.62	55.70	71.31	79.75	49.37	69.62	51.48	57.81	66.24
FOLD 10	70.04	68.78	88.19	88.19	72.15	68.35	55.27	72.57	78.06	59.49	58.23	75.11	62.87	57.38	62.45	56.54	79.75	81.86	75.11	60.34	59.92	75.95	75.11	56.12	63.71	50.21	51.48	67.93
MEAN	**79.12**	**67.90**	**84.01**	**85.19**	**68.03**	**65.04**	**65.10**	**64.90**	**76.77**	**55.81**	**58.38**	**74.16**	**61.04**	**52.75**	**67.18**	**63.02**	**83.46**	**86.50**	**79.67**	**65.03**	**67.42**	**73.02**	**77.78**	**58.07**	**71.16**	**54.21**	**58.55**	**67.68**

Overall Mean Accuracy	**68.96%**

Table A.13: Deep BiLSTM (statistical features) – Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	86.25	90.00	82.50	87.50	80.00	76.25	80.00	71.25	87.50	82.50	66.25	81.25	72.50	61.25	66.25	70.00	95.00	73.75	81.25	91.25	81.25	83.75	75.00	81.25	75.00	67.50	56.25	77.50
FOLD 2	82.50	75.00	87.50	86.25	72.50	83.75	81.25	83.75	87.50	81.25	65.00	85.00	75.00	57.50	65.00	67.50	90.00	71.25	81.25	82.50	82.50	87.50	75.00	80.00	73.75	77.50	58.75	73.75
FOLD 3	87.34	83.54	87.34	79.75	79.75	81.01	77.22	82.28	83.54	82.28	63.29	87.34	75.95	65.82	67.09	68.35	87.34	75.95	78.48	83.54	82.28	81.01	75.95	86.08	72.15	69.62	65.82	77.22
FOLD 4	88.61	86.08	87.34	82.28	77.22	84.81	81.01	78.48	88.61	84.81	77.22	82.28	74.68	56.96	64.56	69.62	89.87	78.48	75.95	83.54	89.87	93.67	77.22	77.22	77.22	74.68	72.15	73.42
FOLD 5	82.28	77.22	92.41	86.08	79.75	88.61	81.01	68.35	79.75	82.28	77.22	78.48	72.15	65.82	72.15	73.42	89.87	79.75	87.34	87.34	84.81	92.41	74.68	81.01	77.22	75.95	64.56	73.42
FOLD 6	81.01	84.81	87.34	86.08	77.22	74.68	83.54	75.95	81.01	86.08	74.68	77.22	77.22	55.70	64.56	74.68	91.14	79.75	84.81	78.48	78.48	86.08	75.95	87.34	72.15	68.35	55.70	62.03
FOLD 7	82.28	83.54	86.08	87.34	73.42	83.54	72.15	78.48	86.08	87.34	62.03	83.54	68.35	64.56	63.29	68.35	91.14	75.95	82.28	78.48	79.75	86.08	74.68	86.08	73.42	71.20	58.23	79.75
FOLD 8	93.67	87.34	93.67	91.14	67.09	79.75	79.75	83.54	81.01	77.22	65.82	82.28	77.22	55.70	64.56	74.68	89.87	77.22	75.95	83.54	83.54	89.87	77.22	88.61	74.68	71.20	68.35	64.56
FOLD 9	89.87	77.22	91.14	82.28	79.75	84.81	81.01	75.95	84.81	79.75	74.68	86.08	60.76	53.16	82.28	79.75	82.28	77.22	86.08	88.61	79.75	89.87	77.22	82.28	70.89	65.82	59.49	82.28
FOLD 10	88.61	87.34	89.87	81.01	67.09	74.68	75.95	74.68	87.34	79.75	72.15	82.28	78.48	59.49	67.09	69.62	89.87	68.35	83.54	86.08	74.68	84.81	69.62	89.87	74.68	62.03	59.49	70.89
MEAN	**86.24**	**83.21**	**88.52**	**84.97**	**75.38**	**81.19**	**79.29**	**77.27**	**84.72**	**82.32**	**69.83**	**82.57**	**73.23**	**59.60**	**67.68**	**71.60**	**89.64**	**75.77**	**81.69**	**84.34**	**81.69**	**87.50**	**75.25**	**83.97**	**74.12**	**69.94**	**61.88**	**73.48**

Overall Mean Accuracy	**78.10%**

Table A.14: Deep BiLSTM (EMD+VMD+Statistical Features) - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	55.04	25.21	69.75	57.14	42.02	56.72	70.59	55.88	88.66	31.93	63.87	74.79	73.53	52.52	15.97	25.21	72.27	24.79	70.17	56.72	74.79	65.97	68.49	45.80	73.11	65.13	31.09	53.36
FOLD 2	97.90	67.23	78.57	95.38	73.11	62.61	79.83	78.57	88.66	75.21	72.69	74.37	78.57	59.66	56.72	46.22	86.97	92.86	86.13	75.21	82.77	69.33	89.08	86.13	81.93	74.37	47.06	69.75
FOLD 3	97.48	84.03	96.64	100.0	72.27	67.65	76.89	77.31	88.66	76.47	78.99	86.13	62.61	51.68	94.96	82.77	97.90	92.44	88.66	78.57	82.77	78.99	81.09	84.45	93.70	90.76	73.95	88.24
FOLD 4	86.13	75.21	95.38	100.0	85.29	73.11	78.57	70.17	80.25	86.55	83.61	87.39	55.46	60.92	94.54	77.73	98.74	96.22	86.55	78.15	78.15	76.05	76.05	79.41	94.12	83.19	76.89	71.01
FOLD 5	84.87	86.55	89.92	98.32	85.71	58.82	75.63	94.96	92.02	93.70	77.73	83.19	78.15	56.30	91.60	86.55	91.18	98.32	90.34	81.51	79.41	87.39	74.79	68.49	91.18	86.13	83.19	72.27
FOLD 6	83.61	54.20	97.90	60.50	76.89	76.47	80.67	47.90	92.86	87.39	72.69	84.03	44.96	47.48	95.38	87.82	89.92	98.74	94.96	83.61	80.25	86.97	92.44	64.71	89.92	57.56	77.31	63.87
FOLD 7	95.78	88.61	96.62	100.0	81.01	89.45	72.15	75.95	87.76	91.14	88.61	86.50	70.89	61.60	94.09	80.17	96.20	100.0	91.14	90.30	78.06	80.59	79.75	83.54	86.50	83.12	73.00	71.31
FOLD 8	94.51	86.92	85.65	98.31	86.08	86.92	68.78	81.86	81.43	90.30	85.65	94.51	95.78	58.23	90.30	84.39	86.08	97.47	78.90	83.97	68.78	91.98	77.22	87.34	88.19	78.90	76.79	67.93
FOLD 9	95.78	83.12	92.41	99.16	88.61	86.50	60.76	71.73	69.20	78.90	58.65	91.14	97.05	67.93	91.14	59.49	94.09	94.94	78.48	83.54	60.76	84.81	81.86	95.78	82.28	77.64	70.89	64.56
FOLD 10	77.22	52.74	97.47	92.83	89.45	83.12	64.56	79.32	73.42	64.98	69.62	79.75	90.72	66.67	72.15	58.65	97.89	85.65	84.81	78.06	73.42	89.45	84.81	96.62	78.48	84.81	59.92	75.11
MEAN	**86.83**	**70.38**	**90.03**	**90.16**	**78.04**	**74.14**	**72.84**	**73.37**	**84.29**	**77.66**	**75.21**	**84.18**	**74.77**	**58.30**	**79.68**	**68.90**	**91.12**	**88.14**	**85.01**	**78.96**	**75.92**	**81.15**	**80.56**	**79.23**	**85.94**	**78.16**	**67.01**	**69.74**

Overall Mean Accuracy	**78.56%**

Ensemble Learning-Based Models Results

- **Stacking**

Table A.15: Stacking – Classification accuracy (%) per subject (S01–S28).

Features	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	Mean
Non-Linear-Mix	93.92	87.84	94.97	98.74	88.89	84.91	88.89	84.28	93.29	90.36	83.65	89.94	90.99	70.02	79.87	84.49	94.97	94.55	87.84	87.84	87.42	89.31	83.23	88.89	90.78	85.32	80.92	75.47	**87.56**
Time-Domain-Mix	96.02	89.52	96.23	98.74	92.24	89.31	90.99	87.21	96.86	93.71	88.47	94.76	92.24	77.99	85.12	84.07	93.92	95.81	90.57	93.08	91.82	92.66	88.47	92.66	94.76	90.36	85.32	79.87	**90.81**
Frequency-Mix	97.27	92.24	96.64	98.95	89.52	89.73	89.52	86.58	93.08	93.08	84.69	90.15	89.94	65.83	93.92	87.63	96.64	98.53	90.78	91.40	90.99	90.15	85.95	90.15	90.15	85.95	84.49	72.12	**89.50**
Wavelet-Energy	97.90	86.79	94.34	99.79	94.76	91.40	87.00	90.15	96.02	95.81	89.10	94.13	93.29	72.54	94.13	84.07	95.39	98.95	94.34	92.45	89.52	94.34	92.45	94.55	92.66	91.20	87.63	76.52	**91.47**
Elite-Melange	98.74	89.93	96.02	98.95	91.20	87.63	89.52	87.00	94.55	92.45	87.00	92.03	90.15	67.30	92.66	86.79	96.02	98.53	91.61	90.36	89.10	89.94	87.42	91.61	92.03	86.16	84.49	75.05	**89.80**
Total-AIO	98.74	93.08	96.64	99.79	93.29	91.82	92.87	92.03	97.06	96.22	90.56	94.96	92.24	75.05	95.38	91.40	97.27	98.95	95.80	94.75	93.92	93.08	91.82	94.13	94.75	91.19	88.05	79.03	**93.00**

- **Global Stacking (Fusion) vs Majority Voting**

Table A.16: Fusion vs Voting – Classification accuracy (%) per subject (S01–S28).

Method	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	Mean
Fusion	99.58	93.50	97.90	100.00	96.86	94.55	95.18	94.97	97.90	97.48	94.13	96.44	94.13	84.07	97.69	94.13	97.90	100.00	96.86	97.27	96.86	96.23	92.66	96.44	96.23	95.81	92.03	81.97	**95.31**
Voting	98.74	90.99	97.06	99.16	93.08	91.40	92.03	90.78	96.86	95.18	88.89	94.55	92.24	75.47	94.13	91.40	96.44	98.74	93.71	93.71	94.97	93.08	90.99	93.50	94.34	91.82	88.26	77.15	**92.45**

Hybrid Deep Learning-Based Models Results

- **CNN + LSTM**

Table A.17: CNN + LSTM - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	70.00	50.00	80.00	11.67	36.67	48.33	86.67	66.67	75.00	21.67	55.00	75.00	50.00	35.00	28.33	26.67	73.33	25.00	75.00	53.33	90.00	80.00	75.00	50.00	75.00	60.00	55.00	51.67
FOLD 2	100.0	26.67	78.33	90.00	25.00	53.33	95.00	75.00	70.00	25.00	26.67	71.67	50.00	50.00	28.33	25.00	68.33	63.33	75.00	81.67	95.00	90.00	75.00	50.00	86.67	53.33	56.67	58.33
FOLD 3	100.0	69.49	100.0	96.61	23.73	54.24	79.66	98.31	61.02	61.02	72.88	84.75	50.85	42.37	25.42	89.83	69.49	89.83	74.58	57.63	89.83	98.31	74.58	49.15	93.22	83.05	79.66	71.19
FOLD 4	100.0	88.14	98.31	98.31	86.44	55.93	74.58	98.31	77.97	91.53	25.42	76.27	50.85	49.15	25.42	79.66	96.61	89.83	64.41	25.42	84.75	98.31	74.58	49.15	88.14	55.93	69.49	50.85
FOLD 5	52.54	79.66	93.22	89.83	25.42	55.93	77.97	94.92	69.49	98.31	44.07	88.14	84.75	49.15	23.73	88.14	100.0	96.61	74.58	71.19	91.53	100.0	74.58	49.15	96.61	71.19	72.88	74.58
FOLD 6	93.22	47.46	100.0	59.32	25.42	55.93	79.66	84.75	88.14	79.66	55.93	72.88	50.85	49.15	25.42	94.92	100.0	98.31	74.58	89.83	94.92	89.83	74.58	49.15	88.14	62.71	37.29	49.15
FOLD 7	100.0	100.0	98.31	100.0	25.42	52.54	93.22	77.97	98.31	25.42	25.42	83.05	62.71	50.85	25.42	100.0	96.61	100.0	74.58	77.97	96.61	88.14	74.58	50.85	74.58	79.66	47.46	52.54
FOLD 8	100.0	96.61	89.83	100.0	25.42	49.15	88.14	67.80	76.27	49.15	25.42	84.75	49.15	52.54	25.42	93.22	74.58	98.31	74.58	54.24	94.92	91.53	57.63	50.85	79.66	83.05	67.80	61.02
FOLD 9	98.31	25.42	96.61	100.0	59.32	71.19	59.32	74.58	88.14	81.36	25.42	76.27	66.10	50.85	25.42	84.75	96.61	98.31	76.27	74.58	88.14	89.83	76.27	50.85	91.53	79.66	62.71	52.54
FOLD 10	100.0	86.44	76.27	98.31	81.36	49.15	44.07	66.10	62.71	77.97	25.42	76.27	49.15	50.85	25.42	72.88	100.0	94.92	76.27	81.36	67.80	88.14	76.27	50.85	93.22	62.71	44.07	64.41
MEAN	**91.41**	**66.99**	**91.09**	**84.40**	**41.42**	**54.57**	**77.83**	**80.44**	**76.70**	**61.11**	**38.17**	**78.90**	**56.44**	**47.99**	**25.84**	**75.51**	**87.56**	**85.44**	**73.98**	**66.72**	**89.35**	**91.41**	**73.31**	**50.00**	**86.68**	**69.13**	**59.30**	**58.63**

Overall Mean Accuracy	**69.30%**

- **CNN + BiLSTM + MHA**

Table A.18: CNN + BiLSTM + MHA - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	61.34	42.26	70.00	11.30	50.21	62.61	76.99	53.14	90.38	36.40	74.48	86.19	78.15	49.79	14.58	0.00	48.12	16.25	76.99	79.92	89.54	69.04	56.90	49.16	88.28	79.92	69.87	57.14
FOLD 2	100	82.85	95.83	90.38	76.57	70.59	92.47	93.31	97.49	85.77	87.45	84.94	91.60	59.41	67.08	62.50	94.14	76.25	91.21	93.72	97.49	81.59	92.05	91.60	87.87	88.70	51.05	79.83
FOLD 3	95.40	84.52	98.75	100	77.82	77.73	92.47	86.61	92.86	95.40	84.94	90.34	96.22	38.91	100	91.25	100	92.92	86.97	82.01	92.47	94.14	94.96	98.32	98.74	99.16	92.05	88.24
FOLD 4	98.33	94.14	100	100	92.05	50.00	87.03	94.56	76.89	98.33	89.96	96.64	92.44	49.79	100	82.50	100	100	91.60	91.21	92.47	100	73.11	88.66	100	99.16	86.61	83.19
FOLD 5	90.34	71.73	93.22	65.40	93.25	76.89	90.30	97.05	98.31	99.16	88.61	76.79	51.68	38.82	100	95.76	94.94	100	86.92	86.50	94.51	100	88.61	86.55	99.58	86.50	96.62	63.03
FOLD 6	89.08	97.05	100	60.34	81.86	77.31	92.41	69.20	98.73	97.89	77.64	90.30	63.03	51.90	100	100	100	100	96.62	87.34	91.56	92.41	75.53	74.37	100	56.96	82.70	52.94
FOLD 7	100	100	100	100	94.94	95.38	86.92	91.98	100	100	93.25	73.00	95.80	57.38	100	99.15	100	100	94.94	100	94.94	90.72	86.08	93.70	95.78	98.73	91.56	72.27
FOLD 8	100	100	83.90	100	89.45	95.38	73.42	92.41	97.89	98.73	92.83	97.05	98.74	54.01	100	100	89.03	99.58	93.25	94.94	89.45	98.73	89.45	96.22	99.16	91.14	92.41	66.81
FOLD 9	100	88.56	97.03	100	96.61	97.88	52.97	90.25	97.47	97.88	78.81	92.83	99.58	27.97	80.08	72.03	97.46	100	88.61	94.92	66.95	89.83	92.41	97.88	98.73	91.53	79.24	58.90
FOLD 10	80.93	92.80	94.07	99.15	89.83	91.10	84.32	86.02	73.00	79.66	87.71	79.32	95.76	50.00	49.15	78.81	92.80	95.34	73.84	90.25	75.42	91.95	77.22	94.07	100	96.61	63.14	50.00
MEAN	**91.54**	**85.39**	**93.28**	**82.66**	**84.26**	**79.49**	**82.93**	**85.45**	**92.30**	**88.92**	**85.57**	**86.74**	**86.30**	**47.80**	**81.09**	**78.20**	**91.65**	**88.03**	**88.10**	**90.08**	**88.48**	**90.84**	**82.63**	**87.05**	**96.81**	**88.84**	**80.53**	**67.24**

Overall Mean Accuracy	**84.72%**

- **CNN + BiGRU + MHA**

Table A.19: CNN + BiGRU + MHA - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	62.50	74.07	96.15	100	78.70	81.25	93.52	82.41	93.75	68.52	79.63	92.86	77.68	51.85	55.77	96.15	95.68	85.58	82.14	79.63	99.07	87.04	65.18	50.00	88.89	85.19	52.78	50.00
FOLD 2	100	81.48	96.15	95.37	92.59	95.54	88.89	98.15	22.52	87.96	87.96	89.19	96.43	58.33	65.38	68.27	94.44	90.38	90.99	95.37	94.44	90.74	98.20	88.39	88.89	88.89	75.00	58.93
FOLD 3	100	89.81	100	100	81.48	97.32	85.19	94.44	83.78	95.37	80.56	74.77	98.21	47.22	100	84.62	100	94.23	94.59	96.30	89.81	97.22	91.89	97.32	99.07	99.07	93.52	51.79
FOLD 4	100	97.22	100	100	93.52	97.32	74.07	96.30	82.88	100	70.37	96.40	98.21	50.93	100	93.27	100	100	85.59	94.44	76.85	99.07	76.58	91.96	100	100	96.30	64.29
FOLD 5	64.55	100	96.15	100	90.65	90.91	67.29	98.13	100	99.07	89.72	99.10	80.91	31.78	100	97.12	97.20	100	90.99	93.46	85.05	100	76.58	92.73	100	86.92	95.33	62.73
FOLD 6	84.55	98.13	100	59.81	82.24	79.09	78.50	89.72	99.10	98.13	90.65	94.59	70.00	42.06	100	100	94.39	100	93.69	98.13	89.72	100	83.78	80.91	100	55.14	58.88	60.00
FOLD 7	100	100	100	100	85.05	79.09	93.46	91.59	97.30	99.07	92.52	97.30	85.45	62.62	100	100	98.13	100	94.59	99.07	94.39	93.46	86.49	90.91	95.33	97.20	79.44	80.00
FOLD 8	100	100	94.00	100	94.29	75.45	86.67	88.57	91.82	100	95.24	82.73	97.27	71.43	100	100	99.05	100	95.45	83.81	94.29	100	96.36	94.55	99.05	92.38	97.14	95.45
FOLD 9	100	85.71	98.00	100	98.10	66.36	88.57	77.14	97.27	99.05	97.14	78.18	90.00	81.90	97.00	99.00	98.68	100	79.09	80.95	92.38	99.05	97.27	98.18	100	95.24	95.24	97.27
FOLD 10	100	97.14	100	99.05	88.57	70.91	86.67	85.71	98.18	94.29	91.43	91.82	68.18	77.14	98.00	84.00	100	98.00	78.18	93.33	95.24	95.24	77.27	99.09	98.10	100	52.38	74.55
MEAN	**91.16**	**92.36**	**98.05**	**95.42**	**88.52**	**83.32**	**84.28**	**90.22**	**86.66**	**94.14**	**87.52**	**89.69**	**86.24**	**57.53**	**91.62**	**92.24**	**97.76**	**96.82**	**88.53**	**91.45**	**91.12**	**96.18**	**84.96**	**88.40**	**96.93**	**90.00**	**79.60**	**69.50**

Overall Mean Accuracy	**88.58%**

- **Transformer + Deep BiLSTM**

Table A.20: Transformer + Deep BiLSTM - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	77.31	61.34	57.56	67.23	40.34	49.58	73.11	57.98	87.39	27.73	64.71	74.79	50.00	50.00	36.97	37.82	71.01	65.13	64.71	71.01	75.21	86.13	69.75	49.58	78.99	60.50	29.41	57.14
FOLD 2	98.74	67.65	84.87	83.61	69.75	56.72	92.02	90.34	81.09	62.61	63.87	74.79	60.50	55.04	27.73	54.62	88.66	81.93	74.79	81.93	73.95	82.35	75.63	66.81	62.18	68.49	50.00	84.03
FOLD 3	97.90	72.27	99.58	95.80	72.69	48.74	72.69	86.97	74.79	79.41	71.85	74.79	31.93	44.96	90.76	63.03	95.38	89.50	74.79	69.33	72.27	93.28	74.79	81.51	87.39	67.23	43.28	63.45
FOLD 4	100.00	83.61	97.48	97.06	76.05	49.58	88.66	91.18	74.79	59.66	80.67	74.79	55.88	52.10	88.24	74.37	95.80	89.92	74.79	82.77	81.09	95.80	83.19	80.67	88.24	66.81	48.74	77.31
FOLD 5	61.34	89.08	90.34	95.38	79.41	50.84	62.18	88.24	85.71	43.28	73.11	75.21	50.00	49.58	89.92	80.25	85.29	54.20	75.21	78.99	73.95	99.58	64.71	56.30	94.54	50.00	49.58	65.55
FOLD 6	93.28	60.50	96.64	63.45	60.08	79.41	74.79	72.69	73.11	59.66	66.81	75.21	50.00	38.66	79.41	86.13	78.99	97.06	75.21	92.02	95.80	91.18	69.33	50.00	87.82	50.00	50.00	62.61
FOLD 7	99.58	97.47	73.84	98.31	73.00	81.43	86.92	82.28	87.76	92.83	81.43	75.11	60.76	64.98	93.25	51.90	94.94	97.47	75.11	84.81	72.15	88.19	78.48	76.37	94.51	68.78	50.21	64.14
FOLD 8	100.00	91.56	74.26	98.73	76.79	70.04	68.35	75.53	75.11	92.41	78.90	75.11	45.57	49.79	90.30	78.90	75.11	49.37	75.11	89.87	86.50	96.62	75.11	73.00	90.72	50.21	53.59	57.81
FOLD 9	99.58	84.81	89.87	74.26	76.79	49.79	85.23	73.42	77.64	86.92	54.85	75.11	68.78	52.32	81.43	70.89	88.61	94.94	75.11	78.90	68.35	92.83	70.46	49.79	88.61	73.84	47.68	66.67
FOLD 10	80.17	58.23	97.89	94.09	75.11	82.70	48.52	76.79	72.57	66.24	72.15	30.38	50.21	59.92	67.51	60.76	90.30	80.59	75.11	66.24	55.70	88.61	68.35	85.23	89.45	62.45	47.26	66.67
MEAN	**90.79**	**76.65**	**86.23**	**86.79**	**70.00**	**61.88**	**75.25**	**79.54**	**79.00**	**67.08**	**70.84**	**70.53**	**52.36**	**51.73**	**74.55**	**65.87**	**86.41**	**80.01**	**73.99**	**79.59**	**75.50**	**91.46**	**72.98**	**66.93**	**86.24**	**61.83**	**46.98**	**66.54**

Overall Mean Accuracy	**73.13%**

- **CNN + MHA + BiLSTM**

Table A.21: CNN + MHA + BiLSTM - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	61.30	88.11	96.88	74.89	67.40	83.04	95.59	82.38	100.00	67.40	78.41	98.70	77.39	50.66	45.54	64.29	94.27	75.00	83.98	81.06	97.80	84.14	73.16	50.00	96.92	92.07	68.72	53.04
FOLD 2	100.00	84.14	95.98	92.51	83.26	94.35	92.07	97.36	90.48	84.58	78.85	77.92	99.57	57.71	73.66	75.00	96.92	87.05	95.24	92.51	99.56	96.04	97.84	90.00	90.31	89.43	75.77	53.91
FOLD 3	98.26	96.48	100.00	100.00	89.87	94.35	82.38	95.59	79.13	98.24	82.38	74.35	97.39	50.66	100.00	91.07	100.00	98.21	93.48	94.71	86.34	97.80	83.04	95.65	100.00	98.68	94.27	60.87
FOLD 4	100.00	97.80	100.00	100.00	95.15	95.22	77.97	99.12	85.65	96.92	72.69	96.52	97.39	52.42	100.00	95.98	100.00	100.00	85.22	96.04	87.22	99.56	81.30	95.22	100.00	100.00	96.48	50.00
FOLD 5	99.13	99.11	99.09	100.00	91.56	82.17	88.89	96.44	100.00	99.11	83.56	96.94	64.35	51.11	100.00	96.36	98.22	100.00	92.58	92.00	92.00	100.00	85.15	90.00	100.00	94.67	73.78	63.91
FOLD 6	91.30	100.00	100.00	59.56	86.22	65.65	80.89	84.00	99.13	96.44	89.33	96.51	53.91	40.00	100.00	100.00	99.11	100.00	95.20	98.22	86.22	91.56	83.84	69.13	100.00	60.89	73.78	56.96
FOLD 7	100.00	100.00	100.00	100.00	86.22	79.57	87.11	93.33	99.56	100.00	90.67	95.63	87.83	51.56	100.00	100.00	100.00	100.00	96.07	99.11	96.00	88.89	80.79	97.83	97.78	98.22	85.33	83.48
FOLD 8	100.00	100.00	96.82	100.00	93.33	87.83	84.89	92.89	92.58	100.00	95.56	85.15	98.70	69.33	100.00	100.00	95.56	100.00	90.39	90.22	93.33	99.56	90.39	98.70	98.67	95.11	91.56	94.35
FOLD 9	100.00	100.00	98.64	100.00	96.43	61.84	86.61	97.32	98.25	97.77	95.98	89.52	99.12	72.32	96.82	99.09	98.21	100.00	93.89	79.46	94.20	99.55	93.01	99.56	100.00	94.20	46.43	86.40
FOLD 10	100.00	98.21	93.18	98.21	99.55	77.63	87.50	92.86	100.00	91.96	95.09	88.65	94.74	65.62	97.27	93.18	100.00	96.82	80.79	82.59	91.96	94.20	85.59	99.56	99.11	98.21	50.89	82.46
MEAN	95.00	96.38	98.06	92.52	88.90	82.16	86.39	93.13	94.48	93.24	86.25	89.99	87.04	56.14	91.33	91.50	98.23	95.71	90.68	90.59	92.46	95.13	85.41	88.56	98.28	92.15	75.70	68.54

Overall Mean Accuracy	**89.07%**

- **CNN + MHA + BiLSTM + Ensemble EL**

Table A.22: CNN + MHA + BiLSTM + Ensemble EL - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD 1	60.00	68.72	97.32	76.65	70.93	77.39	95.59	73.13	95.67	43.17	75.77	93.07	66.96	51.10	37.50	86.16	91.63	83.04	83.98	81.06	99.56	79.74	81.82	46.96	92.51	86.34	67.84	46.52
FOLD 2	100.00	82.38	92.41	93.83	96.04	96.96	92.51	95.59	87.88	80.18	78.41	93.94	97.39	50.66	71.43	73.21	96.48	83.93	94.81	92.07	91.63	92.51	96.97	91.30	90.75	88.11	74.01	70.87
FOLD 3	100.00	93.83	100.00	100.00	89.43	92.61	88.99	92.07	75.65	98.24	80.18	76.96	97.83	51.10	100.00	80.80	100.00	95.98	92.61	93.83	89.87	94.71	88.26	92.17	100.00	96.92	93.83	70.43
FOLD 4	100.00	93.83	100.00	100.00	94.27	91.74	64.32	97.80	83.91	98.68	62.56	91.74	95.65	55.07	100.00	95.54	100.00	100.00	90.87	97.36	89.43	99.12	80.43	91.30	100.00	99.12	94.71	68.26
FOLD 5	98.70	100.00	94.55	100.00	96.00	88.26	85.33	98.22	99.56	99.56	81.33	96.07	91.74	61.78	100.00	95.91	93.78	100.00	90.39	94.67	94.22	100.00	79.48	86.96	100.00	95.56	93.33	67.39
FOLD 6	84.78	99.56	100.00	60.00	74.22	76.52	80.00	65.33	99.13	96.00	90.22	97.82	50.43	47.56	100.00	98.64	96.89	100.00	95.20	98.22	83.56	92.00	81.22	73.48	98.67	70.22	53.33	55.65
FOLD 7	100.00	100.00	100.00	100.00	76.44	74.78	90.22	93.78	97.82	100.00	96.89	95.63	90.87	64.44	100.00	100.00	100.00	100.00	97.82	96.89	95.11	90.67	76.86	96.09	96.89	95.56	81.78	81.30
FOLD 8	100.00	100.00	94.09	100.00	93.33	81.30	85.78	93.33	91.27	99.11	96.89	86.46	94.78	70.67	100.00	99.55	97.78	100.00	91.70	90.22	94.22	99.56	90.83	94.78	99.11	96.89	96.44	94.35
FOLD 9	100.00	77.68	97.73	100.00	94.64	81.1	91.52	95.54	97.82	94.20	95.98	96.07	85.96	66.07	95.91	95.91	98.21	100.00	78.60	84.82	91.52	96.43	96.51	99.12	99.11	91.07	89.29	88.16
FOLD 10	100.00	90.18	99.55	100.00	97.32	75.44	88.39	96.43	99.56	91.96	93.75	90.39	90.35	74.55	89.09	86.82	98.21	96.36	86.03	88.84	85.27	90.18	91.27	98.25	100.00	97.77	72.77	89.91
MEAN	**94.35**	**90.62**	**97.56**	**93.05**	**88.26**	**83.61**	**86.27**	**90.12**	**92.83**	**90.11**	**85.20**	**91.82**	**86.20**	**59.30**	**89.39**	**91.25**	**97.30**	**95.93**	**90.51**	**91.93**	**91.44**	**93.45**	**86.36**	**87.04**	**97.70**	**91.75**	**81.73**	**73.29**

Overall Mean Accuracy	**88.87%**

Self-Training Strategy with Hybrid Deep Learning and Intra-Subject Fine-Tuning Results

Table A.23: Intra-Subject Fine-Tuning With StandardScaler - Classification accuracy (%) per subject (S01–S28) & fold.

FOLDS	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD1	98.74	99.58	–	100.00	97.06	95.38	95.38	98.74	99.58	100.00	95.80	98.32	93.28	91.60	100.00	98.32	99.58	100.00	97.90	94.12	95.80	60.50	75.21	98.32	99.16	85.29	97.90	97.06
FOLD2	100.00	99.58	–	94.96	97.90	97.90	97.06	96.22	99.58	99.16	97.90	97.48	98.74	88.20	97.06	98.32	96.22	99.58	97.06	98.32	98.74	97.48	99.16	98.32	100.00	98.32	97.06	91.60
FOLD3	99.16	100.00	–	94.54	99.16	96.22	95.80	93.70	97.90	99.16	82.35	99.58	100.00	89.00	97.90	99.58	99.58	98.32	97.48	96.22	95.38	97.90	99.16	99.16	100.00	95.80	96.22	92.86
FOLD4	99.58	98.74	–	100.00	96.64	95.38	90.76	48.74	99.58	97.06	81.51	97.90	98.74	95.38	98.74	100.00	100.00	100.00	99.16	96.64	96.64	97.48	99.16	72.69	100.00	88.24	96.22	96.64
FOLD5	99.16	100.00	–	99.58	99.16	94.12	92.86	94.96	99.16	99.16	98.32	99.12	97.90	95.38	98.32	97.48	99.58	99.16	97.90	96.22	47.48	99.58	97.06	99.16	98.32	83.19	63.87	96.64
FOLD6	100.00	100.00	–	100.00	98.74	97.06	98.32	98.32	99.16	25.21	99.58	100.00	99.16	97.06	99.58	99.58	99.58	100.00	98.32	91.18	43.70	98.74	74.79	94.96	97.48	96.22	97.06	91.18
FOLD7	99.58	100.00	–	100.00	99.58	98.31	96.20	97.05	99.16	99.16	97.47	87.34	100.00	39.66	99.58	99.58	100.00	99.58	98.73	94.09	96.62	98.73	75.11	98.73	99.16	97.05	97.05	50.21
FOLD8	100.00	100.00	–	100.00	100.00	98.73	98.31	45.99	100.00	100.00	97.05	95.78	82.28	88.54	98.31	98.73	100.00	99.58	99.16	97.05	98.31	97.05	97.89	97.47	99.58	90.72	86.50	93.25
FOLD9	99.16	100.00	–	100.00	99.16	98.31	95.78	97.89	100.00	98.73	95.36	97.05	99.58	60.76	97.80	99.58	98.31	99.16	99.58	96.62	96.20	97.89	100.00	99.58	100.00	86.50	97.05	99.16
FOLD10	99.58	100.00	–	96.62	97.89	97.47	94.94	43.46	100.00	99.58	98.73	98.73	98.31	89.45	97.05	97.89	99.58	100.00	99.58	24.89	95.78	97.89	75.11	50.21	98.73	99.16	99.58	75.11
MEAN	**99.50**	**99.79**	–	**98.57**	**98.53**	**96.89**	**95.54**	**81.51**	**99.41**	**91.72**	**94.41**	**97.13**	**96.80**	**82.81**	**98.44**	**98.91**	**99.24**	**99.54**	**98.49**	**88.53**	**86.47**	**94.32**	**89.26**	**90.86**	**99.24**	**92.05**	**92.85**	**88.37**

Overall Mean Accuracy	**94.44%**

Table A.24: Intra-Subject Fine-Tuning With RobustScaler - Classification accuracy (%) per subject (S01–S28) & fold.

Subject	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
FOLD1	98.74	100.00	–	100.00	95.38	98.32	97.90	97.90	99.58	99.58	97.48	97.06	98.31	91.60	99.58	98.32	99.58	99.58	96.22	94.96	98.32	100.00	97.90	99.16	99.58	96.22	93.70	95.38
FOLD2	100.00	100.00	–	100.00	95.80	98.32	99.16	98.32	96.64	99.16	97.90	97.48	99.16	94.96	99.16	100.00	100.00	99.58	99.58	95.38	95.38	99.58	98.32	98.32	99.58	92.44	95.80	96.64
FOLD3	98.32	100.00	–	100.00	95.80	96.22	94.54	95.38	97.90	99.16	98.74	99.16	98.74	89.08	98.74	99.16	99.58	100.00	98.32	94.96	98.32	97.48	97.90	100.00	99.58	93.70	95.38	95.38
FOLD4	99.58	99.16	–	100.00	98.74	98.32	93.70	94.54	99.58	97.48	99.16	97.90	98.74	87.82	96.62	99.16	99.16	98.74	97.48	95.80	97.06	97.06	97.89	100.00	99.58	94.54	93.70	98.74
FOLD5	100.00	100.00	–	99.16	98.32	96.64	96.64	98.74	100.00	98.32	97.90	98.74	98.32	89.92	96.64	99.16	100.00	99.58	96.64	96.64	95.80	98.74	97.90	98.74	99.58	97.06	97.48	98.74
FOLD6	100.00	99.58	–	100.00	96.22	99.58	92.86	96.64	100.00	100.00	99.16	98.6	99.58	93.70	94.96	100.00	99.58	99.58	99.16	95.38	90.34	96.64	97.06	99.16	99.58	95.80	98.32	97.48
FOLD7	99.58	99.58	–	100.00	97.47	99.58	95.36	93.25	98.31	100.00	97.47	97.89	98.31	90.30	99.16	100.00	100.00	100.00	97.89	93.67	97.05	98.31	96.20	99.58	100.00	97.89	97.47	94.09
FOLD8	99.58	97.89	–	99.58	97.05	97.47	98.31	96.62	99.58	99.16	95.78	97.89	97.89	94.94	96.63	98.31	100.00	100.00	98.73	93.25	97.47	99.58	97.05	95.36	98.73	94.51	99.58	97.05
FOLD9	99.16	100.00	–	100.00	98.73	96.62	94.09	95.36	99.16	100.00	99.58	97.89	99.16	96.20	98.73	98.31	100.00	100.00	99.15	94.94	97.47	99.16	98.73	98.31	99.58	97.05	93.25	97.89
FOLD10	98.73	100.00	–	98.73	96.62	99.16	97.89	98.31	99.58	95.78	96.62	97.89	97.05	89.45	99.16	99.16	99.58	99.16	99.16	98.31	97.47	95.78	99.16	99.58	98.73	96.62	99.58	95.78
MEAN	**99.37**	**99.62**	–	**99.75**	**97.01**	**98.02**	**96.04**	**96.51**	**99.03**	**98.86**	**97.98**	**98.02**	**98.53**	**91.79**	**97.93**	**99.16**	**99.75**	**99.71**	**98.23**	**95.32**	**96.46**	**98.23**	**97.81**	**98.82**	**99.45**	**95.58**	**96.42**	**96.71**

Overall Mean Accuracy	**97.78%**

Subject-Independent Classification Results

Machine Learning-Based Models Results

- **KNN**

Table B.1: KNN – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	27.94	34.77	52.35	41.02	21.43	33.89	35.99	29.78	50.59	37.42	34.90	51.59	38.67	33.64	30.16	36.54	38.59	34.90	40.81	26.09	30.70	29.91	52.94	34.73	43.58	30.45	35.49	43.29	**36.86**
Time-Domain-Mix	18.25	39.60	30.62	43.20	21.48	26.47	29.74	26.80	52.60	38.25	35.57	52.01	38.46	35.28	23.95	27.14	20.13	25.38	32.76	20.81	34.69	20.55	53.23	36.33	41.86	31.88	37.67	45.09	**33.56**
Frequency-Mix	37.67	33.81	58.93	46.56	24.58	32.76	28.06	31.50	55.37	38.55	37.63	51.55	34.48	34.98	28.57	41.23	65.48	39.22	33.56	18.88	38.13	22.82	44.17	19.59	22.11	19.17	37.71	44.67	**36.49**
Wavelet-Energy	34.94	34.69	55.16	49.83	21.94	37.92	30.29	37.04	52.85	27.10	38.93	41.82	35.95	30.75	40.14	39.51	61.12	63.88	49.75	19.42	34.10	19.67	46.52	31.59	36.58	24.92	28.31	32.80	**37.76**
Elite-Melange	42.62	37.84	64.72	49.92	14.72	29.40	28.31	32.93	48.78	43.71	37.29	50.38	33.98	36.66	29.36	41.48	71.81	55.75	36.62	18.37	39.43	21.81	44.46	18.88	17.20	20.13	40.14	45.97	**37.59**
Total-AIO	35.03	37.12	59.31	46.98	23.57	29.57	28.48	31.50	57.00	40.56	38.21	50.55	33.35	35.49	33.14	44.21	68.08	48.62	40.81	19.80	43.79	22.36	46.52	21.18	32.55	20.55	36.03	44.97	**38.19**

- **LDA**

Table B.2: LDA – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	57.59	51.55	26.89	39.56	28.73	43.54	40.23	46.64	73.99	45.39	49.29	68.54	49.79	29.78	26.17	30.54	26.13	31.04	33.52	23.32	35.65	25.84	73.24	47.86	48.95	49.33	46.73	53.10	**42.96**
Time-Domain-Mix	47.02	48.70	40.73	30.24	22.78	36.70	31.29	48.53	73.41	46.27	50.08	69.76	47.82	28.06	26.64	26.38	29.57	26.43	39.01	5.03	20.05	22.73	69.42	49.92	48.28	45.93	48.24	52.27	**40.40**
Frequency-Mix	15.06	33.35	48.45	40.27	27.43	42.70	24.75	48.28	62.12	47.61	48.70	59.52	56.12	34.40	29.61	31.12	49.66	28.31	17.70	28.52	34.52	22.53	74.25	36.49	49.66	35.57	40.52	30.75	**39.21**
Wavelet-Energy	41.78	28.15	51.05	34.35	24.45	41.02	25.34	43.62	74.20	47.19	51.89	71.85	41.02	28.19	33.39	35.40	45.39	34.19	3.86	10.19	23.57	17.32	68.79	35.70	35.65	38.67	24.29	49.50	**37.85**
Elite-Melange	31.96	29.74	49.87	45.18	28.82	44.46	39.77	48.20	62.79	47.94	52.06	63.34	56.29	35.82	30.41	33.64	52.01	27.47	34.86	33.43	29.61	23.95	73.45	35.91	49.92	33.72	40.60	24.66	**41.42**
Total-AIO	36.54	34.27	58.89	38.34	25.67	44.88	41.78	50.04	60.28	46.77	43.62	36.79	46.39	35.61	35.19	36.33	56.92	33.39	29.91	35.78	36.12	25.38	72.40	37.21	48.78	29.66	32.30	40.10	**41.04**

- **SVM**

Table B.3: SVM – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	58.51	50.17	26.55	38.67	32.13	43.29	38.29	47.94	74.37	46.26	49.70	70.39	50.34	26.00	26.25	28.56	25.96	29.15	30.28	27.09	34.06	25.62	74.12	48.99	49.70	49.53	48.53	50.79	**42.90**
Time-Domain-Mix	47.27	48.02	39.97	29.61	23.82	38.08	28.31	49.33	74.20	47.35	50.08	72.10	48.74	26.17	26.51	26.34	29.19	26.34	37.24	12.54	20.93	23.90	71.69	49.87	48.95	47.69	48.82	52.85	**40.92**
Frequency-Mix	17.07	36.91	50.79	43.45	29.69	43.33	25.13	49.58	60.53	48.40	49.62	57.92	56.38	30.45	30.36	31.20	51.59	28.85	18.91	29.53	30.75	21.43	74.83	37.96	49.75	38.84	44.58	24.83	**39.73**
Wavelet-Energy	43.29	30.28	50.83	31.58	24.28	40.81	25.21	46.18	74.50	47.82	51.71	72.94	45.59	26.88	33.76	32.21	45.39	33.47	5.33	14.85	23.32	21.56	71.60	39.47	40.47	42.15	32.59	50.59	**39.23**
Elite-Melange	31.59	32.63	52.47	46.81	29.40	43.41	38.00	49.58	61.33	48.83	52.81	61.49	57.59	32.47	31.50	32.89	53.65	28.94	35.69	33.64	28.98	22.44	74.49	37.91	49.75	40.64	43.03	16.61	**41.73**
Total-AIO	40.31	38.80	59.68	37.37	26.01	40.72	43.07	51.21	55.53	47.10	39.72	29.57	44.42	31.45	36.36	37.41	52.13	32.34	30.49	38.46	31.87	24.20	74.16	37.03	48.90	33.97	30.20	28.81	**40.04**

- **RandomForestClassifier**

Table B.4: RandomForestClassifier – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	22.73	28.52	26.80	36.53	22.31	43.41	29.44	38.88	71.10	48.28	42.36	65.39	48.99	25.25	25.55	32.13	28.69	28.39	57.89	23.19	23.28	26.30	66.28	43.08	53.52	42.99	46.89	49.87	**39.21**
Time-Domain-Mix	25.96	39.97	31.29	39.76	24.07	43.20	29.87	35.94	71.90	48.44	40.48	65.85	46.81	25.29	25.67	27.77	24.66	22.73	48.36	25.16	25.92	26.00	66.94	46.26	54.15	36.82	43.25	49.37	**38.99**
Frequency-Mix	59.43	44.75	53.78	69.25	25.42	53.90	26.93	29.99	67.70	51.00	25.04	75.04	52.18	41.07	22.65	26.17	31.50	37.12	74.04	27.72	45.72	26.59	47.99	50.13	41.48	46.52	33.43	50.96	**44.19**
Wavelet-Energy	36.91	40.43	61.57	56.46	24.87	49.33	26.09	45.84	67.49	47.48	46.22	70.97	49.12	25.29	35.99	28.06	50.92	50.75	68.33	21.18	21.27	23.24	69.97	41.28	51.80	40.05	45.59	43.88	**44.29**
Elite-Melange	58.89	47.94	45.93	49.50	32.09	49.79	39.47	41.78	72.82	50.33	25.00	74.74	51.17	30.53	25.67	26.68	34.39	33.09	56.29	24.70	50.13	29.02	56.75	47.02	25.63	47.99	43.41	53.27	**43.71**
Total-AIO	58.85	35.36	54.24	62.84	23.53	47.73	34.27	46.69	73.91	50.12	48.53	74.08	44.37	30.20	26.97	29.02	43.54	43.92	65.60	19.58	48.07	28.43	43.79	51.25	51.88	46.77	27.68	54.28	**45.19**

- **XGBoost**

Table B.5: XGBoost – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	28.02	41.82	28.85	34.22	23.48	48.19	29.49	34.01	72.44	47.77	44.59	57.63	47.56	25.21	25.25	28.90	29.06	26.97	48.07	23.28	23.07	26.25	68.83	48.74	52.77	46.64	46.51	49.24	**39.53**
Time-Domain-Mix	30.95	46.43	29.40	35.27	24.58	42.44	29.73	30.53	74.16	48.19	45.51	63.67	47.69	27.89	26.71	27.94	29.40	14.84	34.23	22.86	25.75	24.41	70.39	49.87	54.27	47.86	48.86	46.14	**39.28**
Frequency-Mix	50.83	44.00	55.95	57.46	24.66	45.34	21.98	48.78	74.96	50.04	50.50	72.90	53.98	26.17	25.63	31.33	47.73	37.29	74.96	16.11	25.04	25.96	75.00	50.83	50.67	49.87	41.69	52.93	**45.80**
Wavelet-Energy	58.38	39.21	59.69	55.78	24.37	47.39	26.21	44.42	66.90	47.60	44.00	69.04	45.55	25.92	32.59	31.37	54.06	48.83	66.57	15.39	20.88	21.47	65.68	33.14	37.75	39.09	48.36	46.81	**43.44**
Elite-Melange	55.03	42.28	56.33	65.73	24.32	49.53	26.76	44.50	75.00	50.00	50.16	72.36	56.45	25.83	24.70	30.87	45.80	30.08	75.00	18.49	27.18	25.12	75.00	48.28	48.82	49.87	33.38	61.33	**46.00**
Total-AIO	65.52	38.59	54.91	55.62	19.37	46.09	37.24	40.93	74.79	50.16	48.40	72.77	48.49	25.54	28.02	30.57	54.61	40.77	74.45	20.51	30.24	24.37	75.00	45.42	53.39	49.95	36.40	54.36	**46.30**

- **ExtraTreesClassifier**

Table B.6: ExtraTreesClassifier – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	21.14	40.48	32.76	33.14	23.74	44.09	28.44	39.72	72.57	48.70	44.42	69.42	49.33	25.29	26.09	32.84	28.23	27.01	60.03	24.87	24.12	25.80	69.38	44.38	52.52	43.00	47.61	50.04	**40.32**
Time-Domain-Mix	21.81	41.61	32.59	33.35	24.45	42.45	29.15	37.21	72.23	48.95	44.04	68.54	45.81	26.89	25.50	28.52	23.95	20.01	52.60	25.04	28.98	25.84	69.97	45.13	51.85	38.55	43.88	49.83	**39.24**
Frequency-Mix	48.11	41.11	63.80	50.29	24.96	44.71	24.79	44.09	74.58	50.25	50.08	74.83	47.23	26.43	23.78	32.51	42.66	42.28	73.87	25.04	25.59	24.92	74.62	48.78	46.48	40.02	46.52	50.67	**45.10**
Wavelet-Energy	49.87	40.69	62.12	54.36	24.54	49.45	25.34	48.70	65.02	48.32	50.59	69.59	49.54	25.25	40.02	28.90	53.31	50.88	68.00	23.15	25.34	23.74	67.53	46.27	49.79	40.18	46.35	44.59	**45.40**
Elite-Melange	43.88	36.79	52.77	52.98	24.92	47.19	26.76	46.14	74.83	50.13	46.60	74.87	45.68	26.85	27.39	29.53	37.84	45.05	67.74	24.96	39.81	24.83	73.24	47.27	44.63	28.73	47.65	53.69	**44.38**
Total-AIO	42.45	36.20	55.16	50.80	24.71	48.36	25.46	45.39	74.29	50.04	46.90	74.54	48.36	25.55	28.36	31.21	39.51	47.06	67.41	25.34	26.89	25.42	74.37	48.99	51.72	31.63	43.54	49.20	**44.24**

- **LightGBM**

Table B.7: LightGBM – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
Non-Linear-Mix	24.66	40.14	29.66	36.79	22.44	47.94	30.16	31.63	71.10	47.32	42.49	57.76	46.85	25.92	24.54	28.40	29.28	26.85	52.68	24.41	23.53	26.51	69.63	49.08	53.52	45.30	45.64	48.95	**39.39**
Time-Domain-Mix	35.49	41.40	30.83	36.37	24.58	40.98	31.21	30.29	73.53	48.49	45.51	62.12	44.80	26.38	25.34	27.10	30.37	14.56	31.04	22.02	25.17	21.98	70.22	50.17	52.94	45.72	46.69	47.65	**38.67**
Frequency-Mix	56.12	42.41	50.92	53.48	24.37	36.45	30.62	47.40	74.50	50.04	50.80	68.67	50.88	30.08	22.82	30.20	58.47	36.20	75.00	17.53	28.86	25.00	75.00	50.25	50.00	50.00	27.47	52.47	**45.21**
Wavelet-Energy	62.33	33.89	59.77	55.58	24.29	46.56	28.73	47.65	63.05	46.39	41.61	60.19	44.84	26.93	35.78	35.36	53.10	44.84	67.37	14.77	21.10	21.02	62.46	33.01	35.65	38.59	46.81	45.81	**42.76**
Elite-Melange	64.64	44.09	57.76	73.78	13.17	47.99	26.76	33.98	74.12	50.00	50.04	67.45	53.78	26.55	27.27	29.15	46.31	26.64	74.66	17.95	38.09	25.46	75.00	47.82	50.00	50.00	39.09	61.33	**46.17**
Total-AIO	67.20	39.43	63.63	55.33	17.58	45.81	33.31	37.92	73.41	50.08	48.57	73.53	50.17	25.76	30.45	27.27	53.19	33.10	73.66	14.09	40.10	24.08	75.00	42.16	50.00	50.00	26.72	57.09	**45.66**

Ensemble Learning-Based Models Results

- **Stacking**

Table B.8: Stacking – Classification accuracy (%) per subject (S01–S28).

Blend	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	Mean
Non-Linear-Mix	72.78	66.15	68.25	77.94	55.29	66.48	58.93	55.99	75.00	58.14	52.98	74.75	61.66	52.68	42.28	52.01	78.15	55.03	75.00	61.75	57.63	53.06	75.00	66.36	59.02	61.49	52.64	62.84	**62.47**
Time-Domain-Mix	74.54	63.34	65.10	73.32	56.17	66.57	57.89	54.91	75.67	59.69	56.25	75.80	63.21	54.61	42.87	45.09	70.51	52.98	75.13	66.19	56.75	57.93	75.00	66.95	64.05	63.17	54.36	66.02	**62.65**
Frequency-Mix	79.91	81.80	81.17	95.81	55.45	63.05	60.23	60.15	77.06	61.91	57.63	76.13	64.35	54.78	56.92	71.85	93.04	86.20	75.08	59.82	59.48	60.28	75.00	67.91	68.46	70.81	57.05	67.62	**69.25**
Wavelet-Energy	80.29	82.76	80.08	93.79	57.01	64.93	63.63	57.89	78.94	65.35	54.32	75.71	69.04	53.02	62.79	66.69	91.28	87.50	74.71	58.35	54.32	54.95	74.96	62.08	64.01	67.45	58.31	64.72	**68.53**
Elite-Melange	77.52	79.91	75.25	93.33	55.66	62.12	60.70	54.24	76.09	60.78	58.43	77.27	63.00	55.45	54.61	62.96	91.86	84.52	75.59	61.16	65.18	59.44	75.00	67.91	64.81	67.32	54.87	66.28	**67.90**
Total-AIO	79.91	81.80	81.17	95.81	55.45	63.05	60.23	60.15	77.06	61.91	57.63	76.13	64.35	54.78	56.92	71.85	93.04	86.20	75.08	59.82	59.48	60.28	75.00	67.91	68.46	70.81	57.05	67.62	**69.25**

- **Fusion vs Voting**

Table B.9: Fusion vs Voting – Accuracy (%) per Subject.

Method	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	Mean (%)
Fusion	89.39	91.78	94.51	99.83	74.50	76.09	76.76	73.66	85.36	80.54	73.57	84.73	77.64	63.00	81.46	85.03	97.90	98.03	83.26	78.06	79.74	75.38	80.70	79.40	82.30	80.49	71.98	72.53	**81.70**
Voting	80.54	83.60	85.11	95.05	55.66	68.29	63.38	60.23	76.55	65.86	56.63	75.80	66.32	55.24	66.82	71.14	93.41	89.85	75.00	66.65	59.31	57.17	75.00	71.81	71.10	70.89	56.63	69.04	**70.79**

Hybrid Deep Learning & Intra-Subject Fine-Tuning

Table B.10: Hybrid Deep Learning & Intra-Subject Fine-Tuning – Classification accuracy (%) per subject (S01–S28).

SUBJECT	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	MEAN
EEGNet +Attention	49.00	75.33	49.00	47.00	49.67	24.33	33.33	52.00	38.66	69.33	38.00	27.33	22.33	19.67	44.33	46.33	47.00	73.00	50.66	18.33	26.67	53.00	18.33	22.67	24.00	30.67	22.67	47.66	**39.83**
EEGNet +Fine-Tuning	46.72	45.87	72.90	46.30	25.00	41.67	57.74	40.15	60.02	45.79	44.86	20.53	23.06	26.18	45.53	46.21	72.81	47.05	35.10	25.59	23.73	22.22	47.47	21.38	26.51	21.04	17.08	44.52	**39.04**
EEGNet +Attention +Fine-Tuning	45.92	44.15	71.67	47.73	25.00	49.24	45.03	48.11	64.69	35.52	41.92	21.67	20.24	31.06	45.41	46.67	73.32	49.87	30.51	18.30	27.48	24.20	42.42	26.39	31.78	16.28	35.60	21.84	**38.65**
DAN	44.88	50.17	33.98	33.98	24.83	33.14	65.44	50.76	73.74	46.81	30.70	62.00	23.15	24.50	31.21	26.17	25.25	26.85	43.62	24.08	46.81	25.59	70.72	44.55	49.24	38.67	21.14	49.16	**40.04**
DAN +Augmentation	34.06	50.00	35.82	49.58	25.00	25.76	25.00	42.28	46.64	23.66	22.06	05.03	20.13	41.69	33.64	31.63	37.67	35.32	24.83	21.14	41.02	22.73	32.89	20.13	45.72	22.06	22.23	42.28	**31.42**
DAMGCN +Attention	25.63	37.08	33.42	40.91	16.58	46.84	35.02	45.50	69.40	49.83	49.79	72.31	39.39	25.21	28.62	28.62	28.24	15.28	70.79	22.81	36.07	20.33	66.84	49.96	50.04	48.19	34.93	48.19	**40.56**
LABraM	49.33	31.67	42.33	42.00	28.67	49.67	25.00	50.00	74.33	49.33	48.67	74.00	48.00	24.67	30.33	33.33	37.00	31.67	57.67	23.00	25.33	25.00	72.67	49.33	52.00	50.00	49.00	49.00	**43.68**
CNN+BiLSTM +MHA	46.46	34.76	68.43	46.88	25.33	51.01	42.00	49.15	71.38	46.71	50.42	53.19	49.07	26.01	40.40	46.88	70.20	46.29	45.53	25.75	08.41	32.23	37.54	22.81	35.52	14.81	27.35	31.06	**40.91**
CNN+BiGRU +MHA	61.63	49.49	73.65	60.88	25.67	41.25	67.17	49.83	68.43	39.23	45.29	70.33	32.80	38.38	46.38	48.65	73.48	53.45	33.92	24.16	29.71	24.83	61.72	49.32	51.09	15.15	32.09	42.76	**46.81**
CNN +MHA +BiLSTM	74.96	51.18	74.87	71.30	26.30	50.59	70.50	45.41	73.53	50.29	57.53	72.77	44.49	28.03	44.61	46.09	86.03	52.95	42.42	27.40	36.57	29.25	72.22	51.94	50.51	23.86	34.18	53.66	**51.54**
Intra-Fine-Tuning	-	52.36	-	-	50.42	61.36	-	63.38	-	69.11	58.92	-	44.32	49.54	76.77	49.24	-	82.95	56.48	83.00	46.34	62.92	-	55.88	80.47	58.96	52.06	51.73	**60.31**

Detailed Subject Annotations and Demographics

Table C.1: Subject demographics and self-assessment manikin (SAM) emotional annotations across the four game sessions (G1–G4).

Subjects	Annotation/SAM	G1	G2	G3	G4	Gender	age
Subject 1	Annotation	LANV	LANV	HANV	HANV	Female	24
	SAM	A:3, V:2	A:4, V:5	A:9, V:2	A:8, V:5		
Subject 2	Annotation	LANV	LANV	HANV	LAPV	Male	23
	SAM	A:1, V:1	A:5, V:5	A:9, V=1	A:5, V:7		
Subject 3	Annotation	LANV	LAPV	HANV	HAPV	Male	25
	SAM	A:2, V:1	A:5, V:7	A:9, V:2	A:6, V:9		
Subject 4	Annotation	LANV	HAPV	HANV	HAPV	Male	20
	SAM	A:4, V:2	A:7, V:8	A:9, V:3	A:7, V:6		
Subject 5	Annotation	LANV	HAPV	HANV	HAPV	Female	25
	SAM	A:3, V:3	A:9, V:9	A:9, V:1	A:7, V:9		
Subject 6	Annotation	LANV	HAPV	LANV	HAPV	Female	21
	SAM	A:3, V:2	A:9, V:7	A:5, V:4	A:7, V:7		
Subject 7	Annotation	LANV	HAPV	HANV	HAPV	Male	22
	SAM	A:1, V:3	A:6, V:7	A:7, V:5	A:9, V:9		
Subject 8	Annotation	LANV	HAPV	LANV	HANV	Male	26
	SAM	A:1, V:2	A:7, V:9	A:1, V:1	A:6, V:5		
Subject 9	Annotation	LANV	LANV	HANV	LANV	Male	23
	SAM	A:5, V:5	A:1, V:5	A:7, V:3	A:1, V:5		
Subject 10	Annotation	LANV	LANV	HANV	LAPV	Male	20
	SAM	A:5, V:5	A:5, V:3	A:7, V:3	A:5, V:9		

Table C.1: Subject demographics and self-assessment manikin (SAM) emotional annotations across the four game sessions (G1–G4).

Subjects	Annotation/SAM	G1	G2	G3	G4	Gender	age
Subject 11	Annotation	LANV	LAPV	LANV	HAPV	Male	26
	SAM	A:1, V:5	A:3, V:7	A:4, V:3	A:7, V:7		
Subject 12	Annotation	LANV	LAPV	LANV	LANV	Female	25
	SAM	A:3, V:5	A:5, V:7	A:5, V:5	A:5, V:5		
Subject 13	Annotation	LAPV	LANV	LANV	LAPV	Male	20
	SAM	A:1, V:7	A:3, V:3	A:4, V:3	A:2, V:9		
Subject 14	Annotation	LANV	LAPV	HANV	LAPV	Male	25
	SAM	A:1, V:3	A:5, V:7	A:7, V:4	A:5, V:9		
Subject 15	Annotation	LANV	LAPV	HANV	HAPV	Male	22
	SAM	A:3, V:4	A:3, V:6	A:7, V:2	A:8, V:7		
Subject 16	Annotation	LANV	LAPV	HANV	HAPV	Male	20
	SAM	A:2, V:3	A:2, V:8	A:6, V:1	A:7, V:6		
Subject 17	Annotation	LANV	LAPV	HANV	HANV	Female	26
	SAM	A:1, V:4	A:4, V:7	A:6, V:3	A:7, V:5		
Subject 18	Annotation	LAPV	LANV	HANV	HAPV	Female	26
	SAM	A:2, V:7	A:5, V:5	A:6, V:4	A:6, V:8		
Subject 19	Annotation	LANV	LANV	LANV	LAPV	Female	23
	SAM	A:2, V:5	A:4, V:3	A:5, V:4	A:4, V:8		
Subject 20	Annotation	LANV	LAPV	HANV	LAPV	Female	22
	SAM	A:3, V:2	A:3, V:6	A:8, V:2	A:5, V:7		

Table C.1: Subject demographics and self-assessment manikin (SAM) emotional annotations across the four game sessions (G1–G4).

Subjects	Annotation/SAM	G1	G2	G3	G4	Gender	age
Subject 21	Annotation	LANV	HAPV	HANV	HAPV	Female	20
	SAM	A:4, V:1	A:6, V:7	A:9, V:1	A:8, V:6		
Subject 22	Annotation	LANV	LAPV	HANV	HANV	Male	23
	SAM	A:4, V:3	A:4, V:9	A:6, V:3	A:9, V:5		
Subject 23	Annotation	LANV	LANV	LANV	HAPV	Male	21
	SAM	A:3, V:2	A:2, V:4	A:4, V:2	A:7, V:8		
Subject 24	Annotation	LANV	HAPV	LANV	HAPV	Male	23
	SAM	A:2, V:4	A:6, V:6	A:5, V:4	A:6, V:7		
Subject 25	Annotation	LANV	LANV	HANV	HAPV	Male	27
	SAM	A:1, V:4	A:3, V:5	A:6, V:4	A:6, V:9		
Subject 26	Annotation	LAPV	LANV	HANV	LANV	Male	23
	SAM	A:4, V:6	A:3, V:3	A:7, V:2	A:5, V:4		
Subject 27	Annotation	LANV	LAPV	HANV	LANV	Male	22
	SAM	A:3, V:2	A:1, V:7	A:7, V:1	A:4, V:5		
Subject 28	Annotation	LANV	LAPV	LANV	LAPV	Male	24
	SAM	A:4, V:1	A:2, V:8	A:4, V:1	A:3, V:7		