DEVELOPMENT OF OPTIMAL ALGORITHMS FOR EMOTION RECOGNITION

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

DOCTOR OF PHILOSOPHY

in
Electronics and Communication Engineering

Submitted by
Amit Kumar Dwivedi
Roll No. 2K20/PHDEC/512

Under the Supervision of
Prof. O. P. Verma
Professor

Co-supervision of
Dr. Sachin Taran
Assistant Professor

Department of Electronics and Communication Engineering
Delhi Technological University

Department of Electronics and Communication Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India

December, 2025

Acknowledgement

I want to express my gratitude to Prof. O. P. Verma and Dr Sachin Taran, my supervisors, for their encouragement, support, and direction. I want to express my gratitude to HOD Prof. Neeta Pandey madam for her assistance. I want to express my gratitude to my father, Rajender Kumar Dwivedi, and mother, Pushpa Dwivedi, for their unwavering encouragement and support. I would like to thank my brothers Ashish and Alok, sisters Savita and Pratibha, and sisters-in-law Shalini and Shikha for their support and encouragement.

I want to express my gratitude to my wife for being patient with me and inspiring me to conduct research. My children, Avyaan and Arshini, as well as my nephews and nieces, Ishan, Pranjal, Trisha, Deksha, and Vivaan, all show me love and affection.

Finally, I want to express my gratitude to my friends Deokinandan, Pranjit, Suresh, Yashika, Arjun, Vikram, Ankit, and Rohit for their helpful advice, encouragement, and support.

Amit Kumar Dwivedi
Roll No. 2K20/PHDEC/512

DELHI TECHNOLOGICAL UNIVERSITY

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

CANDIDATE’S DECLARATION

I, **Amit Kumar Dwivedi**, hereby certify that the work which is being presented in the thesis entitled **“Development of Optimal Algorithms for Emotion Recognition”**, in partial fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy**, submitted in the Department of Electronics and Communication Engineering, Delhi Technological University, is an authentic record of my own work carried out during the period from **January 2021 to December 2025** under the supervision of **Prof. O. P. Verma** and **Dr. Sachin Taran**.

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

Candidate’s Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor(s) **Signature of External Examiner**

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

CERTIFICATE BY THE SUPERVISOR(s)

Certified that **Amit Kumar Dwivedi (2K20/PHDEC/512)** has carried out the research work presented in this thesis entitled **"Development of Optimal Algorithms for Emotion Recognition"** for the award of **Doctor of Philosophy** from the **Department of Electronics and Communication**, Delhi Technological University, Delhi, under our supervision. The thesis embodies the results of the original work, and the student himself carried out studies. The contents of the thesis do not form the basis for the award of any other degree to the candidate or anybody else from this or any other university.

Signature

Prof. O.P.Verma

Professor

Department of ECE DTU

Date:

Signature

Dr. Sachin Taran

Assistant Professor

Department of ECE DTU

Date:

ABSTRACT

Emotion recognition from electroencephalography (EEG) is vital to advancing human–computer interaction, affective computing, and mental health assessment, yet its effectiveness is impeded by high dimensionality, noise contamination, and the nonlinear dynamics of neural signals. This thesis addresses these challenges through integrated research objectives: optimal EEG channel selection, optimal filter design for noise elimination, an optimal mode filtering framework, optimal wavelet design using nature-inspired algorithms, and optimally tuned deep learning models for EEG-based emotion recognition.

The first objective is to develop a correlation-based channel selection method that identifies a compact set of electrodes most relevant to emotion recognition. This approach reduces redundancy while preserving discriminative neural activity. The selected channels undergo processing using volt–Kalman sub-band filtering, from which statistical features are extracted in the sub-bands. These features are then refined using the MRMR (Minimum Redundancy Maximum Relevance) algorithm, and the resulting features are input into a classification algorithm, which yields an accuracy of 88%.

Building upon this framework, the second objective focuses on designing a COA-optimized IIR (Infinite Impulse Response) filter to achieve accurate rhythm extraction, thereby isolating clean δ, θ, α, β, and γ rhythms for subsequent feature analysis. From these refined rhythms, Hjorth parameters and entropy-based features are computed, which, then classified using a cubic and quadratic SVM (Support Vector Machine), deliver robust performance with accuracies of 94.4% and 93.0%, respectively.

The second objective complements the framework by introducing a PSO-based adaptive IIR filtering approach for the extraction of clean δ, θ, α, β, and γ rhythms from EEG signals. The filter parameters are adaptively selected using Particle Swarm Optimization (PSO), where the optimization criterion is defined as the minimization of the mean squared error (MSE) between the original and reconstructed rhythms. Subsequently, the extracted rhythms are converted into time–frequency images using adaptive superlets (ASLT). Building upon these representations, a hybrid CNN (Convolutional Neural Network) and ResNet architecture (HCRNet) is proposed. When trained on ASLT-derived time–frequency images, the proposed HCRNet achieves a classification accuracy of 93%, thereby validating its effectiveness in advancing temporal-spectral modeling for emotion recognition.

To further improve EEG quality, the thesis proposes an optimal mode filtering framework comprising two complementary strategies. In the first, Variational Mode Decomposition (VMD) is used to separate EEG signals into intrinsic modes, with Pearson's correlation coefficient identifying modes that preserve meaningful neural dynamics. Further, the clean EEG is decomposed into the canonical δ, θ, α, β, and γ rhythms using the Wavelet Packet Transform (WPT), thereby ensuring precise rhythm isolation. Joint time–frequency scattering applied to the filtered rhythms produces robust feature representations that yield high clustering separability, revealing the dominance of low-frequency rhythms in negative-emotion states and high-frequency rhythms in high-arousal emotions. JTFS-derived features are then evaluated across multiple classification algorithms, where δ and θ rhythms consistently outperform higher-frequency bands. Notably, the combination of δ and θ rhythms achieves the highest classification accuracy of 80.39% on Dataset 1 and 99.35% on Dataset 2, thereby showing the effectiveness of the proposed framework in emotion recognition. The second strategy employs Group Sparse Mode Decomposition (GSMD) with Bhattacharyya Distance to suppress modes that lack discriminative power. The refined signals are transformed using Superlet and Adaptive Superlet Transforms, and a dedicated super-resolution neural network (SRNET) designed for fine-grained time–frequency patterns achieves 99.63% accuracy while outperforming standard deep architectures.

Complementing these contributions, the thesis develops nature-inspired optimized wavelet frameworks to enhance time–frequency feature extraction. An adaptive flexible analytic wavelet transform (AFAWT) optimized using Particle Swarm Optimization automatically tunes wavelet parameters to EEG characteristics, achieving 90.3% accuracy on video game emotion datasets. Similarly, a crayfish-optimized tunable-Q Wavelet Transform (TQWT) adapts decomposition resolution for emotion-specific structure, yielding state-of-the-art performance on SEED-IV (91.47%) and DEAP (77.81%). A final extension introduces a PSO-optimized Wavelet Scattering Transform, improving scattering feature quality and achieving accuracies of 97.6% and 99.4% across two benchmark datasets.

The thesis culminates in the fourth objective, which focuses on developing optimally tuned deep learning models for emotion recognition. EEG signals are first converted into time–frequency images (spectrogram, scalogram, and SPWVD), which are then processed by an optimized CNN/GoogleNet architecture. Given the sensitivity of deep learning performance to hyperparameter settings, this work incorporates Bayesian Optimization (BO) to automatically identify the optimal hyperparameters. BO efficiently explores the parameter space, achieving faster convergence and significantly higher accuracy compared to conventional manually tuned models, thereby validating the robustness of the proposed framework. The optimized GoogleNet-derived features are classified using an SVM, achieving robust performance with accuracies exceeding 82% and AUC values above 0.95.

Overall, the optimized filtering, decomposition, wavelet design, and BO-enhanced deep learning strategies collectively yield robust and high-accuracy emotion recognition. The proposed framework demonstrates strong potential for deployment in brain–computer interfaces, affective computing, adaptive gaming environments, and mental health monitoring systems.

List of Publications

Published Journals

1. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "Adaptive Flexible Analytic Wavelet Transform for EEG-Based Emotion Recognition," *IEEE Sensors Journal*, vol. 24, no. 18, pp. 28941–28951, Sep. 2024, doi:10.1109/JSEN.2024.3429523.
2. Amit Kumar Dwivedi, Om Prakash Verma, and S. Taran, "Joint Time–Frequency Wavelet Scattering Transform–Based Framework for Emotion Recognition Enhancement," *Signal, Image and Video Processing*, Sep. 2025, doi:10.1007/s11760-025-04656-w.
3. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "Group Sparse and Super-Resolution Time–Frequency-Based Method for Emotion Recognition," *Digital Signal Processing*, Elsevier, Nov. 2025, doi:10.1016/j.dsp.2025.105761.
4. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "Decoding Human Feelings: A Dynamic Fusion of Adaptive IIR Filtering and HCRNet for Emotion Recognition," *Signal, Image and Video Processing*, Springer, Jan. 2026, doi: 10.1007/s11760-026-05142-7.
5. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "Crayfish Optimized Wavelet for Enhancing Emotion Recognition using EEG signals," *The Journal of Supercomputing*, Springer, May 2026, doi:10.1007/s11227-026-08612-4.

Conference Proceedings

1. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "EEG-Based Emotion Recognition Using Optimized Deep-Learning Techniques," in *Proc. 11th Int. Conf. on Signal Processing and Integrated Networks (SPIN)*, Noida, India: IEEE, Mar. 2024, pp. 372–377, doi:10.1109/SPIN60856.2024.10512074.
2. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "EEG-Based Emotion Recognition Using Optimum Number of Channels," in *Proc. IEEE Int. Conf. on Smart Power Control and Renewable Energy (ICSPCRE)*, Rourkela, India: IEEE, Jul. 2024, pp. 1–6, doi:10.1109/ICSPCRE62303.2024.10675012.
3. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, "EEG-Based Mental States Prediction of Driver," in *Proc. IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Slovakia, Jan. 2025, pp. 429–434,

doi:10.1109/SAMI63904.2025.10883297.

4. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, ”Emotion Recognition Using Electroencephalogram Signals,” in *IEEE DELCON - International Conference on Recent Smart Technologies in Engineering for Sustainable Development, New Delhi,* India, 2025, pp. 1-6, doi: 10.1109/DELCON68055.2025.11400205.

5. Amit Kumar Dwivedi, Sachin Taran, and Om Prakash Verma, “Optimized Filter Design for EEG-Based Emotion Recognition,” in *Lecture Notes in Electrical Engineering*, vol. 1449, *Advancements in Signal, Image and Video Processing*, Chapter 18.

Communicated

1. Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran, “ Optimized Wavelet Scattering Transform for Robust EEG-based Emotion Recognition” *Biomedical Signal Processing and Control*, Elsevier.

Contents

6 Optimal Wavelet Design using Nature-Inspired Algorithms 89

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

AI	Artificial Intelligence
EEG	Electroencephalogram
ML	Machine Learning
DL	Deep Learning
PSO	Particle Swarm Optimization
QDA	Quadratic Discriminant Analysis
JTFS	Joint Time–Frequency Scattering
GSMD	Group Sparse Mode Decomposition
SLT	Superlet Transform
ASLT	Adaptive Superlet Transform
VAD	Valence–Arousal–Dominance
PANAS	Positive and Negative Affect Schedule
SAM	Self-Assessment Manikin
CMQ	Current Mood Questionnaire
HR	Heart Rate
HRV	Heart Rate Variability
GSR	Galvanic Skin Response
EMG	Electromyography
ECG	Electrocardiogram
FACS	Facial Action Coding System
MEG	Magneto-encephalography
fMRI	Functional Magnetic Resonance Imaging
CNNs	Convolutional Neural Networks
LSTMs	Long Short-Term Memory Networks
HCI	Human–Computer Interface
IIR filter	Infinite Impulse Response filter
FAWT	Flexible Analytic Wavelet Transform
COA	Crayfish Optimization Algorithm
TQWT	Tunable Q Wavelet Transform
VMD	Variational Mode Decomposition
WST	Wavelet Scattering Transform
EOG	Electrooculography
DEAP	A Database for Emotion Analysis using Physiological Signals
DREAMER	A Database for Emotion Recognition Through EEG and ECG Signals

	From Wireless Low-cost Off-the-Shelf Devices
LUMPED-2	Loughborough University Multi-modal Emotion Database-2
ICA	Independent Component Analysis
MSE	Mean Squared Error
PCC	Pearson Correlation Coefficient
BCIs	Brain-Computer Interfaces
PSD	Power Spectral Density
FFT	Fast Fourier Transform
STFT	Short-Time Fourier Transform
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
WPT	Wavelet Packet Transform
EWT	Empirical Wavelet Transform
GA	Genetic Algorithms
GWO	Grey Wolf Optimization
SVM	Support Vector Machine
k-NN	k-Nearest Neighbors
ANNs	Artificial Neural Networks
MLP	Multilayer Perceptron
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
RMSProp	Root Mean Square Propagation
BO	Bayesian Optimization
DBNs	Deep Belief Networks
TFR	Time-Frequency Representation
SPWVD	Smoothed Pseudo Wigner Ville Distribution
ELM	Extreme Learning Machine
IMFs	Intrinsic Mode Functions
WNN	Wide Neural Networks
PCA	Principal Component Analysis
SNR	Signal-to-Noise Ratio
FPR	False Positive Rate
TPR	True Positive Rate
VGG	Visual Geometry Group (a type of CNN architecture)
HHT	Hilbert Huang Transform
LDA	Linear Discriminant Analysis
RF	Random Forest
CV	Cross Validation
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
ROC	Receiver Operating Characteristic
PRE	Precision

RC	Recall
MCC	Matthews Correlation Coefficient
KP	Cohen’s kappa Coefficient

Chapter 1

Introduction

Emotions are complex, multidimensional phenomena that include physiological arousal, subjective experiences, and expressive behaviors in reaction to either internal or external stimuli. They are essential to human functioning and have an impact on perception, social relationships, cognition, decision-making, and survival strategies. Emotion recognition is a significant area of research because it enhances our understanding of human behavior, cognition, and decision-making processes. It facilitates natural and adaptive interactions between humans and machines. This improves applications in human-computer interfaces, virtual assistants, and affective computing. In healthcare and medicine, emotion recognition aids in the early diagnosis of mental health issues, including anxiety, depression, and stress. Emotional analysis is important in other areas such as education, entertainment systems, and personalized learning, as it allows for adaptation to users' emotional states. Additionally, emotional awareness is useful for sectors including security, marketing, and autonomous systems.

1.1 Emotion Recognition Models

Emotion recognition is based on various psychological models that elucidate the origins of emotions, their classification, and the methods by which they can be interpreted. The primary models that elucidate the process of emotion recognition are outlined as follows.

(a) Circumplex Model: The Circumplex Model [1] represents emotions along two dimensions—arousal and valence (Figure 1.1). Within this framework, a circular emotional space captures the full range of emotions, where the clustering of similar emotions allows for smooth transitions between different emotional states.

(b) PAD Model – Mehrabian and Russell: PAD Model – Mehrabian and Russell model [3] defines emotions in 3 dimensions naming dominance, arousal, and valence. These aspects clarify the subtle differences among similar emotions, which are valuable for modeling affect in both consumer and environmental psychology. The Valence–Arousal–Dominance (VAD) model is the most widely utilized among these models. Since changes in brain signals occur continuously rather than in isolation during EEG-based studies, this model is well-suited for offering a comprehensive description of emotional experiences.

Valence is defined as the degree of pleasantness or unpleasantness of an emotion, ranging

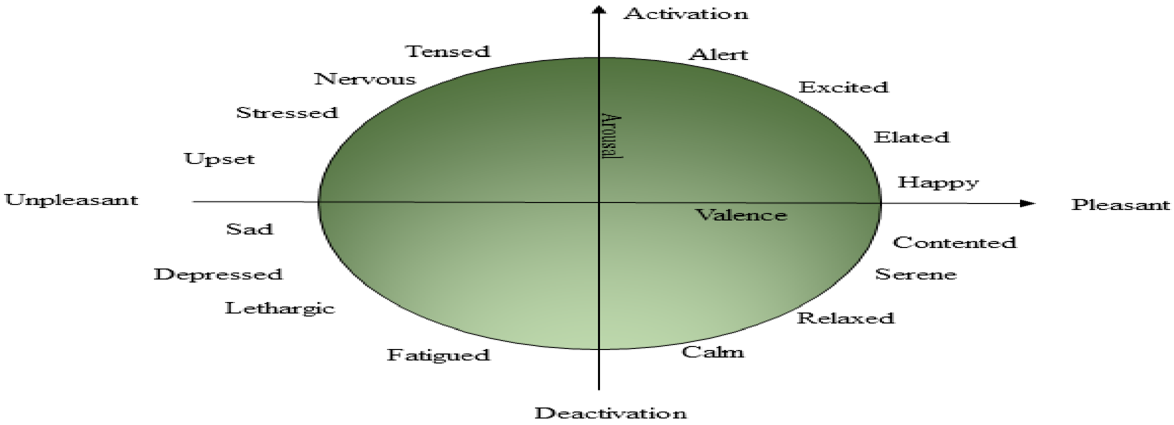


Figure 1.1: The dimensions of valence and arousal constitute the 2D model of emotion [2].

from negative (e.g., sadness) to positive (e.g., joy).

Arousal is defined as the range of emotional activation intensities, from low (such as calm) to high (such as excitement).

The degree of power or influence one feels in a situation is reflected in dominance, which can range from submissive (such as fear) to dominating (such as fury).

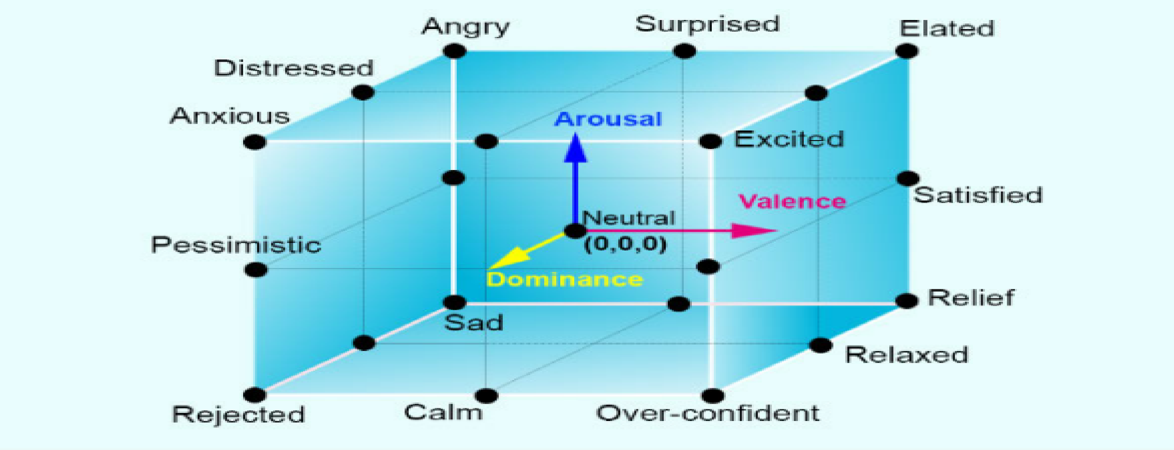


Figure 1.2: Valence, arousal, and dominance are the three dimensions that constitute the 3D model of emotion [2].

The two types of emotion models are differentiated by their dimensionality: the 2D model, which utilizes valence and arousal dimensions shown in Figure 1.1, and the 3D model, which adds a dominance dimension shown in Figure 1.2.

(c) Paul Ekman Model of Discrete Emotions: Paul Ekman Model [4] models classify emotions into a limited set of basic or prototypical emotional states. These emotions are regarded as universal and evolutionarily conserved across cultures. Each emotion corresponds to specific physiological responses and unique facial expressions. The most notable of these is Ekman's six basic emotions, which are shown in Figure 1.3.

---

[1] Source: `https://www.twinkl.co.in/resource/emoji-emotion-cards-au-t-10003409`.

(a) Happiness (b) Sadness (c) Anger

(d) Fear (e) Surprise (f) Disgust

Figure 1.3: Emotional expressions: happiness, sadness, anger, fear, surprise, and disgust.[1]

(d) James-Lange Theory of Emotions: The James-Lange Theory [5], suggests that emotions arise from our interpretation of physiological responses to stimuli. This theory challenged previous beliefs by suggesting that bodily changes happen before the awareness of emotions. Despite criticism, it shaped modern understanding of how physical states influence emotions.

(e) Cannon-Bard Theory: The Cannon-Bard Theory [6] of emotion posits that emotional experiences and physiological responses occur simultaneously and independently when triggered by a stimulus. It highlights the role of the thalamus in sending parallel signals to the brain and body. Unlike James-Lange, it argues emotion does not depend on perceiving bodily changes. This theory helped pave the way for central nervous system models in affective neuroscience.

(f) Schachter-Singer Two-Factor Theory: The Schachter-Singer [7] Two-Factor Theory defines emotion as a combination of physiological arousal and cognitive interpretation. Emotional labeling depends on contextual evaluation of the arousal state. This theory underscores the cognitive shaping of emotion and informs context-aware affective systems.

1.2 Modalities for Emotion Assessment

Measuring emotions is challenging. Emotions are very personal and can vary from person to person, but they often show up in behaviors and physical changes. A variety of methods are used for capturing these complex experiences.

1.2.1 Physiological Measurements

The objective of physiological measurements is to record physical reactions that are highly related to emotional states. These methods are a useful addition to subjective assessment measures since they provide unbiased information based on measurable biological signals rather than subjective emotions.

(a) Heart Rate and Heart Rate Variability: Heart Rate (HR) and Heart Rate Variability (HRV) are sensitive indicators of emotional arousal [8]. For instance, an increase in HR or a decrease in HRV may indicate heightened levels of anxiety, fear, or excitement. Conversely, higher HRV is generally associated with calmness and positive emotional states.

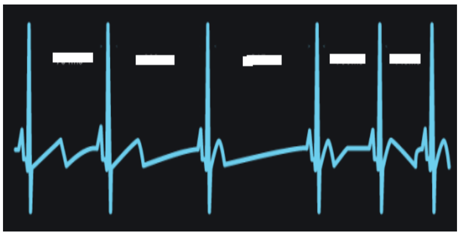


Figure 1.4: Heart Rate Variability[2]

(b) Galvanic Skin Response: Galvanic Skin Response (GSR), also known as electrodermal activity [9], measures changes in sweat gland activity, which typically increase during emotional arousal or stress due to activation of the sympathetic nervous system.

Figure 1.5: Galvanic Skin Response [10]

(c) Facial Electromyography: Facial electromyography detects subtle electrical activity in facial muscles [9], like the zygomatic (associated with smiling) and corrugator (associated with frowning) muscles. This capability enables precise identification of emotional expressions, even when they are not visible.

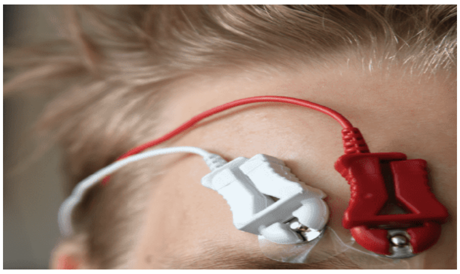


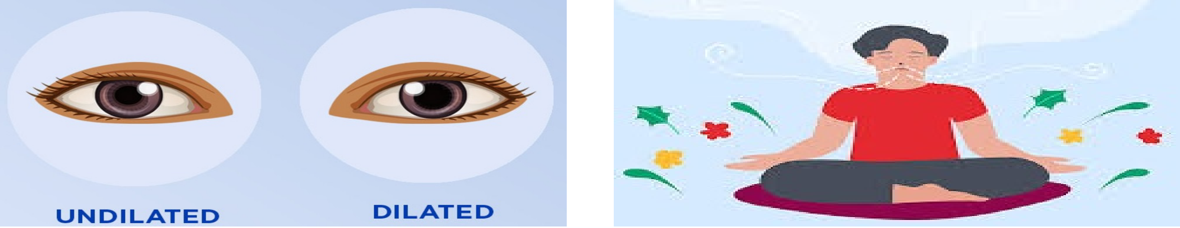
Figure 1.6: Facial Electromyography[3]

(d) Pupil Dilation and Respiration Rate: Pupil size generally expands in response to emotional [11] intensity or cognitive load, whereas breathing rate can fluctuate significantly depending on emotional states, such as fear, stress, or relaxation.

---

[2] Source: `https://ouraring.com/blog/what-is-heart-rate-variability/`

[3] Source:`https://imotions.com/blog/learning/research-fundamentals/electromyography-101/`



(a) Pupil Dilation

(b) Respiration Rate

Figure 1.7: Physiological signals: Pupil Dilation and Respiration Rate.[42]

Physiological signals provide quantitative, objective information free of personal bias. They are crucial to dynamic emotion research because they provide high temporal accuracy for real-time tracking of rapid emotional changes.

Physiological signals may be influenced by non-emotional factors such as physical activity, temperature, or caffeine intake, which can confound results. Their accurate measurement also demands specialized equipment and technical expertise. This requirement increases both the complexity and cost of emotion research.

1.2.2 Behavioral and Observational Methods

Emotion identification emphasizes the interpretation of external behavioral cues, such as body language [12], vocal intonation, and facial expressions, rather than relying on internal physiological data, like EEG or GSR. This method is extensively employed in fields such as human-computer interaction, emotional computing, and psychology.

(a) Facial Expression Analysis: Facial expression [13] analysis uses the Facial Action Coding System (FACS) and advanced deep learning algorithms to determine emotions based on facial muscle activations and micro-expressions. These methods can detect both obvious facial emotions and subtle, involuntary facial movements that reveal underlying emotional states.

Figure 1.8: Facial expressions showing different emotions. [5]

---

[1] Source (Pupil Dilation): `https://oscarwylee.ca/glasses/eye/dilated-pupil-exam`

[2] Source (Respiration Rate): `https://medium.com/@smaha6170/improvement-of-anger-issues-as-a-goal-dbb4d477dfb1`

[5] Source (`https://www.istockphoto.com/photos/different-emotions`).

(b) Body Language and Posture: This method centers on comprehending emotional states by analyzing body posture [14], movement dynamics, and gestures. A slouched posture, for example, may indicate feelings of melancholy, exhaustion, or low energy. Such bodily indicators offer significant nonverbal clues regarding a person's emotional state.

Figure 1.9: Body language and posture showing different emotions. [6]

(c) Eye Tracking and Gaze Behavior: Gaze behavior and eye tracking [15] measure how long and where people concentrate their visual attention. To determine emotional and cognitive states, they analyze fixations, saccades, pupil dilation, and blink rate. This approach is frequently used in behavioral investigations, emotion research, and human-computer interface.

Figure 1.10: Eye tracking and gaze behavior [16].

(d) Speech Signal Characteristics: Speech Signal Characteristics analysis [17], which is independent of speech content, analyzes acoustic components such as tone, pitch, rhythm, and pronunciation to determine emotional states. Variations in characteristics, like a slower rhythm or a higher pitch, represent emotions like happiness, fury, or despair. It is frequently utilized to detect small, involuntary emotional variables in affective computing, psychology, and human-computer interaction.

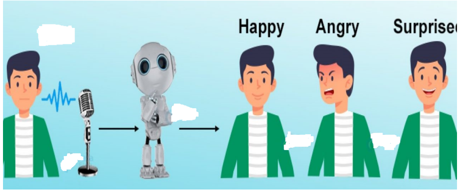


Figure 1.11: Speech features classify emotions [7].

(e) Behavioral Patterns: Behavioral patterns [18] include interaction timing, task performance, and decision-making behavior in structured settings. These behaviors may represent emotional states, such as delayed reactions during anxiety. Analyzing such patterns can help

---

[6] Source (`https://thebftonline.com/2023/11/20/insights-with-dzigbordi-k-dosoo-the-power-of-alignment/`).

[7] Source (`https://tarun-lohex.medium.com`).

one comprehend a person's emotional and cognitive condition.

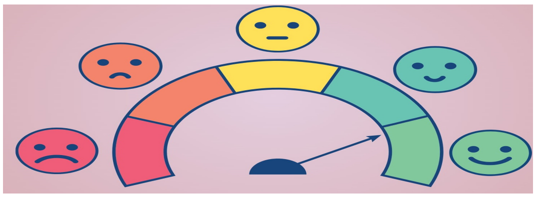


Figure 1.12: Behavioral patterns [8]

Non-invasive behavioral and observational methods provide objective, real-time indicators of emotional states. They record nonverbal clues that self-report techniques could overlook, such as facial expressions, tone of voice, and eye movements. When combined, these approaches provide a thorough comprehension of emotional reactions.

Individual variance, environmental factors, and cultural variances can all affect their precision. For accurate measurement, particular equipment and regulated environments are frequently needed. Non-emotional elements like illness or exhaustion may bias the results.

1.2.3 Neurophysiological Measurement Methods

Neurophysiological Measurement Methods [19] provide significant novel perspectives on the neurological processes that underlie emotional processing. An organized summary of the main techniques for emotion recognition is provided below:

(a) Magneto-Encephalography: Magnetic fields [20] from synchronized neural activity, primarily in cortical regions, are measured using MEG. For surface mapping, it provides superior spatial accuracy and excellent temporal resolution. MEG is used to examine brain oscillations and emotional reactions. However, it is very costly, non-portable, and has limited ability to detect deep brain structures.

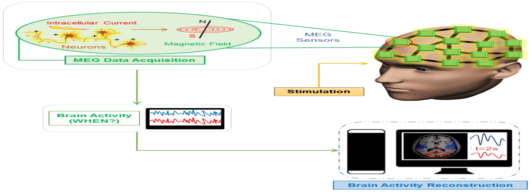


Figure 1.13: Magnetoencephalography [9]

(b) Functional Magnetic Resonance Imaging: Functional magnetic resonance imaging (fMRI) [21] for emotion recognition relies on changes in brain blood flow to identify emotional states by analyzing activity in regions such as the amygdala and prefrontal cortex. When integrated with artificial intelligence models like convolutional neural networks (CNNs) and long short-term memory networks, it can classify emotions with moderate accuracy. It is non-invasive and has high spatial resolution. fMRI has limitations of being expensive and noninvasive. It requires large datasets and complex analysis.

---

[8] (`https://www.psychologytoday.com`)

[9] Source (`https://www.ntu.edu.sg/`).

[10] Source (`https://xrayinterpreter.com/use-case/`).

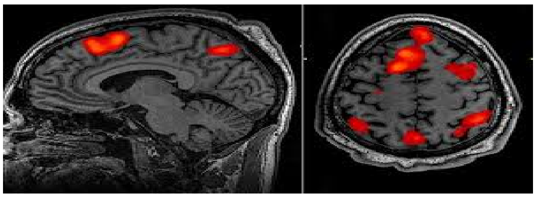


Figure 1.14: Functional Magnetic Resonance Imaging [10]

(c) Electroencephalography: Electroencephalography (EEG) [22] captures the brain's electrical activity using electrodes placed on the scalp. EEG signals primarily arise from the postsynaptic potentials in pyramidal neurons situated in the cerebral cortex. These neurons are organized in such a manner that their collective electrical activity can be detected on the scalp. Neurons communicate through synaptic transmission, in which neurotransmitters trigger the opening of ion channels. This process results in ionic currents flowing across the neuronal membrane, leading to fluctuations in voltage. When large groups of neurons fire in synchrony, their combined electrical fields become detectable as EEG signals. The EEG captures neural activity across different frequency ranges.

Peripheral physiological data are useful for evaluating the autonomic and somatic expressions of emotional states, while EEG is better for examining the brain dynamics and specificity of emotions with high temporal accuracy. Their integration with EEG provides a comprehensive view of the emotion process, bridging brain activity and bodily responses.

Behavioral and observational techniques are used to record the expressive, communicative, and socially significant elements of the phenomenon. While behavioral techniques reveal a state's socially mediated external manifestations, EEG shows its raw neurological origins.

MRI is superior for anatomical accuracy, fMRI for functional localization, and EEG for dynamic neuronal tracking. Their integration in multimodal neuroimaging frameworks is increasingly being studied to enhance research and diagnostic abilities.

EEG is the most efficient method for real-time cognitive monitoring as it directly measures cortical electrical activity with superior temporal resolution, while being portable, non-invasive, and economically viable. In contrast to MRI and fMRI, which are limited by lengthy acquisition times and high costs, EEG allows for the dynamic monitoring of cognition and emotion in practical, scalable settings.

1.3 Motivation

Emotion recognition finds applications in diverse domains. It enables machines to interpret and respond to human emotions, enhancing user experience and adaptability.

(a) Healthcare and Assistive Technologies: The objective of emotion recognition-based mental health monitoring is to identify early warning signs of stress, anxiety, depression, and emotional instability, enabling prompt intervention [23]. Advanced feature extraction and machine learning models categorize these states, allowing for continuous and precise monitoring. Persistent abnormal patterns may trigger alerts, suggest coping strategies, or facilitate connections to professional assistance. This technology, when incorporated into wearable, mobile applications and telehealth systems, provides proactive, real-time mental health support that goes beyond traditional self-report methods.

(b) Human–Computer Interaction: Human–Computer Interaction (HCI) is an adaptive user interface that responds to emotional states [24]. It assists in the development of personalized learning systems in education by adjusting content according to the learner's emotions. Moreover, virtual reality and gaming systems utilize emotion recognition technologies that are guided by personalized learning experiences.

(c) Business and Marketing: Analyzing customer emotions to inform product reviews and advertising [25]. Enhancing customer service by utilizing virtual assistants and chatbots equipped with emotional intelligence. Adjusting services based on consumers' emotional states to improve their overall experience.

(d) Law Enforcement and Security: Recognizing tense, dishonest, or suspicious behaviour in interviewees [26]. Utilizing advanced transportation technologies to detect driver sleepiness and fatigue. Putting in place surveillance systems for public safety that employ emotional computing.

(e) Education: Intelligent tutoring systems [27] that adapt their methods based on the emotions of their students. The incorporation of emotive feedback to enhance e-learning systems.

(f) Media and Entertainment: Individualized suggestions for movies and music according to emotions [27]. Interactive media and narratives that are emotionally charged.

(h) Robotics and Assistive Technologies: Social robots can perceive and address human emotions with compassion [28]. Equipment that helps those with disabilities and the elderly.

1.4 Problem Statement

EEG-based emotion recognition gained substantial attention for applications in human computer interaction and psychological examination. However, its reliability remains low due to several challenges. EEG signals are highly susceptible to noise and artifacts, which reduce signal quality and prevent reliable recognition of emotional states. Moreover, conventional wavelet and time-frequency analysis methods frequently display poor temporal and spectral resolution, constraining their capacity to detect nuanced emotion-related EEG features. Existing deep learning methodologies, particularly transfer learning models, sometimes exhibit significant computational complexity and lack effective generalization across different EEG datasets. Therefore, an integrated framework is required to efficiently denoise EEG signals, produce distinctive time-frequency representations, and achieve reliable emotion classification with improved accuracy, computational efficiency, and generalization capabilities.

1.5 Datasets

1.5.1 DEAP Dataset

The DEAP dataset comprises [29] recordings from 32 individuals who watched 40 music video clips selected as emotional stimuli. Physiological signals and videos of participants' faces were recorded throughout this time. The multimodal dataset includes 40 physiological channels, such as respiration signals, GSR, EEG, EMG, and EOG (electrooculography). On a scale of 01 to 09, participants rated each film based on four emotional dimensions: dominance, liking, arousal, and valence. The EEG signals were down-sampled to 128 Hz after their initial sampling at 512 Hz. The EEG data is organized into 32 channels per participant, with each channel consisting of 63 × 128 samples. Each trial has a duration of 63 seconds, with the initial three seconds serving as the baseline EEG.

1.5.2 DREAMER Dataset

In the DREAMER dataset [30], EEG and ECG signals were recorded from 23 subjects as they viewed 18 emotionally charged movie clips. After viewing each clip, participants rated their levels of dominance, valence, and arousal on a 5-point scale. The EEG signals underwent downsampling to 128 Hz. This research exclusively utilizes EEG signals.

1.5.3 SEED-IV Dataset

The study used the SEED-IV [31] dataset. The dataset is available publicly on request. The dataset has the EEG recordings of 15 subjects. Among them, 7 are males and 8 are females. An ESI neuroscan with 62 channels was used to record the EEG signals. EEG signals were recorded while subjects were watching videos that evoked four emotions: fear, happiness, neutrality, and sadness. The video was 2 minutes in length. Each participant viewed 24 trials in each video across the three independent sessions that constituted the data collection process. Each session involves three steps that include a 5-second cue to begin, a two-minute video clip, and a 45-second self-evaluation stage. EEG signals were captured at a 1000 Hz sampling frequency.

1.5.4 Institutional Dataset

The India Institute of Technology, Design and Manufacturing (IIIT), located in Jabalpur, India, recorded the dataset. EEG recordings were obtained from twenty healthy students, with a mean age of 22.5. Short audio-video clips from Indian movies were used as stimuli. Twelve videos were chosen in all, three for each of the four emotions (fear, sadness, happiness, and relaxation). The Self-Assessment Manikin (SAM) system evaluated the EEG data, acquired at a sample frequency of 256 Hz using a 24-channel EEG recorder, to determine the ground truth for emotional experience. Details on the dataset can be found online in the references [32], [33], and [34].

1.5.5 GAMEEMO Dataset

GAMEEMO is the publicly available [35] database. The dataset contains recorded EEG signals of 28 subjects. A 14-channel EMOTIV EPOC+ Mobile EEG device was employed to record EEG signals from 28 healthy individuals aged between 20 and 27 in the Software

Engineering Department at Firat University. The device records samples at a rate of 2048 Hz but subsequently reduces this to 128 Hz. The signal bandwidth ranges from 0.16 Hz to 43 Hz. A built-in fifth-order sinc filter eliminates motion and external artifacts. EEG data, encompassing both raw and preprocessed formats, were gathered while participants engaged in four audiovisual computer games, each lasting five minutes. These games were classified as boring, calm, horror, and funny. Consequently, this process yielded a total of 20 minutes of EEG data for every participant.

1.5.6 LUMPED-2 Dataset

The Loughborough University Multi-modal Emotion Database-2 (LUMED-2) [36] is a dataset that encompasses various forms of emotional expression. It was created using audiovisual stimuli from a total of 13 subjects, comprising 6 females and 7 males. The stimuli consist of brief video clips, each lasting 8 minutes and 50 seconds, designed to elicit specific emotions. To provide a break, there are 20-second grey screens between segments. Participants recognized that emotional responses can differ from person to person and reported their emotions as ”happy”, ”neutral”, or ”sad” at the end of each session.

1.6 Research Questions

This research is guided by the following questions to address the challenges associated with emotion identification using EEG:
Q1: How can optimal EEG channel selection be used to identify the most emotion-discriminative brain areas while lowering computational complexity and retaining high classification accuracy?
Q2: How can noise and artifacts in EEG signals be efficiently minimised while preserving emotion-related information?
Q3: How can improved time-frequency representations provide more discriminative and informative EEG features than conventional methods?
Q4: How can adaptive signal decomposition be used to effectively eliminate noise and artifacts from EEG signals while maintaining emotional information?
Q5: How can optimal feature selection algorithms be used to identify the most informative EEG features and eliminate redundancy?
Q6: How can an effective deep learning framework be developed to reliably capture subtle temporal and spectral EEG properties?

1.7 Research Objectives

The objective of this research is to develop a robust and adaptive EEG-based emotion recognition framework by integrating advanced signal processing and optimization techniques. The study specifically aims to enhance feature extraction utilizing wavelets, wavelet scattering, and time-frequency analysis. It also seeks to improve signal quality through mode filtering and to employ optimization algorithms for the adaptive selection of decomposition parameters. The research aims to enhance classification accuracy by utilizing machine learning and deep learning algorithms with optimized hyperparameters. This approach contributes to dependable emotion recognition, which has potential applications in mental health monitoring, affective computing, and human–computer interaction. The following are the four main research objectives that have been proposed:

Objective 1: Channel selection optimization: Design an algorithm to optimize the data-recorded channels for emotion recognition.

Objective 2: Noise and artifacts removal Optimization: To design an optimized filter for noise and artifacts removal.

Objective 3: Signal decomposition algorithm optimization: To design an optimize signal decomposition algorithm for emotion recognition.

Objective 4: Deep learning algorithm optimization: To design an optimize deep-learning model for emotion recognition.

The main objective of this thesis is to provide a cohesive and enhanced EEG-based emotion recognition framework that increases recognition accuracy, resilience, and computing efficiency. This work systematically addresses critical challenges in EEG analysis through four interconnected stages: optimal channel selection to minimize redundancy; suppression of noise and artifacts for reliable EEG representation; adaptive signal decomposition; wavelet parameter optimization for effective feature extraction; and deep learning-based hyperparameter optimization for improved classification performance. Each contribution gradually improves the preceding stage, providing a cohesive and integrated methodology for effective emotion recognition.

1.8 Contributions of Research

This research makes the following key contributions:

(a) Efficient EEG Channel Selection for Redundancy Reduction and Computational Optimization: This study presents a computationally efficient EEG channel selection framework that reduces redundancy while preserving signal discriminability. It identifies optimal channel subsets. The approach enhances scalability for real-time systems and demonstrates robust performance across benchmark datasets.

(b) Adaptive Denoising Framework for EEG Signal Integrity: Incorporated mode filtering for noise reduction, improving the quality of EEG signals and enhancing the reliability of emotion recognition systems.

(c) Adaptive Decomposition through Optimization: Proposed the use of optimization algorithms to adaptively select wavelet and IIR filter parameters, ensuring efficient decomposition and more discriminative feature generation.

(d) Enhanced Feature Extraction Framework: Developed a robust feature extraction pipeline for EEG signals using wavelets, wavelet scattering, and time–frequency analysis, enabling improved representation of the non-stationary and multi-scale nature of brain dynamics.

(e) Optimized Machine Learning and Deep Learning Models: Implemented and systematically optimized the hyperparameters of ML and DL algorithms, leading to improved classification accuracy, robustness, and generalization across multiple emotional states.

(f) Integrated Multi-Stage Framework for EEG Feature Extraction and Classification: Designed a comprehensive EEG-based emotion recognition system that combines advanced signal processing, adaptive decomposition, and optimized learning models, bridging the gap between handcrafted features and deep representations.

(g) Benchmarking and Comparative Analysis: Conducted extensive experiments and comparative evaluations with existing methods, demonstrating the superiority of the proposed approach in terms of accuracy and computational efficiency.

1.9 Organization of Thesis

This thesis comprises eight chapters, each addressing different aspects of emotion detection through EEG signals. It highlights the effectiveness of signal processing, machine learning, and deep learning methods.

Chapter 1 Introduction: This chapter introduces the background of emotional studies, the importance of emotion recognition, various models of emotion, and methods for measuring emotions. It also outlines the research objectives, motivation, and contributions of the study.

Chapter 2 Literature Review: A comprehensive review of the current literature on emotion recognition studies is presented, identifying the gaps in research and emphasizing the objectives of the study.

Chapter 3 Optimal EEG Channel Selection for Emotion Recognition: Introduces channel selection methods using Pearson's correlation coefficient and Volt-Kalman filtering. Includes experimental results and comparisons with previous studies.

Chapter 4 Optimal Filter Design for EEG Noise Elimination: The optimal filtered design emphasizes the creation of an optimized Butterworth bandpass IIR filter using COA. It also addresses feature extraction, classification, and performance metrics. Particle swarm–driven filtering optimizes filter design using PSO, applied to DEAP and DREAMER datasets. Includes signal decomposition, signal to image conversion, and classification using deep neural networks.

Chapter 5 An Optimal Mode Filtering Framework for EEG Denoising: Joint time-frequency scattering (JTFS) based framework Presents a novel framework using JTFS for emotion recognition enhancement. Includes silhouette tests and classification results across multiple datasets. Group sparse and super-resolution time-frequency method introduces group sparse mode decomposition and adaptive superlet transform for improved emotion recognition. Includes deep learning model comparisons and computational complexity analysis.

Chapter 6 Optimal Wavelet Design using Nature-Inspired Algorithms: Emotion recognition using optimized wavelet focuses on the use of flexible analytic wavelet transform optimized using Particle Swarm Optimization. It includes detailed methodology, statistical tests, feature selection, and comparative analysis. Optimization in wavelet using crayfish optimization algorithm presents the use of the COA with tunable Q wavelet Transform for emotion recognition. Includes subject-wise analysis and comparisons with deep learning models. Optimization in wavelet scattering transform using PSO describes the use of variational mode decomposition and WST optimized using PSO. Covers feature extraction, classification, and performance evaluation.

Chapter 7 Optimized Deep Learning Models for Emotion Recognition: This chapter introduces various methodologies, including the continuous wavelet transform, smoothed pseudo Wigner-Ville distribution, short-time Fourier transform, and Bayesian Optimization.

Chapter 8 Conclusion: Summarizes the findings, contributions, and potential future directions of the research.

Chapter 2

Literature Review

Emotion recognition has emerged as a pivotal research domain within affective computing, representing a fundamental capability for next-generation human-computer interaction systems. The clinical implications are particularly profound, offering transformative potential for psychiatric assessment and therapeutic interventions, especially for populations with emotional expression deficits such as autism spectrum disorder, schizophrenia, and major depressive disorder [37]. Beyond clinical applications, robust emotion recognition systems are becoming increasingly indispensable across diverse sectors including adaptive virtual reality environments, automotive safety systems that monitor driver states, emotionally intelligent robotics, and personalized human-computer interfaces [38].

The multimodal nature of emotional expression has led researchers to investigate various physiological and behavioral channels for affect detection. Peripheral physiological signals including electrocardiogram, galvanic skin response, blood volume pulse, and heart rate variability provide valuable indicators of autonomic nervous system activity [39]. Behavioral modalities such as facial expression analysis, vocal prosody, and gestural patterns offer additional channels for emotion inference [40]. However, these peripheral measures share a common limitation: their susceptibility to voluntary suppression or social masking. In contrast, EEG provides direct measurement of central nervous system activity with millisecond temporal resolution, capturing neural correlates of emotional processing that are largely immune to conscious control [24]. This neurophysiological foundation, combined with the non-invasive nature and relatively low cost of modern EEG systems, has established EEG as the most promising modality for reliable, real-time emotion recognition [41].

The development of robust EEG-based emotion recognition systems faces several interconnected methodological challenges that have impeded consistent progress and cross-study validation. A fundamental issue concerns the lack of standardization across multiple dimensions of experimental design and implementation. There remains no consensus regarding optimal stimulus selection, with different research groups employing static images from standardized databases, audio clips, video segments, or multimodal stimuli [42]. The temporal dynamics of emotional response elicitation vary significantly across these stimulus types, complicating direct comparison of results. Furthermore, the field lacks standardized electrode montages, with systems ranging from minimal configurations with 2-8 electrodes to high-density arrays with 64-256 channels [43]. This variability in spatial sampling creates significant challenges for reproducing and validating findings across different laboratories.

Electrode selection represents a critical preprocessing challenge with substantial implications for system performance and practicality. High-density EEG systems capture comprehensive spatial information but introduce significant computational burdens, increase vulnerability to artifacts, and heighten the risk of overfitting in machine learning models [44]. Consequently, channel selection methodologies have emerged as an essential component of the EEG processing pipeline. These approaches can be broadly categorized into three methodological families:

Filter methods operate independently of any specific classifier, employing statistical measures to rank channels according to their presumed relevance for emotion discrimination. Common approaches include F-score analysis of between-class versus within-class variance [45], mutual information measures [46], and Relief-based algorithms that weight features according to their ability to distinguish neighboring instances [47]. While computationally efficient, filter methods may select suboptimal channel subsets due to their disregard of classifier-specific interactions.

Wrapper methods employ a specific learning algorithm to evaluate the performance of different channel subsets, typically using search strategies to navigate the combinatorial space of possible configurations [48]. These approaches generally yield superior performance to filter methods but incur substantial computational costs, particularly for high-dimensional EEG data.

Embedded methods integrate the channel selection process directly within classifier training, often through regularization techniques that drive weights associated with irrelevant channels toward zero [49]. Sparse Common Spatial Pattern (SCSP) represents a prominent example, simultaneously optimizing spatial filters and channel selection through L1-norm regularization [50]. While embedded methods offer an attractive balance of performance and efficiency, they often require specialized optimization procedures and may converge to local minima.

The non-stationary character of EEG signals presents fundamental challenges for traditional signal processing approaches. Emotional responses manifest as transient, frequency-specific neural oscillations that evolve over time, necessitating analytical methods that can simultaneously resolve temporal and spectral dynamics. Early approaches relied on temporal or spectral features extracted from predefined frequency bands (delta, theta, alpha, beta, and gamma), but these methods discarded critical information about the temporal evolution of oscillatory activity [51].

The Short-Time Fourier Transform (STFT) represents the most straightforward approach, computing Fourier transforms over sliding temporal windows to generate a spectrogram representation [52]. While computationally efficient and easily interpretable, STFT suffers from the Heisenberg uncertainty principle, manifesting as a fixed trade-off between temporal and frequency resolution that is determined by the window length. This limitation is particularly problematic for emotional EEG analysis, where different frequency bands may require different resolution trade-offs.

Wavelet-based approaches address this limitation through multi-resolution analysis, employing scalable kernel functions that provide high temporal resolution for high-frequency components and high frequency resolution for low-frequency components [53]. The discrete wavelet transform and continuous wavelet transform have been extensively applied to generate time-frequency representations (scalograms) for both manual feature extraction and deep learning applications [54]. However, conventional wavelet transforms suffer from several limitations, including empirical parameter selection, vulnerability to mode mixing, and limited invariance to signal transformations.

Recent advances have produced more sophisticated time-frequency representations that address these limitations:

The wavelet scattering transform creates a deep convolutional network with fixed wavelet filters, generating stable representations that are invariant to translations and deformations while preserving high-frequency information [55]. This property is particularly valuable for emotion recognition, where the absolute timing of emotional responses may vary across trials while their relative spectral patterns remain discriminative.

The Joint Time-Frequency Scattering (JTFS) extends WST by capturing not only spectral content but also modulation patterns, effectively characterizing how spectral components evolve over time [56]. This capability to capture spectral modulations is particularly relevant for emotion recognition, as affective states often manifest as specific patterns of cross-frequency coupling and amplitude modulation.

A particularly promising research direction integrates decomposition methods with nature-inspired optimization algorithms to adaptively tune transform parameters. The tunable Q wavelet transform optimized automatically adjusts the Q-factor to match the oscillatory characteristics of emotional EEG components [57]. Feature discriminability is similarly maximized through adaptable settings for the analytical wavelet transform's dilation and redundancy [53], [32]. These optimization driven methods reduce the dependency on empirical parameter selection while enhancing decomposition quality and ultimately improving classification performance.

The conversion of EEG signals into time-frequency representations has catalyzed the dominance of deep learning architectures in emotion classification, with convolutional neural networks (CNNs) emerging as the predominant approach. The hierarchical feature learning capability of CNNs aligns naturally with the multi-scale structure of time-frequency representations, enabling automatic discovery of discriminative patterns without manual feature engineering [58].

Early CNN architectures applied to EEG emotion recognition demonstrated the feasibility of the approach while revealing fundamental challenges. A 2D-CNN with four convolutional and seven fully connected layers achieved accuracies of 76.56% for arousal and 80.46% for valence on the DEAP dataset, establishing a competitive baseline [59]. Subsequent architectural innovations have focused on enhancing spatial characterization, with regional-asymmetric CNNs (RACNN) employing separate feature extractors for different scalp regions to capture hemisphere-specific emotional processing, achieving remarkable performance up to 95% accuracy [60].

The temporal dynamics of emotional responses have motivated the development of hybrid architectures that integrate CNNs with recurrent components. Convolutional Recurrent Neural Networks (CRNNs) combine spatial feature extraction via convolutional layers with temporal modeling through recurrent layers (LSTM, GRU), effectively capturing the spatiotemporal evolution of emotional states [61]. Bidirectional architectures further enhance temporal modeling by incorporating both past and future context, which is particularly valuable for capturing the extended time course of emotional responses to sustained stimuli such as music videos [62].

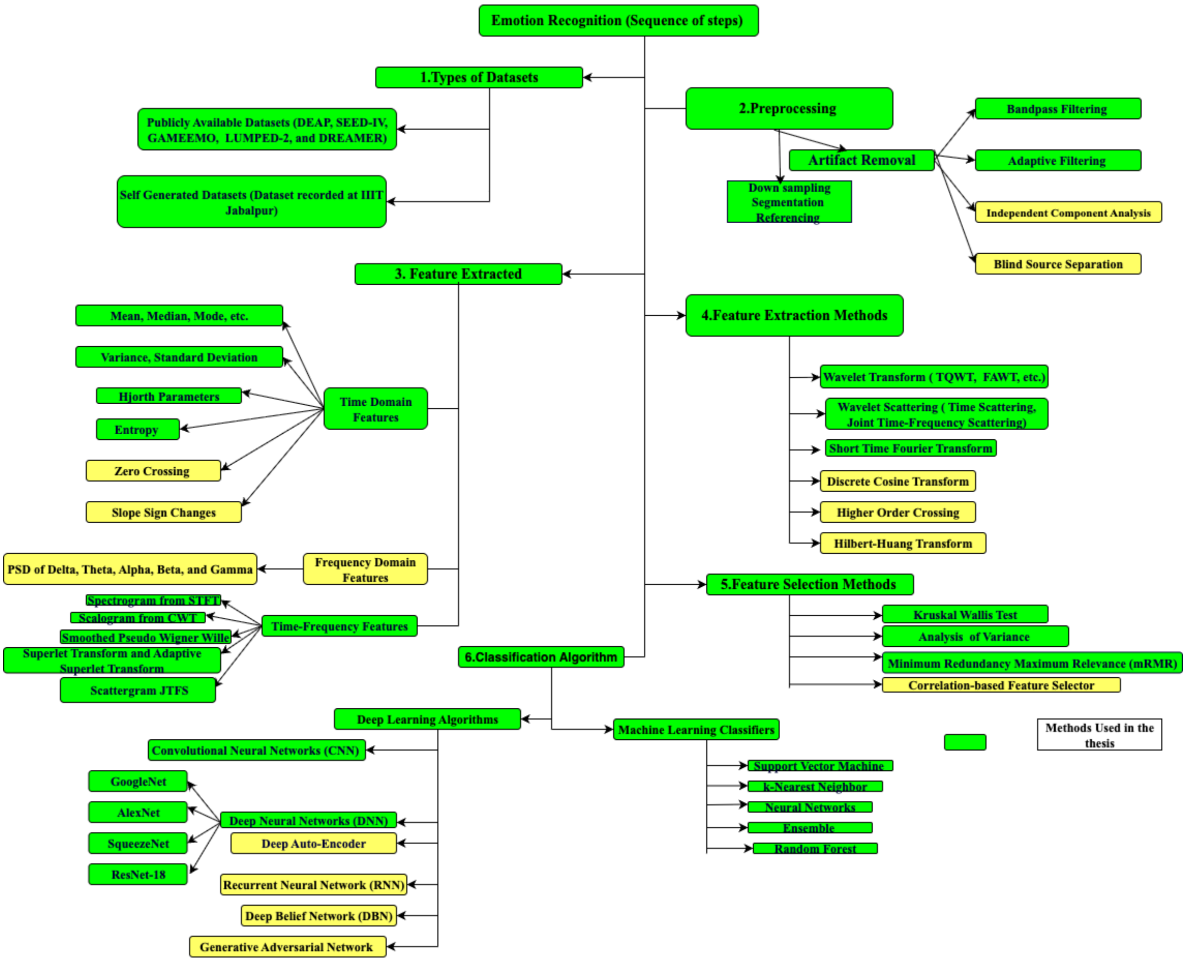


Figure 2.1: General Steps used in EEG-Based Emotion Recognition.

Recent architectural innovations have incorporated attention mechanisms, transformer modules, and sophisticated multiscale processing frameworks to further enhance performance. Models such as Multi-scale Dynamic CNN, Gated Transformer, Convolution Network (MS-DCGNet) and AEF-DL implement multi-scale decomposition before feature extraction, processing different frequency bands with dedicated pathways that are subsequently integrated [63]. Attention mechanisms enable dynamic weighting of relevant spatial regions, temporal intervals, and spectral components, effectively focusing computational resources on the most discriminative aspects of the signal [64]. These advanced architectures have pushed reported accuracies above 99% on specific datasets, approaching what might be considered the theoretical upper limit for binary emotion classification.

However, these performance gains come with substantial costs that limit practical implementation. The computational demands of deep learning models necessitate high-performance hardware, creating barriers to real-time deployment on embedded systems or consumer devices [65]. Extensive hyperparameter tuning requires significant expertise and computational resources, while the massive parameter counts of complex models increase vulnerability to overfitting, particularly given the limited size of publicly available EEG emotion datasets [66]. Most critically, the ”black box” nature of deep learning models creates interpretability challenges that are particularly problematic in clinical applications, where understanding the basis for emotional state classification is essential for trust and adoption [67].

2.1 Taxonomic Approaches to EEG-Based Emotion Recognition

EEG-based emotion recognition methods can be systematically classified into several interconnected stages, including signal preprocessing, feature extraction, feature selection, and classification, as shown in Figure 2.1. Signal preprocessing is designed to improve the EEG signal quality using filtering, artifact elimination, and adaptive signal decomposition methods. Feature extraction techniques are usually classified into time-domain, frequency-domain, and time-frequency domain methods, with advanced representations including the wavelet, wavelet scattering transform (WST), joint time-frequency scattering (JTFS), and superlet-based transforms exhibiting superior discriminative power. Feature optimization uses channel selection, feature selection, and dimensionality reduction approaches to reduce redundancy and computational complexity. Classification frameworks include both shallow machine learning methods and deep learning structures like CNN, CNN-LSTM, and transformer-based models. Finally, hyperparameter optimization techniques such as grid search, Bayesian optimization, and nature-inspired algorithms are utilized to improve classification performance.

Table 2.1: Critical Comparison of Time–Frequency Methods

Method	**Strengths**	**Limitations**	**Research Gap**
STFT	Simple implementation	Fixed time–frequency resolution	Limited adaptability for non-stationary EEG signals
CWT	Multi-resolution analysis	Sensitive to mother wavelet selection	Dependency on empirical parameter tuning
WST/ Wavelets	Translation-invariant and stable feature representation	High computational complexity	Limited capability in capturing temporal modulation patterns
JTFS	Captures temporal modulation and spectral evolution	Increased computational burden	Challenging for real-time implementation
SLT and ASL	High temporal and frequency resolution	Limited studies in EEG-based emotion recognition	Requires further validation across diverse datasets and conditions

Table 2.2: Critical Comparison of Deep Learning Models

Model	**Advantages**	**Limitations**	**Research Gap**
CNN	Automatic feature extraction	Weak temporal dependency modeling	Limited sequential representation
LSTM	Captures temporal information	High computational cost	Difficult optimization
CNN–LSTM	Spatial–temporal learning	Risk of overfitting	Increased complexity
Transfer Learning	High accuracy	Large parameter size	Poor interpretability
Transformer	Long-range dependency modeling	Data intensive	Computationally expensive

2.2 Research Gaps

Existing optimal channel selection algorithms for EEG-based emotion recognition analysis often fail to maintain consistency across cross-subject datasets. While using all available channels can improve robustness, it introduces significant computational overhead and redundancy, making real-time applications impractical. Existing approaches to signal decomposition and time-frequency representation (Table 2.1) often rely on manually tuned parameters or conventional optimization techniques, which may not adequately capture the complex, nonlinear characteristics of biomedical signals. This limitation can result in suboptimal decomposition performance and loss of critical information. Conventional filtering techniques used to remove diverse types of noise and artifacts from EEG recordings often lead to the unintended loss of clinically relevant information, thereby limiting the accuracy of subsequent analysis. The classification accuracy achieved by both deep learning (DL) (Table 2.2) and traditional machine learning (ML) algorithms often depends heavily on manually tuned hyperparameters. This reliance can limit generalizability and efficiency, as suboptimal parameter choices may reduce performance and hinder reproducibility. Current EEG frameworks largely rely on 1D signal representations, which fail to capture spatial–temporal dependencies. Deep learning performance is highly dependent on hyperparameter tuning, yet current manual and conventional optimization methods are computationally expensive, often suboptimal, and lack adaptability.

2.3 Research Objectives

The objectives of this research are centred on advancing EEG-based emotion recognition through methodological and algorithmic innovation. This work aims to develop advanced feature extraction techniques capable of capturing multidimensional emotional patterns, thereby enhancing accuracy, reliability, and generalizability. In parallel, an adaptive channel-selection algorithm will be designed to balance computational efficiency with consistent cross-subject performance. To address signal quality, an improved filtering framework will be introduced that enables precise rhythm decomposition while effectively suppressing noise and artifacts. Complementing this, nature-inspired optimization algorithms will be employed to tune decomposition parameters, ensuring that physiologically meaningful rhythms are preserved while unwanted noise is minimized. Furthermore, the study seeks to establish a standardized approach for transforming one-dimensional EEG signals into two-dimensional spatial–temporal representations, supporting richer and more comprehensive analyses. Finally, automated hyperparameter optimization combined with adaptive learning schemes will be integrated to strengthen the accuracy, robustness, reproducibility, and scalability of machine and deep learning models applied to EEG emotion recognition.

Based on the comprehensive study framework, four primary objectives have been established to methodically address the practical challenges, as outlined below.

Objective 1: Design an algorithm to optimize the data-recorded channels for emotion recognition.

Objective 2: To design an optimized filter for noise and artifacts removal.

Objective 3: To design an optimized signal decomposition algorithm for emotion recognition.

Objective 4: To design an optimized deep-learning model for emotion recognition.

2.4 Summary

The development of reliable EEG-based emotion recognition systems requires careful consideration of data acquisition, preprocessing, feature extraction, and classification. EEG devices vary from low-density (1–8 channels) to high-density (up to 128 channels), making channel selection critical for reducing redundancy, improving computational efficiency, and enhancing classification accuracy. Preprocessing is essential due to noise and physiological artifacts in raw EEG signals. Common steps include down-sampling, re-referencing, and artifact removal using bandpass filtering, adaptive filtering, ICA, and blind source separation. Optimizing artifact removal ensures high-quality feature extraction and improves classifier performance. Feature extraction spans time, frequency, and time-frequency domains. Wavelet-based methods (CWT, DWT) capture temporal and spectral characteristics, while non-adaptive (WPT, TQWT) and adaptive (EMD, VMD, FAWT) approaches have been explored. However, conventional wavelets often suffer from suboptimal parameters. Optimization algorithms like PSO and COA can tune wavelet parameters to enhance feature discriminability and classification performance. For classification, both machine learning (SVM, k-NN, ensemble methods) and deep learning (MLP, CNN, RNN/LSTM, Transformers) models are used. Deep learning performance strongly depends on hyperparameter optimization, which significantly improves accuracy and generalizability. Overall, optimization at multiple stages including channel selection, noise removal, wavelet parameter tuning, and hyperparameter selection is crucial for developing high-performance EEG-based emotion recognition systems. This research focuses on integrating these strategies into a unified and reliable framework.

Chapter 3

Optimal EEG Channel Selection for Emotion Recognition

3.1 Introduction

Emotion recognition plays a vital role in social interactions and daily life, with applications that include assisting patients who are unable to express emotions, supporting psychiatric research and diagnosis, enhancing human-computer interfaces, and developing emotionally intelligent robots. EEG-based systems generally involve feature extraction, selection, and classification [43], with signals recorded via invasive or non-invasive electrodes, the latter being preferred. The number of electrodes ranges from 2–256, though many contain redundant data; optimizing channel selection improves efficiency [68], often guided by event-related potentials [44]. Channel selection techniques fall into three categories: wrapper, filter, and embedded methods. Filter methods use statistical measures to select channels without learning algorithms; examples include normalized mutual entropy [46], asymmetric amplitude ratio (AAR), asymmetric variance ratio (AVR), and the F-score index for between/within-class variation [45]. Relief-based methods also identify emotion-relevant channels while reducing redundancy [47]. Wrapper methods use learning algorithms to evaluate subsets, while filter methods are computationally cheaper but may lose accuracy. Deep learning-based wrappers use mean absolute weight distribution to identify optimal channels[48]. Multi-objective swarm optimization methods (dragonfly, cuckoo search, grey wolf, particle swarm optimization) have also been applied to identify important channels. Embedded methods integrate feature selection with classifier training, e.g., sequential search strategies in BCI systems [50]. Sparse Common Spatial Pattern (SCSP) [49] optimizes channel selection for maximum accuracy with minimal channels, either by reducing redundancy or eliminating noisy inputs. However, most methods are computationally expensive. The proposed work adopts threshold PCC for channel selection, followed by the VKF algorithm to decompose data into frequency sub-bands for extensive feature extraction. Maximum relevance minimum redundancy (MRMR) is then applied to select the most relevant features, which are subsequently classified using various algorithms.

3.2 Methodology

The block diagram in Figure. 3.1 displays the sequential process of the proposed methodology.

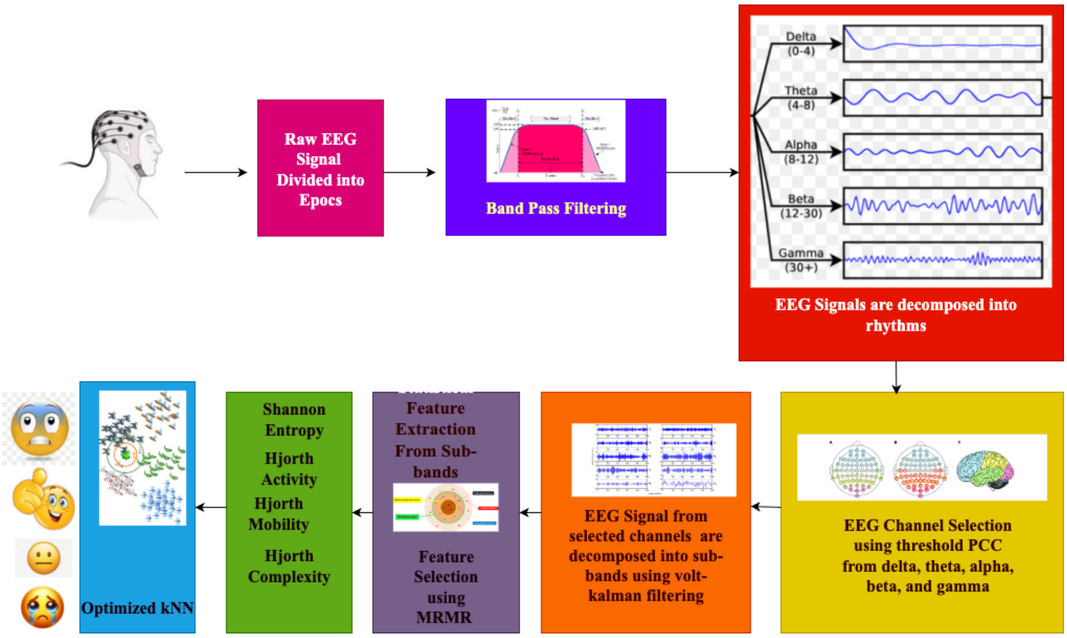


Figure 3.1: Block diagram of the proposed framework for optimal EEG channel selection in emotion recognition.

3.2.1 Volt-Kalman Filtering

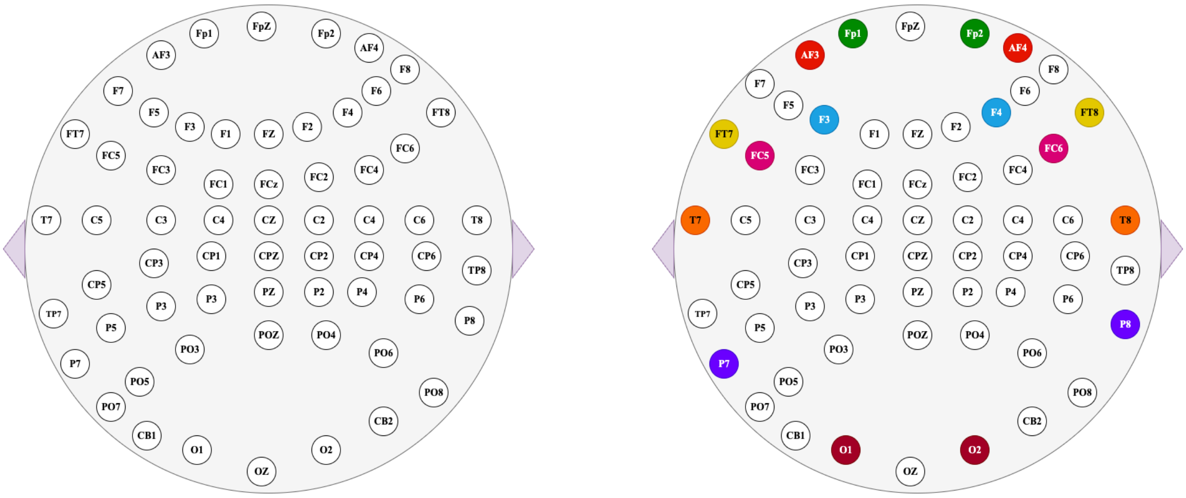
Hvard Vold and Jan Leuridan developed the Vold–Kalman Filter (VKF) method, which uses frequency vectors to separate non-stationary signal components [69]. Widely applied in various fields, it effectively isolates desired sequences and is particularly suited for complex multi-component signals such as EEG [70]. Unlike EMD and VMD, which rely on decomposition and filtering, VKF extracts components directly from the time domain, avoiding phase shifts from time-frequency conversion. It also adapts centre frequencies dynamically to instantaneous frequency, reducing overlap compared to EMD. Moreover, VKF efficiently tracks orders in systems with multiple shafts, it facilitates extraction of waveforms, amplitude, and phase while avoiding beating interactions common in conventional methods.

3.2.2 Pearson's Correlation Coefficient

The Pearson correlation coefficient (PCC) [71] quantifies the linear relationship between two variables. It is defined as the covariance of variables M and N divided by the product of their standard deviations:

$$\rho(M, N) = \sum_{i=1}^{n} \left(\frac{M_i - \bar{M}}{\sigma_M} \right) \left(\frac{N_i - \bar{N}}{\sigma_N} \right) \tag{3.1}$$

Here, $\bar{M}$ and $\bar{N}$ are the mean values of M and N. The coefficient ρ ranges from -1 to 1, indicating both the strength and direction of the relationship: positive values represent a direct association, negative values an inverse one. The ranges and interpretations are listed in Table 3.1. Pearson correlation is particularly suitable when variables follow a normal



(a) SEED IV dataset's 62 EEG electrode setup (b) Electrode setup with channel selection

Figure 3.2: SEED-IV 62-channel EEG electrode layout with selected channels highlighted.

distribution, as in EEG signals.

Table 3.1: Pearson's correlation coefficient interpretation.

PCC Value	Correlation Intensity
0.8–1.0	Very Strong Correlation
0.6–0.8	Strong Correlation
0.4–0.6	Medium Correlation
0.2–0.4	Weak Correlation
0.0–0.2	No Correlation

3.2.3 Statistical Time Domain Features

Hjorth parameters [72] are statistical measures widely applied in neuroscience and biomedical signal processing to describe the time-domain characteristics of EEG and other physiological signals. They comprise three key parameters, including activity, mobility, and complexity.

3.2.3.1 Hjorth Complexity (HC)

Hjorth complexity reflects the predictability or irregularity of a signal. It is calculated as the ratio of the mobility of the signal's first derivative to the overall mobility of the signal. Mathematically, HC is defined as [72]

$$HC = \frac{\text{Mobility}\left(\frac{ds(t)}{dt}\right)}{\text{Mobility}\ (s(t))} \tag{3.2}$$

Where s(t) denotes the signal.

3.2.3.2 Hjorth Mobility (HM)

Hjorth mobility quantifies the signal's frequency characteristics or rate of variation. It is computed as the ratio between the standard deviation of the signal's first derivative and the standard deviation of the signal itself. The first derivative is typically estimated using finite differences or other numerical methods. Mathematically, HM is defined as [72]

$$HM = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\dot{s}_i(t) - \overline{\dot{s}(t)}\right)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(s_i(t) - \overline{s(t)}\right)^2}} \tag{3.3}$$

Where N denotes the number of samples, $s(t)$ denotes the signal, $\overline{s(t)}$ denotes the mean of the signal, $\overline{\dot{s}(t)}$ denotes the derivative of the mean, and $s(t)_i$ denotes the i-th sample of $s(t)$.

3.2.3.3 Hjorth Activity (HA)

Hjorth activity represents the overall power or energy of a signal. It is calculated as the variance of the signal. Mathematically, HA is defined as [72]

$$HA = \frac{1}{N}\sum_{i=1}^{N}\left(s(t)_i - \overline{s(t)}\right)^2 \tag{3.4}$$

Where N denotes the number of samples, $s(t)$ denotes the signal, and $\overline{s(t)}$ denotes the mean of the signal.

3.2.3.4 Shannon Entropy (S_{en})

A higher Shannon entropy value signifies greater uncertainty in the outcome. It quantifies the level of unpredictability associated with the occurrence of an event. The formula for its computation is given as [73]:

$$S_{\text{en}} = -\sum_{i=0}^{N-1} \left(p(s)_i\right)^2 \left(\log_2\left(p(s)_i\right)\right)^2 \tag{3.5}$$

Where S is the random variable, p(s(t)) is the probability density function of S, and s(t) represents the signal in the time domain.

“‘latex

3.3 Proposed Method and Results

3.3.1 Preprocessing

The SEED IV [31] dataset provides both raw and preprocessed EEG data, with this study utilizing the raw recordings. Each subject’s file contains 24 recordings (trials), characterized by the number of electrodes and samples, ranging from 8000 to 45000 samples across 62 electrodes. Since emotions emerge at the end of trials, the last 4.68 s of each recording (4680 samples at 1000 Hz) are used. The signals are first band-pass filtered in the 0.4–60 Hz range and subsequently amplitude-normalized across all epochs to streamline computation.

3.3.2 Channel Selection

The SEED IV dataset records EEG using 62 electrodes, but not all are necessary for emotion classification, as they increase redundancy and computational cost. Algorithm 3.1

Table 3.2: Interpretation of EEG Features

Feature	**What it Measures**	**EEG Emotion Relevance**	**Why Selected**
Hjorth Activity	Signal power / variance	Cortical activation and neural intensity	Quantifies EEG energy variations across emotional states
Hjorth Mobility	Mean frequency variation	EEG rhythm transitions during emotions	Sensitive to temporal changes in brain activity
Hjorth Complexity	Signal waveform complexity	Nonlinear EEG irregularities	Characterizes complexity variations among emotions
Shannon Entropy	Signal uncertainty / randomness	Irregular EEG patterns linked to emotions	Represents EEG information content and complexity

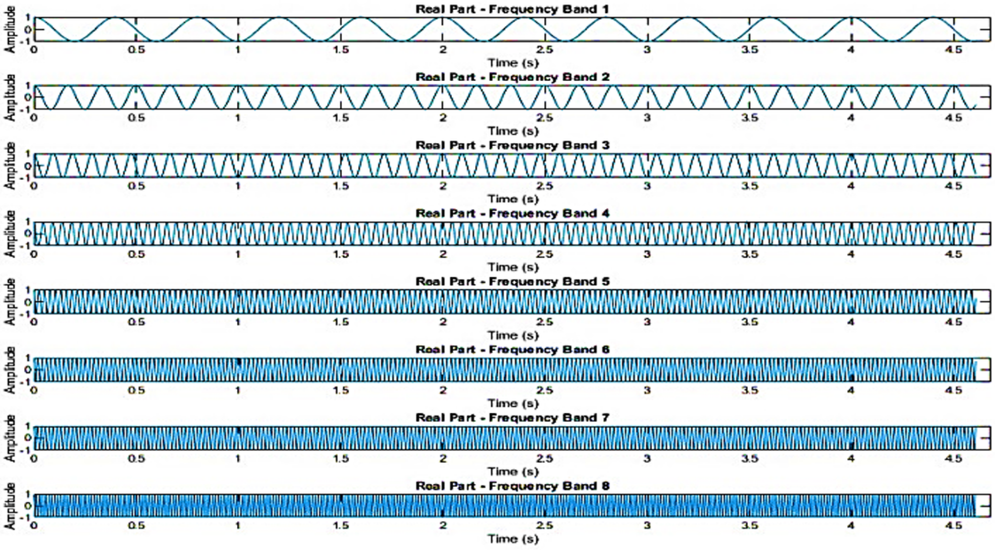
outlines the selection of the most relevant channels. EEG signals are decomposed into five bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–60 Hz, occasionally above 60 Hz). Alpha, beta, and gamma bands show significant variation in Pearson Correlation Coefficient (PCC), where values above 0.4 indicate moderate-to-high similarity. Channels Fp1, Fp2, FC5, FC6, O1, O2, P7, P8, F3, F4, F7, F8, T7, T8, AF3, and AF4 exhibit average PCC values of 0.4–0.7 in these bands, highlighting the frontal and occipital regions as most active, while other regions contribute little to emotion classification. Figure 3.2a illustrates all 62 channels used for EEG signal recording, while Figure 3.2b presents the selected channels obtained through Algorithm 3.1.

Algorithm 3.1 Optimal EEG Channel Selection Using Pearson Correlation Coefficient (PCC)

```
Set PCC threshold range: [0.4, 0.7]
Define EEG frequency bands: $\delta$ (1–4 Hz), $\theta$ (4–8 Hz), $\alpha$ (8–13 Hz), $\beta$ (13–30 Hz), $\gamma$ (30–60 Hz)
Initialize empty set of selected channels: $C_{sel}$
for each EEG electrode $E_i$, $i = 1, \ldots, 62$ do
    Decompose EEG signal into $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$ bands using bandpass filtering
    for each frequency band do
        Compute PCC value for the EEG signal
    end for
    Compute average PCC in $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$ bands
    if any PCC value lies in $[0.4, 0.7]$ then
        Add channel $E_i$ to $C_{sel}$
    end if
end for
Output: Selected EEG channels $C_{sel}$
```

3.3.3 Signal Decomposition using Vold–Kalman Filter

The VKF algorithm extracts filters of order 1–4 based on a given frequency vector, isolating signal components within specific frequency ranges. Higher orders yield sharper roll-off and more accurate filtering, while flexibility in construction is achieved through boundary conditions and lower-order coefficients. VKF also controls spectral properties by allowing constant (scalar) or time-varying (vector) bandwidth specification. In this work, the default filter order of 2 is used with a 1000 Hz sampling rate. EEG signals are decomposed into eight bands: 1–4 Hz, 4–8 Hz, 8–13 Hz, 13–30 Hz, 30–45 Hz, 45–50 Hz, 50–55 Hz, Figures 3.3 and 3.4 display the real and imaginary parts of the decomposed sub-bands, which are then used for statistical feature extraction.



Real Part - Frequency Band 1

Real Part - Frequency Band 2

Real Part - Frequency Band 3

Real Part - Frequency Band 4

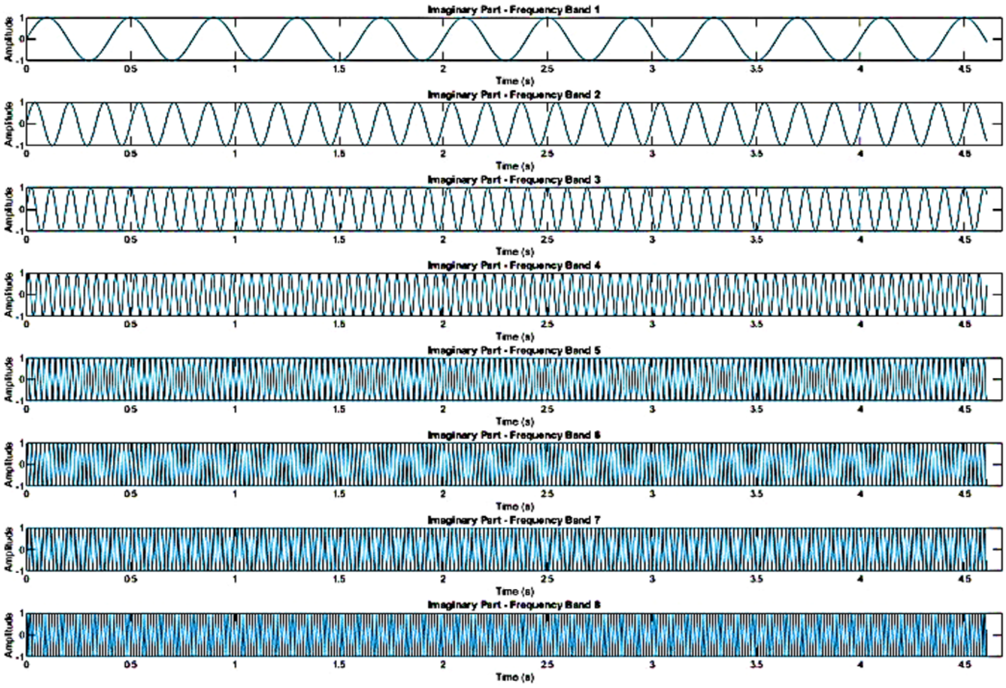
Real Part - Frequency Band 5

Real Part - Frequency Band 6

Real Part - Frequency Band 7

Real Part - Frequency Band 8

Figure 3.3: Real components of the eight frequency bands obtained using VKF decomposition.



Imaginary Part - Frequency Band 1

Imaginary Part - Frequency Band 2

Imaginary Part - Frequency Band 3

Imaginary Part - Frequency Band 4

Imaginary Part - Frequency Band 5

Imaginary Part - Frequency Band 6

Imaginary Part - Frequency Band 7

Imaginary Part - Frequency Band 8

Figure 3.4: Imaginary components of the eight frequency bands obtained using VKF decomposition.

3.3.4 Feature Extraction and Selection

This section investigates the selection of optimal statistical features for EEG-based emotion analysis using a subject-independent approach, where data from all subjects are combined for feature extraction. Initially, seventeen statistical features are evaluated, including mean, skewness, kurtosis, root mean square, Shannon entropy, log energy entropy, mean curve length, Hjorth activity, Hjorth mobility, and Hjorth complexity. Additional features such as mean energy, approximate entropy, log root of sequential variation, and mean total energy are also considered. To identify the most discriminative features, the Minimum Redundancy Maximum Relevance (MRMR) algorithm is applied. Based on feature importance scores, four features emerge as most relevant—Shannon entropy, Hjorth activity, Hjorth mobility, and Hjorth complexity.

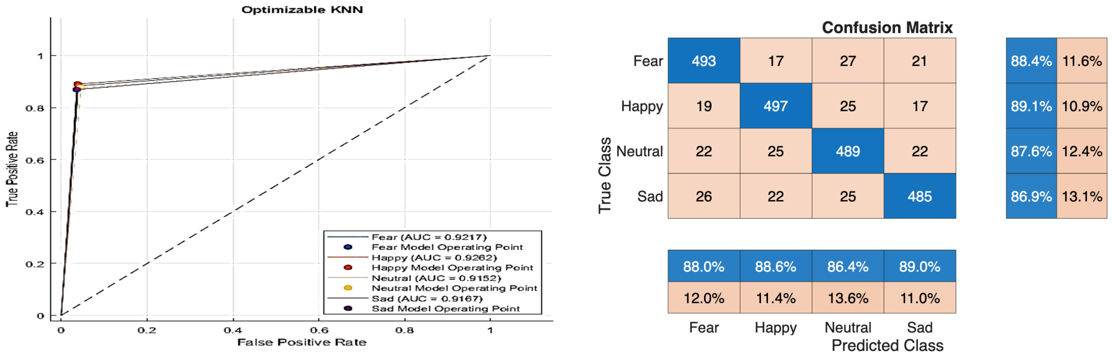


Figure 3.5: Performance evaluation of the O-KNN classifier for emotion recognition using optimally selected EEG channels: (a) ROC curve and (b) confusion matrix.

3.3.5 Classification using Multiple Machine Learning Models

Following the identification of the most relevant features through the MRMR approach, these features are evaluated using various classification algorithms. The selected features are sequenced randomly. The randomly sequenced training and testing sets ratio is 3:1 or 75:25. Seventy-five percent of the features are used as the training set and 25% as the testing set (subject-independent). Table 3.3 displays the results of the test set, including all channels and selected channels.

Table 3.3: Classification accuracy before and after channel optimization.

Model	**All Channels (%)**	**Selected Channels (%)**	**Improvement (%)**
Optimizable KNN	85.2	88.0	+2.8
Fine KNN	83.7	86.8	+3.1
Wide Neural Network	80.1	80.2	+0.1
Fine Gaussian SVM	80.0	81.2	+1.2
Weighted KNN	79.0	80.8	+1.8
Ensemble Bagged Trees	77.5	78.7	+1.2
Kernel SVM	76.4	78.1	+1.7
Cubic SVM	76.1	82.5	+6.4
Average	**79.5**	**82.0**	**+2.5**

Table 3.3 shows that selecting the most relevant channels significantly improves classi-

fication accuracy. Among all classifiers, optimizable KNN (O-KNN) achieves the highest accuracy. Within the SVM group, the fine Gaussian model performs best, while wide neural networks and ensemble bagged trees outperform other variants of neural networks and ensemble methods, respectively.

3.3.6 ROC and Confusion Matrix

The Receiver Operating Characteristic (ROC) curve in Figure 3.5 (a) illustrates the relationship between the true positive rate and the false positive rate. The Area Under the Curve (AUC) quantifies the discriminative capability of the classifier. The ROC curve of the optimizable K-Nearest Neighbor (O-KNN) classifier achieves an AUC value exceeding 0.91, indicating high accuracy in emotion recognition. Figure 3.5 (b) present the confusion matrix and sensitivity matrix obtained after applying the O-KNN classifier.

The diagonal elements of the confusion matrix highlight the classifier's effectiveness in correctly predicting the majority of true labels. Sensitivity analysis further confirms the classifier's performance, with all classes achieving sensitivity values above 88%. This result demonstrates the model's strong ability to reliably identify all emotion categories within the dataset.

Table 3.4: Comparison of emotion recognition methods using EEG signals.

No.	Author(s)	Year	Method	Accuracy
1	Pei Xiang Zhong [74]	2021	Regularized graph neural network (RGNN)	79.37% ± 10.54%
2	Jiyao Liu et al. [75]	2022	Transformer	83.27% ± 8.37%
3	Sijin Zhou, Dongmin Huang [76]	2022	Subjective and objective feature fused neural network (SOFNN)	86.27% ± 10.16%
4	**Proposed Method**	**2024**	**VKF, Optimizable KNN, HC, HM, HA, and SE**	**88.00%**

3.4 Comparison with Previous Studies

Table 3.4 summerizes recent previous studies with SEED-IV dataset. Zhong et al. [74] proposed a Regularized Graph Neural Network (RGNN), which integrates two regularized networks NodeDAT and EmotionDL and achieved a mean classification accuracy of 79.37%. Liu et al. [75] developed four transformer-based frameworks that exploit both temporal and spatial information, reporting an average classification accuracy of 83.27%. Zhou et al. [76] introduced a Subjective-Objective Feature-Fused Neural Network (SOFNN), trained on spatial-temporal EEG features combined with eye movement data, yielding a mean classification accuracy of 87.27%.

In contrast, the proposed method performs emotion recognition using a minimal number of EEG channels, thereby reducing data redundancy and computational complexity. Furthermore, the application of the VKF algorithm for signal decomposition effectively addresses mode mixing and sub-band resolution limitations commonly encountered in traditional decomposition approaches.

3.5 Summary

The first objective of this research focuses on developing an EEG-based emotion recognition method with an optimal set of electrodes. Channel selection is guided by Pearson's correlation coefficient applied to EEG rhythms, allowing the identification of the most informative channels while eliminating redundancy and reducing hardware requirements. The selected EEG signals are then decomposed into multiple frequency sub-bands using the volt Kalman filter order algorithm, from which statistical features, including Hjorth parameters and Shannon entropy, are extracted. To ensure only the most relevant features are retained, the MRMR algorithm is employed for feature selection. Classification using an optimizable kNN model achieves an accuracy of 88%, demonstrating improved performance over existing methods. This strategy not only reduces the number of required channels but also facilitates the analysis of brain regions associated with specific emotional states.

Building on this foundation, the next research objective addresses optimization in noise and artifact removal, as raw EEG signals are often contaminated by physiological and environmental artifacts that can degrade feature quality and classifier performance. By integrating optimized preprocessing techniques, including adaptive filtering and adaptive mode filtering, the proposed framework ensures cleaner signals, thereby enhancing the reliability and accuracy of subsequent feature extraction and classification.

Chapter 4

Optimal Filter Design for EEG Noise Elimination

4.1 Introduction

EEG signals are commonly decomposed into five rhythms, including δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–35 Hz), and γ (>35 Hz). Each associated with specific brain states ranging from deep sleep to high cognitive arousal [77, 78, 79, 80]. These rhythms are complementary in capturing emotional states, making rhythm-specific analysis a foundation for EEG-based classification [81, 82, 83, 84, 85, 86].

Various studies extract features from EEG rhythms for emotion recognition. Statistical and entropy-based features achieve 95.71% accuracy with k-NN and 80.32% with Random Forest [87], while lower-entropy segments improve precision. Using CWT features from α, β, and γ bands further increases k-NN accuracy to 98% [88]. The α–β power ratio has been used for subject-dependent classification of valence, arousal, and dominance [89]. Power spectral features from eight emotion-related EEG channels enable SVM to reach 94.27% accuracy [90]. Emotion-specific electrode epochs across multiple frequency bands yield 77.21%–99.48% accuracy, while phase synchronisation indices from θ, α, β, and γ bands support valence–arousal classification [91]. Multivariate fast iterative filtering extracts multichannel rhythmic features, with γ rhythms classified using ensemble methods achieving 93.5–99.8% accuracy [92].

EEG-based emotion recognition is further enhanced using advanced methods: common spatial patterns with linear SVM classify happiness and sadness in the γ band with 93.5% accuracy for 3-second trials and 93.0% for 1-second trials [93], the RT-ERM with LSTM networks classifies emotions across rhythms and time scales [94], and gated transformer networks leverage spectral features of five frequency bands and 3D sequences [95]. All approaches typically decompose EEG signals into rhythms using bandpass filters and wavelets.

Despite these advances, challenges remain. Wavelet- and scalogram-based methods are sensitive to temporal variability, spectrograms may omit phase information, and deep networks often require high computational resources [55, 96, 97].

To address these limitations, this study proposes an optimized IIR band-pass filter framework for EEG-based emotion recognition. The filter order and cutoff frequencies are adaptively tuned to minimize reconstruction error while preserving rhythm-specific content.

4.2 Selection of Optimization Algorithms

The properties of the EEG filtering difficulty, including nonlinearity, multimodal search spaces, and parameter selection sensitivity, influenced the choice of optimization methods. The Crayfish Optimization Algorithm (COA) was selected because of its high global exploration capabilities and adaptive search mechanism, which helps to avoid premature convergence during filter parameter optimization. Because EEG rhythm extraction requires accurate boundary selection across nonstationary frequency bands, COA maintains an appropriate balance between exploration and exploitation. In contrast, Particle Swarm Optimization (PSO) was selected for IIR filter optimization due to its computational simplicity, rapid convergence behavior, and efficacy in continuous parameter optimization issues.

4.3 Optimal Filter Design using Cray Fish Optimization Algorithm

4.3.1 Methodology

The block diagram for proposed methodology is displayed in the Figure 4.1.

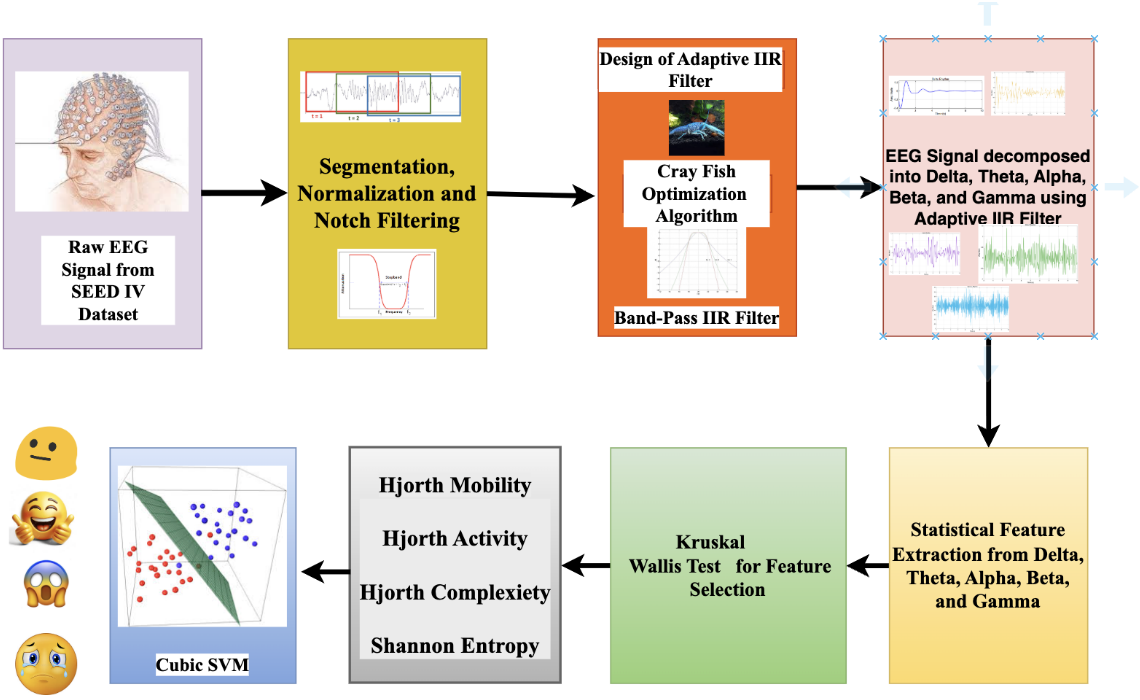


Figure 4.1: Block diagram of the proposed methodology for optimal EEG filter design using the crayfish optimization algorithm.

4.3.1.1 Butterworth Bandpass IIR Filter

The infinite impulse response (IIR) bandpass filter [98] allows specific frequency bands to pass while attenuating others, using an iterative process based on current and past inputs and outputs. It is defined by lower (f_L) and upper (f_H) cutoff frequencies, which suppress signals below and above the passband. The Butterworth filter, characterized by a monotonic response without ripples, uses filter order to control transition steepness, with higher orders producing sharper roll-off. Its digital transfer function in the z-domain is given by [98]

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{4.1}$$

where a_1, a_2 are denominator coefficients and b_0, b_1, b_2 are numerator coefficients. Band-pass design depends on sampling frequency, cutoff frequencies, and filter order to determine the coefficients and desired characteristics.

4.3.1.2 Crayfish Optimization

The crayfish optimization algorithm [99] is a nature-inspired method that simulates crayfish foraging, survival, and competition behaviors to solve multidimensional optimization problems. By alternating between exploration and exploitation, it efficiently converges to the optimal solution.

(a) Initialization:

The population of crayfish is initialized as a matrix $\mathbf{X}$, where each element $X_{i,j}$ is generated randomly within the specified bounds [99]:

$$X_{i,j} = \text{lb}_j + (\text{ub}_j - \text{lb}_j) \times \text{rand}, \tag{4.2}$$

where ub_j and lb_j are the upper and lower bounds for the j-th dimension, and rand is a random number in the range $[0, 1]$.

(b) Foraging Intake and Temperature:

The foraging intake p is represented as a normal distribution influenced by temperature [99]:

$$p = C_1 \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\text{temp} - \mu)^2}{2\sigma^2}\right) \tag{4.3}$$

where σ and C_1 are parameters controlling intake, and μ is the optimal temperature.

The environmental temperature is defined as [99]:

$$\text{temp} = \text{rand} \times 15 + 20 \quad (20^\circ\text{C} \leq \text{temp} \leq 35^\circ\text{C}) \tag{4.4}$$

(c) Summer Resort Stage (Exploration):

Crayfish seek shade (optimal solutions) and migrate towards a ”cave” when the temperature rises to or above 30°C. The shaded location X_{shade} is defined as [99]:

$$X_{\text{shade}} = \frac{X_G + X_L}{2} \tag{4.5}$$

where X_L is the local best position and X_G is the global best position. Crayfish approach the cave according to the following equation [99]:

$$X_{i,j}^{t+1} = X_{i,j}^{t} + C_2 \times \text{rand} \times (X_{\text{shade}} - X_{i,j}^{t}) \tag{4.6}$$

$$C_2 = 2 - \frac{t}{T} \tag{4.7}$$

where C_2 is a linearly decreasing coefficient with iterations.

(d) Competition Stage (Exploitation):

Crayfish compete for the cave by shifting their location toward another randomly selected crayfish X_z when the temperature exceeds 30°C and rand ≥ 0.5 [99]:

$$X_{i,j} = X_{i,j} + \text{rand} \times (X_z - X_{i,j}) \tag{4.8}$$

This stage enhances diversity and broadens the search patterns, improving exploration capability.

(e) Foraging Stage (Exploitation):

At temperatures below 30°C, crayfish concentrate on foraging by moving toward the food source X_{food}, representing the optimal solution [99]:

$$X_{i,j}^{t+1} = X_{i,j}^{t} + \text{rand} \times (X_{\text{food}} - X_{i,j}^{t}) \tag{4.9}$$

Foraging behavior is influenced by food size Q. If $Q > \frac{C_3+1}{2}$, crayfish use a sine-cosine-based strategy to alternate directions and finely shred the food. Otherwise, they approach and consume the food directly. The Crayfish Optimization Algorithm (COA) dynamically switches between exploration and exploitation phases based on environmental factors like temperature and food size. This natural mimicry ensures adaptability and provides fast convergence toward optimal solutions.

4.3.1.3 Statistical Characteristics

Four statistical features including, Hjorth Mobility, Hjorth Activity, Hjorth Complexity, and Shannon Entropy have been used in the study.

4.3.2 Results and Discussion

This section discusses the obtained results and their interpretation. It also shows the empirical values used in the different stages of obtaining the final results.

4.3.2.1 Preprocessing

The SEED IV [31] dataset includes both preprocessed and raw EEG recordings. Each subject file contains 24 trials, with recording dimensions defined by channels and sample length, ranging from 8,000 to 47,000 samples. For analysis, the final 4.68 seconds, where emotions are most evident, are used, yielding 4,680 samples per epoch at a 1,000 Hz sampling rate.

4.3.2.2 Optimized Filter Design using Cray Fish Optimization Algorithm

The key design parameters of the IIR Bandpass Butterworth filter are its order and cutoff frequencies. The optimization algorithm aims to determine their optimal values based on EEG data by decomposing signals into five frequency bands (delta, theta, alpha, beta, gamma). The objective function minimizes the mean squared error (MSE) between the original signal and the sum of the filtered bands. As shown in Table 4.1, the COA algorithm uses T (maximum iterations), N (number of search agents/crayfish), dim (number of parameters), and lb, ub (search space bounds for filter order and cutoff frequencies).

Table 4.1: Parameters and their values used for optimization.

Parameter	**Value**
dim (Dimension)	11
T (Maximum Iteration Count)	100
N (Number of Search Agents)	50
lb (Lower bounds)	[2, 0.5, 1, 4, 5, 8, 9, 12, 13, 30, 31]
ub (Upper bounds)	[10, 4, 5, 8, 12, 13, 30, 31, 50, 63, 64]
Sampling Frequency	1000 Hz

Table 4.2: Optimized parameters for IIR filter design.

Parameter	**Optimized Value**
Filter Order	5.93517
Delta Frequency Band	1.9591 Hz to 4.60128 Hz
Theta Frequency Band	4.61615 Hz to 11.933 Hz
Alpha Frequency Band	11.9326 Hz to 24.1123 Hz
Beta Frequency Band	24.1089 Hz to 31.115 Hz
Gamma Frequency Band	31.2047 Hz to 50.1338 Hz
Minimum MSE (Mean Squared Error)	0.029211

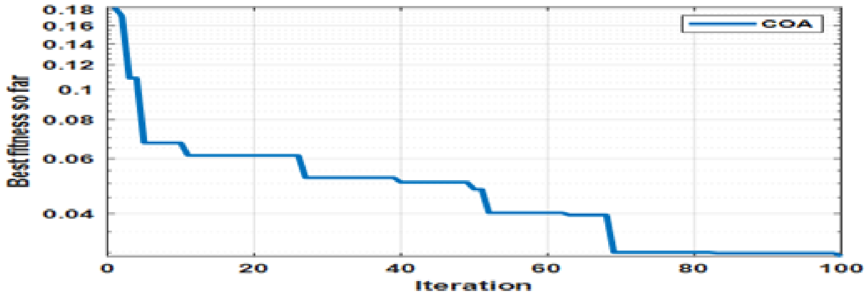
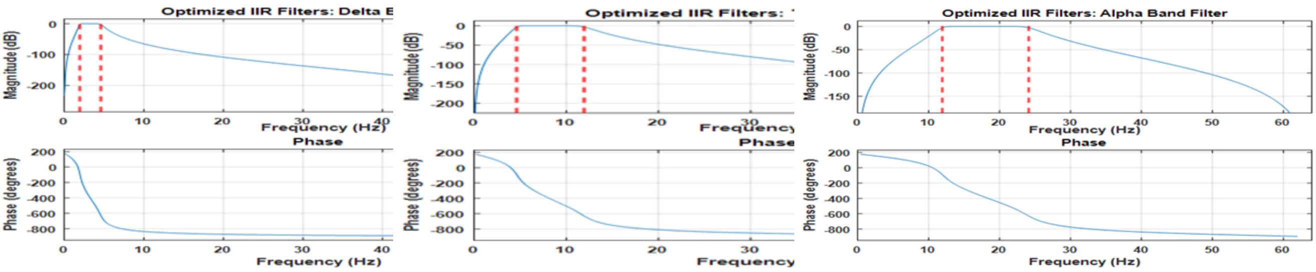


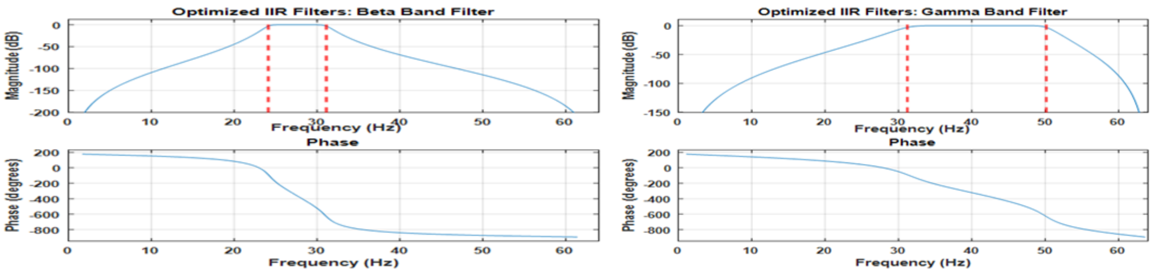
Figure 4.2: Fitness curve (MSE) during optimization using COA.

Figure 4.2 shows the convergence curve, illustrating the minimization of MSE between the original and reconstructed signals across iterations. MSE decreases from 0.18 before optimization to 0.0292111 after optimization, indicating minimal information loss. This confirms that the optimized band-pass filter effectively decomposes EEG signals into frequency bands without distortion, with the reconstructed signal closely matching the original, validating the filter design.

Table 4.2 presents the optimized filter design parameters obtained from the COA algorithm, with the filter order optimized to 5.93517. The filter order balances frequency selectivity and computational efficiency, as higher orders improve selectivity but increase complexity. The optimized magnitude and phase responses for decomposing EEG into five bands are shown in



(a) Delta Band (b) Theta Band (c) Alpha Band



(d) Beta Band (e) Gamma Band

Figure 4.3: Magnitude and phase responses of the optimized bandpass filters for delta, Theta, Alpha, Beta, and Gamma EEG bands.

Figures 4.3a, 4.3b, 4.3c, 4.3d, and 4.3e with cutoff frequencies of 1.9591–4.60128 Hz (delta), 4.61615–11.933 Hz (theta), 11.9326–24.1123 Hz (alpha), 24.1089–31.115 Hz (beta), and 31.2047–50.1338 Hz (gamma). These ranges align with characteristic EEG rhythms while being precisely adapted to the dataset.

4.3.2.3 Features Extraction

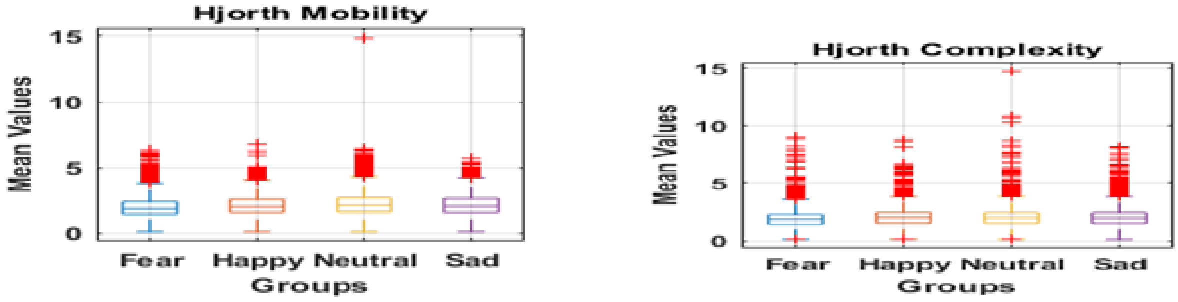
EEG signals are decomposed into five frequency bands using an optimized Bandpass IIR filter, and statistical features are extracted. The Kruskal-Wallis algorithm is applied to evaluate feature importance. Shannon Entropy, Hjorth Mobility, Hjorth Complexity, and Hjorth Activity achieve the highest relevance scores. Their p-values are 1.17×10^{-9}, 7.31×10^{-18}, 2.04×10^{-11}, and 1.32×10^{-14} shown in Figures 4.4a, 4.4b, 4.4c, and 4.4d, respectively—all below the 0.05 threshold—confirming that these features significantly distinguish between the four emotions.

4.3.2.4 Classification

Four features are extracted from five frequency bands and split into training (90%) and testing (10%) sets using a subject-independent analysis. SVM and its variants achieve the highest test accuracy, with cubic and quadratic SVM reaching 94.4% and 93.0%, respectively. Neural networks, ensembles, k-NN, and discriminants also demonstrate high classification accuracy.

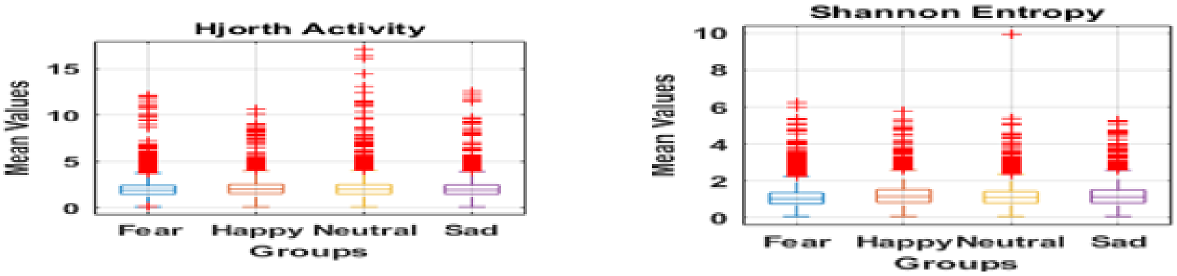
4.3.2.5 Performance Metrics

Model performance is evaluated using three metrics. The confusion matrix (Figure 4.5) (a) shows diagonal values above 91.4%, indicating most true labels are correctly predicted. The ROC curve of the cubic SVM (Figure 4.5) (b) demonstrates strong discrimination, with AUCs



(a) Boxplot of Hjorth Mobility.

(b) Boxplot of Hjorth Complexity.

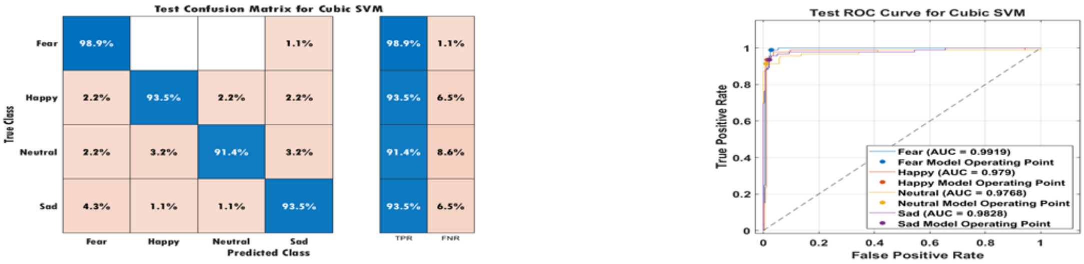


(c) Boxplot of Hjorth Activity.

(d) Boxplot of Shannon Entropy.

Figure 4.4: Kruskal–Wallis test results for different features.

of 0.9919 (fear), 0.979 (joyful), 0.9768 (neutral), and 0.9828 (sad). The precision–recall curve further illustrates performance, highlighting the trade-off between precision and recall across thresholds and its utility for imbalanced datasets. The PR-AUC values of 0.9464 (fear), 0.9062 (joyful), 0.9693 (neutral), and 0.9733 (sad) confirm the model's robustness in classifying all four emotions.



(a) Confusion Matrix

(b) ROC Curve

Figure 4.5: Evaluation of the cubic SVM model using the confusion matrix and ROC curve.

4.3.3 Comparison with Previous Study

Table 4.3 presents key findings from prior studies using the SEED IV dataset. Zheng et al. [31] developed an emotion meter with multimodal data, achieving 85.11% accuracy compared to 72.33% with single-modality data. Wai-Cheong et al.[100] introduced the RODAN deep neural network with 70.28% accuracy, while Ziyu et al.[101] proposed SST-Emotion Net, a two-stream spatial–temporal model, reaching 84.92%. Zhong [102] designed a regularised graph neural network (RGNN) with NodeDAT and EmotionDL modules, obtaining 79.37%. Jiyao [75] proposed four transformer frameworks exploring temporal and spatial features, achieving 83.27%. Sijin et al. [76] developed SOFNN, fusing EEG and eye movement data, which yielded 87.27% accuracy. Other works decomposed EEG into five rhythms via VKF

Table 4.3: Summary of previous studies using the SEED-IV dataset.

Author	**Feature Extraction Method / Deep Neural Model / Machine Learning Classifier**	**Accuracy (%)**
Wei-Long Zheng [31]	Deep Neural Network for Single Modality Data	70.33 ± 14.45, 67.82 ± 18.04
Wei-Long Zheng [31]	Deep Neural Network for Multimodal Data	85.11 ± 11.79
Wai-Cheong Lincoln Lew [100]	Regionally-Operated Domain Adversarial Network (RODAN)	70.28 ± 9.70
Xiyang Cai [101]	SST-Emotion Net	84.92 ± 6.66
Pei Xiang Zhong [102]	Regularized Graph Neural Network (RGNN)	79.37 ± 10.54
Yanxi Zhao [75]	Transformer-based Model	83.27 ± 8.37
Sijin Zhou, Dongmin Huang [76]	Subjective and Objective Feature Fused Neural Network Model (SOFNN)	86.27 ± 10.16
A. K. Dwivedi [103]	VKF Filtering, Hjorth Parameters, and Optimizable kNN	88.00
Proposed Method	Optimized Bandpass Filter, Hjorth Activity, Hjorth Mobility, Hjorth Complexity, Cubic SVM	**94.40**

filtering, identified relevant channels using threshold PCC, and classified statistical features with optimized kNN [103]. In comparison, the proposed method decomposes EEG into delta, theta, alpha, beta, and gamma bands using an optimized band-pass filter, extracts Hjorth Mobility, Hjorth Activity, Hjorth Complexity, and Shannon Entropy, and classifies with cubic SVM, achieving 94.4% accuracy.

4.4 Decoding Human Feelings: A Dynamic Fusion of Adaptive IIR Filtering and HCRNet for Emotion Recognition

4.4.1 Methodology

The proposed method is validated using the DEAP and DREAMER datasets.

4.4.1.1 Particle Swarm Optimization

Particle Swarm Optimization [104] is a nature-inspired, population-based technique modeled on the social behavior of fish schools and bird flocks. It iteratively refines candidate solutions (particles) using both individual and collective experiences. Each particle i has a position $x_i^{(t)}$, velocity $v_i^{(t)}$, personal best p_i (local best), and the swarm's best position g (global best), which are updated according to defined rules [104].

$$v_i^{(t+1)} = wv_i^{(t)} + c_1r_1(p_i - x_i^{(t)}) + c_2r_2(g - x_i^{(t)}) \tag{4.10}$$

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)} \tag{4.11}$$

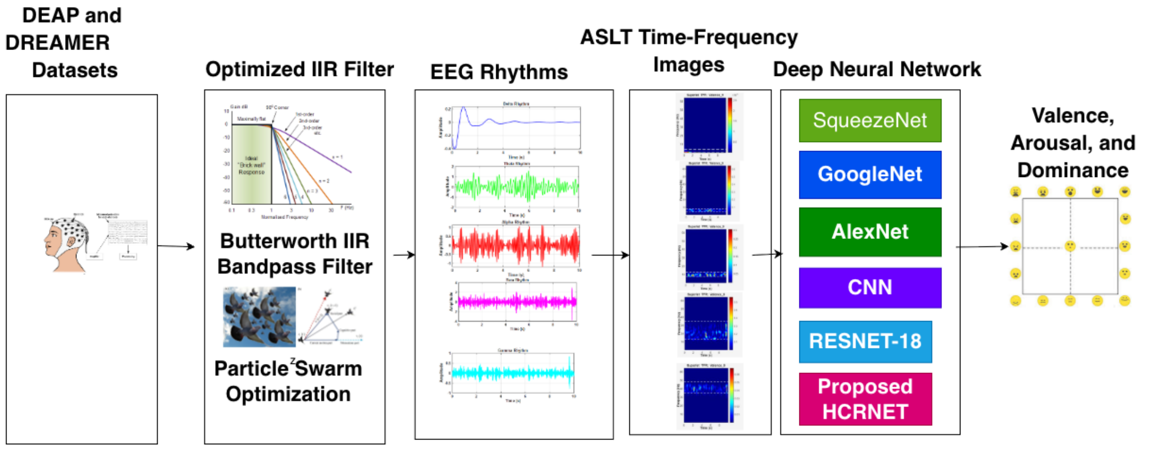


Figure 4.6: Block diagram of the proposed methodology for EEG-based emotion recognition using PSO-optimized IIR filter design and superlet time–frequency analysis with HCRNet.

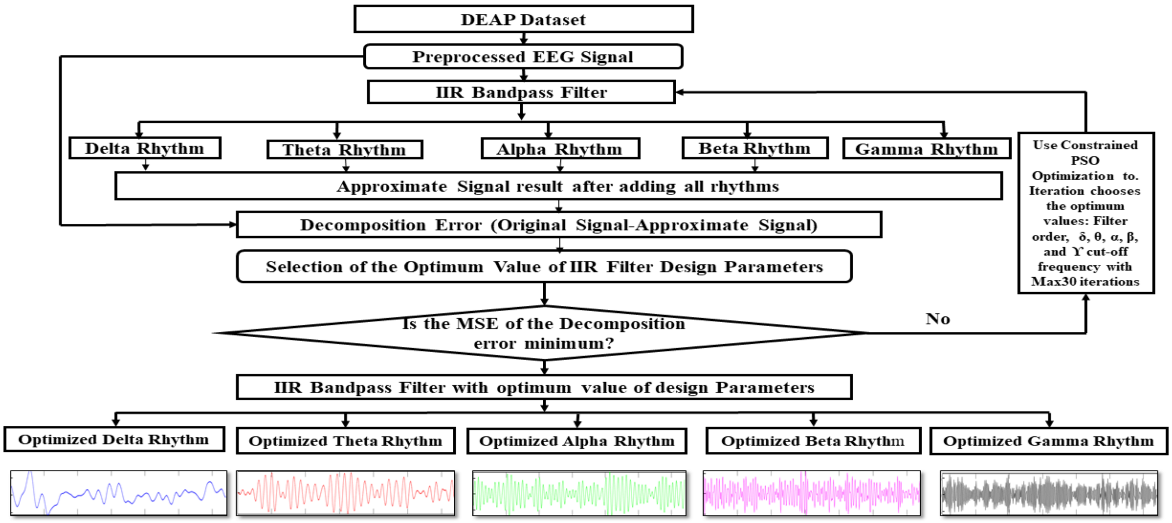


Figure 4.7: Flowchart of the PSO-based methodology for optimizing filter parameters.

Here, v_i and x_i denote the velocity and position of particle i, w is the inertia weight balancing exploration and exploitation, c_1 and c_2 are cognitive and social acceleration coefficients, and r_1 and r_2 are random numbers in $[0, 1]$. PSO efficiently converges to optimal or near-optimal solutions.

4.4.1.2 Particle Swarm Optimization Based IIR Filter Design for EEG Decomposition

EEG signals can be analyzed by decomposing them into frequency bands, but they are often affected by noise. Efficient bandpass filtering reduces this noise. The proposed method employs PSO to optimally design IIR filters for Delta, Theta, Alpha, Beta, and Gamma bands, ensuring valid, non-overlapping frequency ranges and robust filters. PSO optimizes the filter order and cutoff frequencies for each signal to achieve the most accurate reconstruction of the original signal from its filtered bands.

To ensure consistent energy levels, each EEG signal is normalized. Algorithm 4.1 details the stepwise decomposition using the optimized filter design, storing the optimal filter configuration and corresponding MSE for each signal. Zero-phase Butterworth IIR filters extract subbands from signals with the lowest reconstruction error per subject. The reconstructed signal, original EEG, and its five subbands are visualized for evaluation. This adaptive, signal-specific filtering enables precise EEG decomposition. Compared to a standard Butterworth filter (fixed order 4, specified cutoff), the optimized filter showed a 38% reduction in MSE (1.2208 to 0.7554), a 0.92 dB increase in SNR (0.44 to 1.36 dB), and a 70% decrease in computational time (0.501 to 0.153 s) across 32 subjects, demonstrating consistent advantages.

Algorithm 4.1 : PSO-Based IIR Filter Design for EEG Signal Decomposition

```
Require: $D$: Number of subjects; $fs$: Sampling rate; $[lb, ub] \in \mathbb{R}^{11}$; $N$: Swarm size; $T$:
    Max iterations; $dim = 11$; EEG files: $\{\texttt{DS1SUB\_d.mat}\}_{d=1}^{D}$
Ensure: $\mathbf{P}^* \in \mathbb{R}^{11 \times S \times D}$: Optimized filters; $\mathcal{E}^* \in \mathbb{R}^{S \times D}$: Minimum MSEs; $\mathcal{F} \in \mathbb{R}^{T \times S \times D}$:
    Fitness curves
for $d \leftarrow 1$ to $D$ do
    Load $\texttt{EEG\_signals} \leftarrow \texttt{load(DS1SUB\_d)}$, $S \leftarrow \texttt{length(EEG\_signals)}$
    for $s \leftarrow 1$ to $S$ do
        $\mathbf{x} \leftarrow \texttt{EEG\_signals}(s)$; Normalize $\mathbf{x} \leftarrow \mathbf{x}/\|\mathbf{x}\|_2$
        Define $f(\mathbf{p})$: $\texttt{round}(p_1) \rightarrow n$, $[p_2, \ldots, p_{11}] \rightarrow \mathbf{f}_c$
        if any $\mathbf{f}_c \notin (0, fs/2)$ or bands overlap then return $\infty$
        end if
        $\mathbf{x}_{\text{rec}} \leftarrow \mathbf{0}$
        for each EEG band $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$ do
            $[b_l, b_u] \leftarrow \texttt{band\_edges}(b)$
            $[\mathbf{B}, \mathbf{A}] \leftarrow \texttt{butter}(n, [b_l, b_u]/(fs/2), \text{'bandpass'})$
            $\mathbf{x}_b \leftarrow \texttt{filtfilt}(\mathbf{B}, \mathbf{A}, \mathbf{x})$
            $\mathbf{x}_{\text{rec}} \mathrel{+}= \mathbf{x}_b$
        end for
        Normalize $\mathbf{x}_{\text{rec}}$: $\mathbf{x}_{\text{rec}} \leftarrow \mathbf{x}_{\text{rec}} \cdot \|\mathbf{x}\|_2/\|\mathbf{x}_{\text{rec}}\|_2$
        $\texttt{MSE}(\mathbf{x}, \mathbf{x}_{\text{rec}}) \rightarrow$ fitness
        $(\mathbf{p}^*, \mathcal{E}^*_{s,d}, \mathcal{F}_{:,s,d}) \leftarrow \texttt{PSO}(f, N, T, dim, lb, ub)$
        $\mathbf{P}^*_{:,s,d} \leftarrow \mathbf{p}^*$
    end for
end for
for $d \leftarrow 1$ to $D$ do
    $s^* \leftarrow \arg\min_s \mathcal{E}^*_{s,d}$; $\mathbf{p}^* \leftarrow \mathbf{P}^*_{:,s^*,d}$
    Design filters with $\mathbf{p}^*$, decompose $\mathbf{x}_{s^*}$, and plot results
end for
```

4.4.1.3 Adaptive Superlet Transform

The superlet transform (SLT) [105] is a modified Morlet transform that enhances time-frequency resolution (TFR) compared to STFT, CWT, and AWT. It achieves this improvement using a set of constrained-bandwidth wavelets. The SLT is chosen for its superior time-frequency resolution and robustness to noise, whereas CWT [106] is less accurate and causes signal smearing. By adaptively combining wavelets across scales, SLT enhances EEG feature extraction for applications such as depression diagnosis [107] and emotion recognition [108],

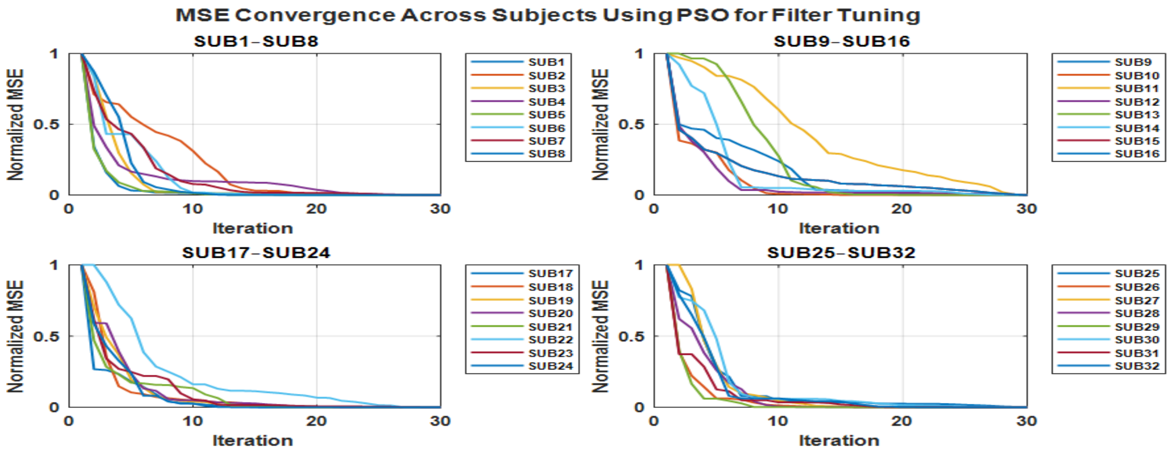


Figure 4.8: Convergence curve for minimizing MSE during filter parameter optimization using PSO on the DEAP dataset.

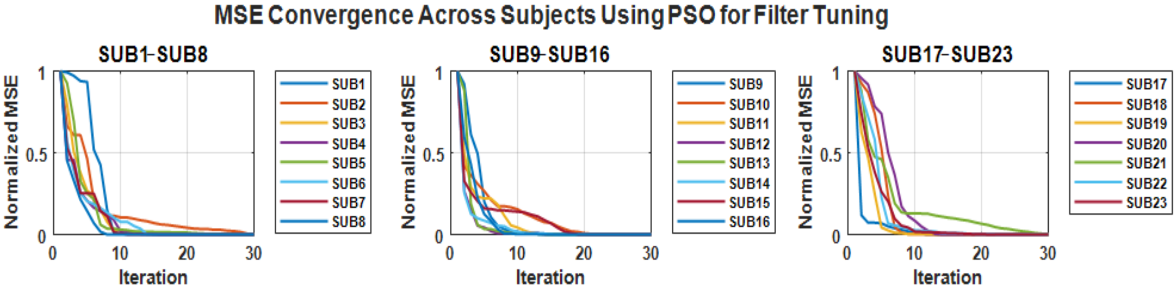


Figure 4.9: Convergence curve for minimizing MSE during filter parameter optimization using PSO on the DREAMERS dataset.

overcoming the time–frequency localization trade-off present in SSWT [107].

SLT improves the analysis of transient signals in noisy EEG data, while SSWT enhances CWT through time-frequency reassignment.

STFT is limited by fixed windows, and AWT's adaptive wavelets often reduce frequency resolution. Although CWT provides variable resolution, its high computational cost and noise sensitivity limit detection of abrupt signal changes. The Morlet wavelet is defined as [105]

$$\psi_{\text{fcc,Cy}}(t) = \frac{1}{B_w\sqrt{2\pi}} e^{-\frac{t^2}{2B_w^2}} e^{-j2\pi f_{cc}t} \tag{4.12}$$

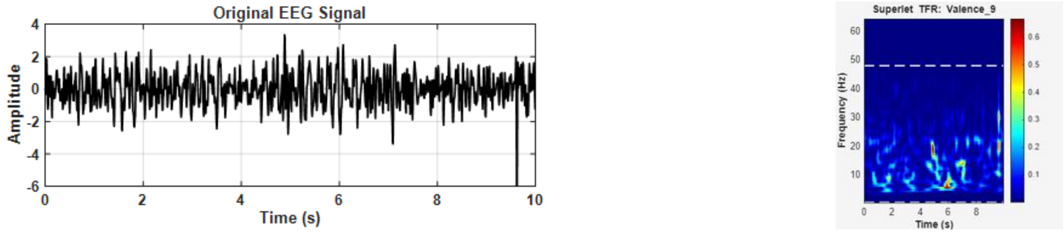
The Morlet wavelet function is defined by its central frequency (f_{cc}) and the number of cycles (C_y) it contains. Its bandwidth (B_w) corresponds to the width of the Gaussian window.

$$B_w = \frac{C_y}{5 f_{cc}} \tag{4.13}$$

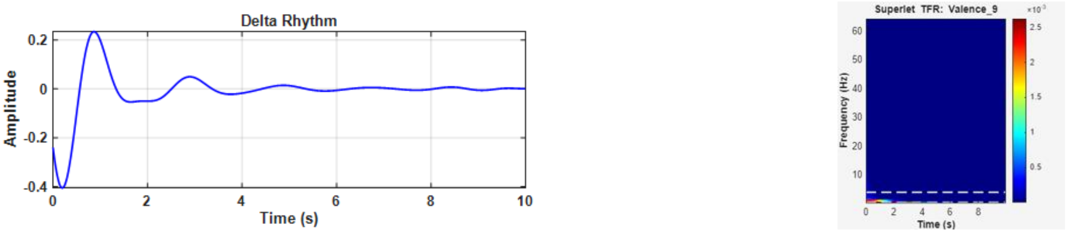
Equation (4.13) relates the number of cycles C_y to bandwidth B_w, showing that increasing f_{cc} reduces frequency resolution when C_y is constant.

Using multiple Morlet wavelet cycles improves frequency resolution but lowers temporal resolution. The superlet enhances both resolutions by combining Morlet wavelets with a common central frequency f_{cc} but varying numbers of cycles. Its mathematical formulation is given in [105]

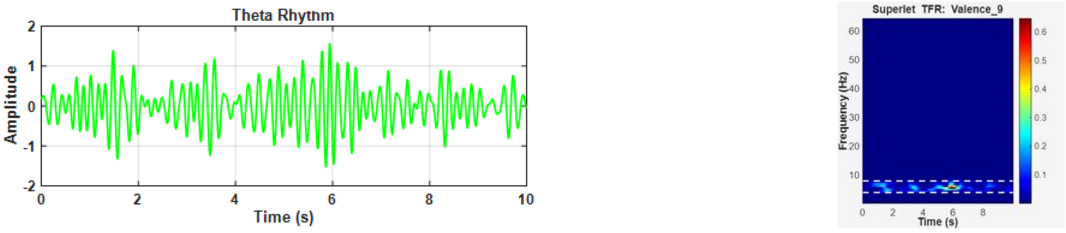
$$\text{SLT}_{f_c,M} = \left\{\psi_{f_c,C_i} \;\middle|\; C_i \in \{C_1, C_2, \ldots, C_M\}\right\} \tag{4.14}$$



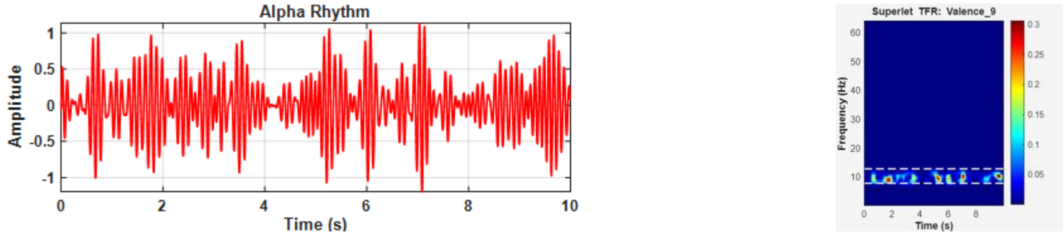
(a) Original 1D Signal (b) Time-Frequency Representation



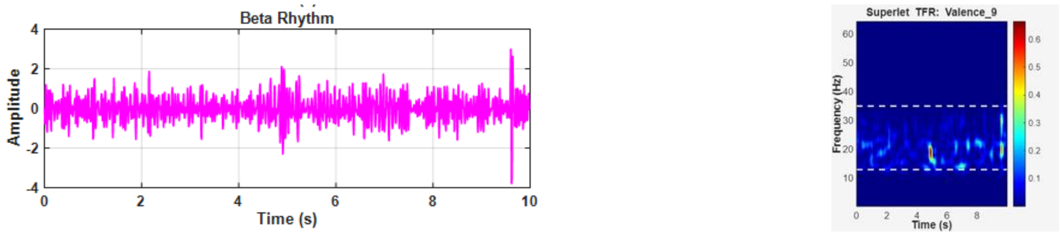
(c) δ 1D Signal (d) δ Time-Frequency Representation



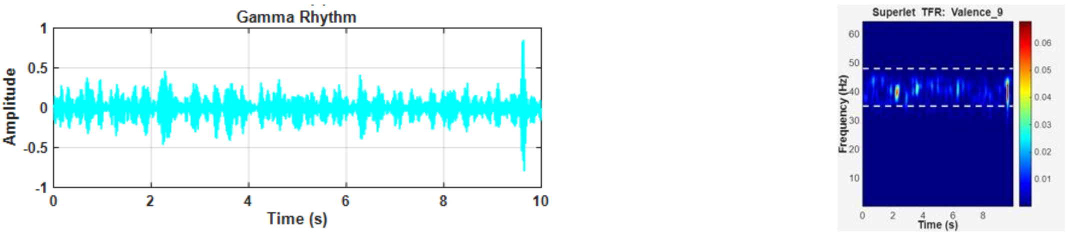
(e) θ 1D Signal (f) θ Time-Frequency Representation



(g) α 1D Signal (h) α Time-Frequency Representation



(i) β 1D Signal (j) β Time-Frequency Representation



(k) γ 1D Signal (l) γ Time-Frequency Representation

Figure 4.10: Comparison of 1D EEG signals (left) and corresponding adaptive superlet TFRs (right) across EEG rhythms.

The Adaptive Superlet Transform is defined as [105]

$$\begin{aligned} \mathrm{ASLT}_{f_0} &= \mathrm{SLT}_{f_0,w} \\ w &= w_{\min} + (w_{\max} - w_{\min}) \frac{f_0 - f_{0,\min}}{f_{0,\max} - f_{0,\min}} \end{aligned} \tag{4.15}$$

Where, $w_{\max}$ and $w_{\min}$ represent the maximum and minimum values of the order, which are modulated according to the maximum and minimum central frequency $f_{0\max}$ and $f_{0\min}$ respectively. The superlet transform (SLT) is defined by its order M, which denotes the number of wavelets, with the cycles determined either additively or multiplicatively. The overall response is obtained as the geometric mean of the individual wavelets. Since the time-frequency resolution depends on the central frequency f_{cc}, the order M must vary accordingly. The adaptive superlet transform addresses this by continuously adjusting M with f_{cc}, thereby enabling either constant bandwidth or improved resolution. This frequency-controlled adaptation allows ASLT to precisely balance time-frequency resolution in accordance with the signal's behavior.

4.4.2 Experiment

This section presents a detailed description of the methodology adopted to obtain the experimental results, outlining the procedural steps, empirical parameter settings, and an in-depth analysis of the outcomes and their implications.

4.4.2.1 Preprocessing

The study employs publicly available DEAP and DREAMER datasets containing physiological and multi-channel EEG data from 40 trials. EEG signals are extracted from each channel and labeled based on the dominant emotion (valence, arousal, or dominance), determined by the highest score on a 1–9 emotional scale. Data and labels are organized in a structured format, with EEG signals segmented into 10-second windows using 50% overlap. Z-score normalization standardizes each segment's amplitude, and each segment is assigned a label corresponding to its dominant emotion.

4.4.2.2 EEG Signal Decomposition into Rhythms using Optimized Filter Design

EEG signals are decomposed using an IIR bandpass filter whose parameters are optimized with the PSO algorithm. As shown in Figures 4.8 and 4.9, optimization significantly reduces MSE. The optimization process is implemented using Algorithm 4.1, with PSO parameters listed in Table 4.5. The resulting subject-specific optimized IIR filter parameters are provided in Table 4.4, and are used to decompose each EEG signal into delta, theta, alpha, beta, and gamma rhythms.

4.4.2.3 Signal to Image Conversion

The proposed method converts one-dimensional EEG rhythms into two-dimensional time-frequency images using the adaptive superlet transform, which enhances resolution across frequency bands. By employing a range of frequencies and adaptive cycles, ASLT effectively captures both low- and high-frequency components. The resulting representations are

Table 4.4: Optimized filter parameters and corresponding MSE for each subject in the DEAP and DREAMER datasets.

S. No.	Subject	Order	Delta (Hz)		Theta (Hz)		Alpha (Hz)		Beta (Hz)		Gamma (Hz)		MSE
			L	H	L	H	L	H	L	H	L	H	After Optz
1	SUB1	2.5051	0.5	1	4.0001	7.4997	8	11.7951	12.2987	34.9980	35.4989	55	5.05E-05
2	SUB2	2	0.5	3.4995	4	8	8.5012	12	12.5215	35	36.1310	55	9.19E-05
3	SUB3	2	0.6214	3.4976	4	8	8.5001	11.4998	12	34.4739	35	46.8932	9.43E-05
4	SUB4	2	1.2325	3.4983	4	7.5520	8.1120	11.8809	12.6303	34.3837	35	55	1.06E-04
5	SUB5	2.1972	1.6403	3.4929	4	7.4634	8.0026	11.5128	12.0205	34.9815	35.4851	55	5.38E-05
6	SUB6	2	0.9889	1	4	7.4988	8	12	12.501	35	35.50268	55	0.00015
7	SUB7	2	1.5142	3.4937	4	7.5633	8.0639	11.496	12	33.5262	35.00819	54.9903	0.00019
8	SUB8	3.0345	0.5	3.4883	4	7.4999	8	11.8781	12.3814	35	35.50001	52.3809	5.11E-05
9	SUB9	2	0.5	3.4994	4	8	8.5147	12	12.5233	35	35.595	55	0.00017
10	SUB10	2	1.3761	3.4972	4.000023	7.4996	8	11.40001	12	35	35.5104	54.7967	6.52E-05
11	SUB11	2	0.5013	3.4028	4.00109	7.4999	8	11.4975	12.0029	35	35.5008	48.7624	0.00012
12	SUB12	2.0749	0.5147	3.9202	4.4227	7.4999	8.5754	11.9657	12.4662	33.8794	35.1774	53.0726	0.00015
13	SUB13	3.1212	0.8801	1.6597	4.00023	7.5103	8.01036	11.8129	12.3135	34.4160	35.2278	53.6464	9.95E-05
14	SUB14	2	1.03454	3.4746	4	7.4982	8	12	12.7219	35	36.2166	53.2220	0.000102
15	SUB15	2	3	4	4.5043	7.4944	8	11.4752	12	34.4343	35	55	2.43E-05
16	SUB16	2	0.5	3.3938	4	8	8.5047	12	12.5408	35	35.7874	55	0.0001175
17	SUB17	2	0.5	3.3207	4	8	8.50008	12	12.5219	35	35.5013	55	1.95E-05
18	SUB18	2.4302	0.5942	3.4795	4	8	8.5056	11.979	12.4826	34.3920	35	52.4324	7.34E-05
19	SUB19	2	0.7958	3.7413	4.2416	7.8881	8.3882	12	13.7998	34.98806	35.5723	55	7.63E-05
20	SUB20	2	0.5	3.4994	4	8	8.5003	12	12.5327	35	36.4746	55	0.000157
21	SUB21	4.4357	2.6933	2.70108	4.0023	7.4580	8.00170	11.3916	12	34.4910	35	47.5301	2.17E-05
22	SUB22	2	1.7609	3.48702	4	7.4911	8	12	13.33001	35	35.5068	49.4109	0.000122
23	SUB23	2.4921	2.0819	3.9978	4.4988	7.4969	8.00022	11.488	12.0001	34.9987	35.9120	53.83225	4.53E-05
24	SUB24	4.8401	2.6447	3.3934	4	7.4998	8	11.4998	12	34.4989	35	55	4.31E-05
25	SUB25	2	0.5013	3.4970	4	7.4942	8	12	12.9260	34.4730	35	53.40989	0.000112
26	SUB26	2.5017	2.6237	2.7936	4	8	8.50014	12	12.5008	35	35.6753	55	3.30E-05
27	SUB27	2	0.5	3.4980	4	8	8.5016	11.4035	12	33.6127	35	55	8.21E-05
28	SUB28	2	0.9431	3.2132	4.0265	8	8.5005	12	12.6396	35	35.5044	55	0.000126
29	SUB29	4.5861	2.2650	3.34062	4	7.9987	8.4988	11.8705	12.3705	35	36.0496	55	5.94E-05
30	SUB30	2.0107	0.7345	3.5213	4.0226	7.9905	8.4907	11.9743	12.9294	34.8594	36.6140	54.7326	0.000127
31	SUB31	2	0.5	1	4	8	8.5011	11.4998	12	34.4896	35	55	0.000102
32	SUB32	3.112	0.5	2.87964	4.01207	7.4998	8	11.5539	12.0544	34.9429	35.70069	50.5193	0.000109
DREAMER Dataset													
33	SUB1	2.4837	0.5	2.1352	4.5287	7.0901	10.3251	11.0135	23.4256	34	35	48	8.31E-05
34	SUB2	2	0.5	3.4826	4.2248	7.7409	8.2496	12	12.5559	34	37.1145	43.2519	3.28E-04
35	SUB3	2	0.5	3.1200	4	8	8.5635	12	13.2365	34	35	48	1.37E-04
36	SUB4	3.0457	3	4	5	5.0340	8.5305	10.4499	21.1971	34	35	48	4.36E-05
37	SUB5	4.5196	1.7549	4	7	7.0286	9.1916	12	19.7832	34	45.6938	48	4.77E-05
38	SUB6	2	0.5	4	4.5005	8	8.5003	12	12.7415	34	35	48	1.17E-04
39	SUB7	2.1504	0.5	3.3832	4	7.2634	8	11.3215	12	34	35	48	9.10E-04
40	SUB8	2	0.6843	4	4.5190	8	8.7752	11.0263	12	34	35	48	8.57E-05
41	SUB9	2.8929	0.5689	3.8815	4.3822	7.4051	8	11.4928	12	34	35	45.4976	9.07E-05
42	SUB10	2	0.5801	3.4011	4	7.4197	8	12	12.5844	34	35	48	9.59E-05
43	SUB11	2.1049	0.5	3.4994	4	7.4788	8.0104	11.4995	12	34	35	48	7.68E-05
44	SUB12	5.1847	0.5	4	4.5000	7.9184	8.4187	9	12	34	35	48	4.71E-04
45	SUB13	3.8727	3	3.0486	6.7103	8	11	10.9555	19.3780	34	35	47.3330	3.01E-05
46	SUB14	2.0095	0.5004	3.5010	4.0013	7.4997	8	11.4998	12	34	48	47.9957	2.36E-04
47	SUB15	2.8561	0.5	1	4.9964	5	8	12	21.0982	34	35	48	5.73E-05
48	SUB16	2	0.5	3.4977	4	8	8.5001	11.4997	12	27.1962	35	48	1.82E-04
49	SUB17	2.4288	0.5	3.4522	4	7.5449	8.0449	11.4992	12	21.7264	40.9375	45.7459	8.14E-05
50	SUB18	2	0.5	4	4.5000	8	8.7327	12	15.9151	34	35	48	3.88E-04
51	SUB19	2	1.0491	4	4.5003	8	8.5531	11.4987	12	20.9141	40.6190	48	1.35E-04
52	SUB20	2.4290	0.5	3.4999	4	7.3850	8	11.4994	12	34	35	48	4.99E-04
53	SUB21	2	0.5	4	5.6666	7.7310	8.2338	12	13.0625	34	35	48	1.37E-04
54	SUB22	2	0.5	3.4962	4	8	8.5057	12	12.8453	34	35	43.6200	2.24E-04
55	SUB23	2	0.5	4	4.5020	8	8.5044	11.4801	12	34	35	48	1.49E-04

normalized, mapped into RGB images to emphasize contrast and structure, and resized to 224×224. Time-frequency resolution is balanced using seven cycles with orders from 1 to 7. Figure 4.10 illustrates EEG rhythms and their corresponding 2D images. These images are then used by deep-learning algorithms to extract complex temporal and spatial patterns.

Table 4.5: Filter design parameters and their optimization ranges.

Parameter	Defined Value
Filter parameters (Order plus Band Edges)	11
Maximum iterations (T)	30
Sampling frequency	128 Hz
Order limits	[2, 6]
Delta band limits	[0.5, 1] to [3, 4]
Theta band limits	[4, 5] to [7, 8]
Alpha band limits	[8, 9] to [11, 12]
Beta band limits	[12, 13] to [34, 35]
Gamma band limits	[35, 36] to [48, 55]

Table 4.6: Training hyperparameters used for deep neural network training.

Parameter	Value
Optimizer	Adam
Initial Learning Rate	0.001
Learning Rate Schedule	Piecewise
Learning Rate Drop Factor	0.1
Learning Rate Drop Period	10 epochs
Max Epochs	20
Mini-Batch Size	64
Validation Data	Validation Images
Validation Frequency	Every 5 iterations
Shuffle	Every epoch
Plots	Training progress
Verbose	False
Execution Environment	GPU
Validation Patience	50

4.4.2.4 Classification using Deep Neural Network Algorithms

A deep neural network (DNN) consists of multiple hidden layers that progressively learn complex features through weights and activation functions, making it highly effective for emotion recognition. Several DNN architectures including GoogleNet, AlexNet, SqueezeNet, CNN, ResNet-18, and HCRNet are evaluated on ASLT images. The dataset is split into 70% for training, 15% for validation, and 15% for testing. The hyperparameter settings for each model are provided in Table 4.6.

(a) AlexNet:

AlexNet [109] is an eight-layer network comprising five convolutional and three fully connected layers, originally trained on ImageNet. It extracts hierarchical image features and accepts inputs of size 227×227×3, with images resized using augmentedImageDatastore. For this study, its final three layers (fully connected, softmax, and classification output) are modified to classify the three dataset classes.

(b) GoogleNet:

GoogleNet [58] is a 22-layer convolutional neural network pre-trained on ImageNet, valued for its efficient architecture and feature extraction. For transfer learning, its final classification layers are replaced with fully connected, softmax, and classification layers. Using replaceLayer, the classification layers are customized for emotion classes, while pre-trained feature extraction layers are retained to leverage prior knowledge.

(c) SqueezeNet:

SqueezeNet [110] is a lightweight convolutional neural network optimized for fast inference and low memory usage, with an input size of 227×227×3. It employs fire modules consisting of a squeeze layer (1×1 convolutions) followed by expand layers (1×1 and 3×3 convolutions), reducing parameters while maintaining performance. For this study, the final convolutional layer (conv10) is replaced with a new 1×1 classification layer.

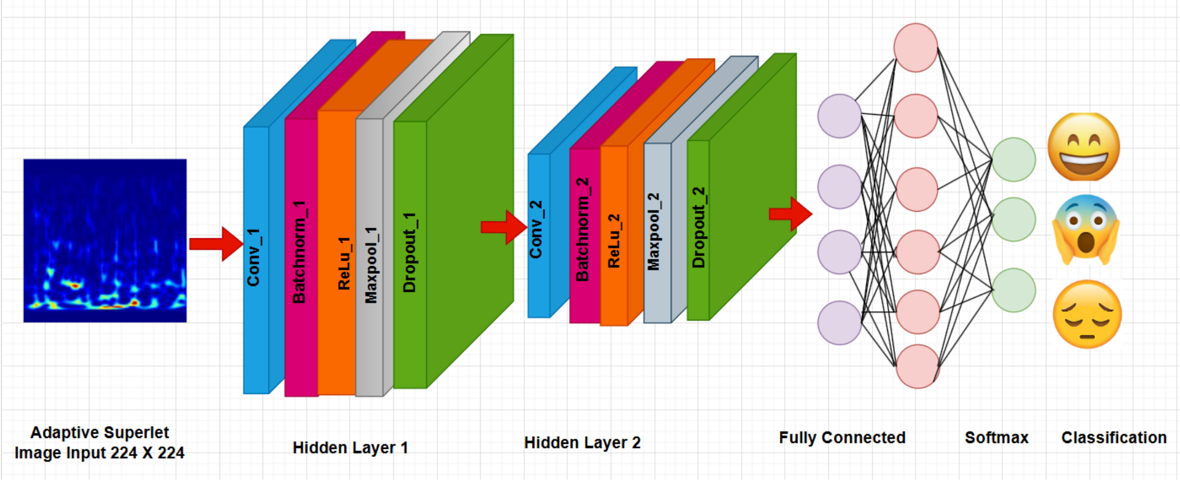


Figure 4.11: Convolutional neural network architecture.

(d) Convolutional Neural Network:

The CNN [111] classifies three emotional classes from 224×224 RGB images (Fig. 4.11). It has two convolutional blocks, each with a 2D convolution layer (8 or 16 filters), batch normalization, max pooling, ReLU activation, and 20% dropout to prevent overfitting. Extracted features are fed to a fully connected layer with three outputs, followed by a classification layer that computes loss and accuracy, effectively capturing and enhancing spatial characteristics.

(e) ResNet-18:

ResNet-18 [112] is an 18-layer convolutional neural network comprising convolutional layers, batch normalization, ReLU activation, pooling, and fully connected layers. It employs a residual learning approach, using identity shortcut connections to add the input directly to the output, effectively learning the difference between input and desired output.

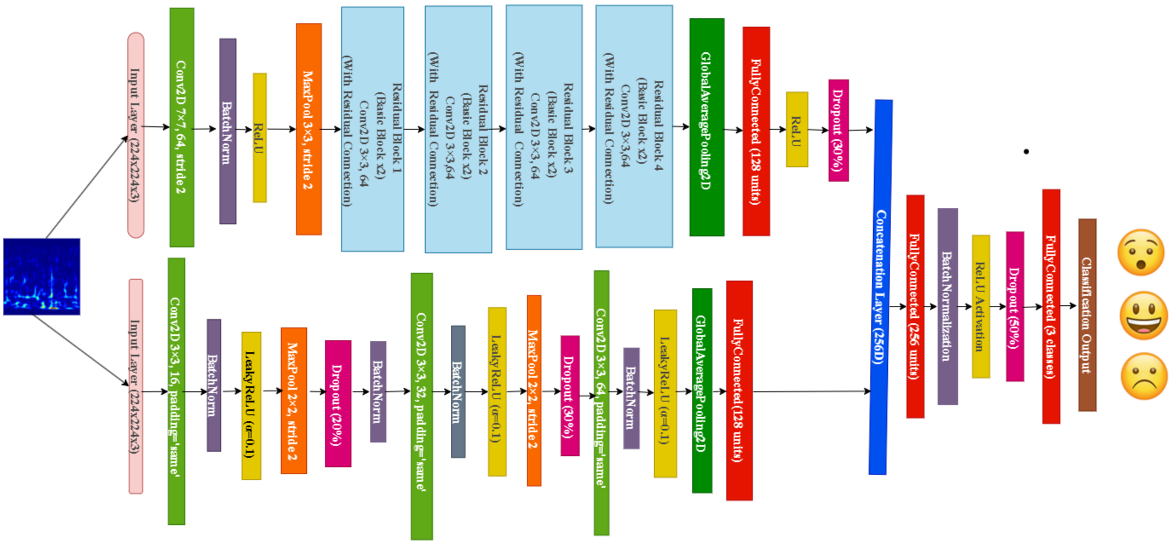


Figure 4.12: HCRNet architecture.

(f) Proposed HCRNet:

HCRNet (Hybrid CNN-ResNet18, Fig. 4.12) combines ResNet-18 and a CNN to enhance feature extraction and classification of EEG time-frequency images. It employs a dual-branch design: the CNN branch uses hierarchical convolutional layers with Leaky-ReLU and progressive dropout for domain-specific EEG features, while the ResNet-18 branch leverages transfer learning and residual connections for general visual patterns. Both branches produce 128-dimensional features, concatenated into a 256-dimensional representation. This hybrid design enables multi-scale feature learning, improved generalization, noise robustness, and faster training, providing an effective solution for EEG time-frequency classification.

The proposed HCRNet combines a custom CNN branch for spatial pattern extraction with a pre-trained ResNet-18 for semantic feature extraction. Features from both branches are merged and passed to a final classification layer. This dual-path design leverages transfer learning and custom feature learning to enhance classification accuracy. A detailed step-by-step description is provided in Algorithm 4.2.

4.4.3 Discussion

4.4.3.1 Subject-Wise Performance Metrics of Deep Neural Networks

Figures 4.13 and 4.14 show subject-dependent rhythm accuracy of various DNNs evaluated on 15% of images excluded from training and validation. The dataset contains 14,080 images across three classes, with an average distribution of 5,456 (valence), 4,136 (arousal), and 4,488 (dominance) images per class. DNN training uses 9,856 images, while the remaining 4,224 images are reserved for testing and validation.

(a) AlexNet Results:

Figure 5.17 presents AlexNet's subject-wise classification accuracies for EEG-based emotion recognition across combined, delta, theta, alpha, beta, and gamma rhythms. Combined rhythms yield the highest performance, with an average accuracy of 74.49% and a standard deviation of 16.37. Alpha and beta rhythms outperform delta, theta, and gamma. Subjects 27,

Algorithm 4.2 : Structured Pseudocode for HCRNet Framework

```
Require: Dataset $D$ with $C$ classes
Ensure: Model $M$, Test accuracy $A$
Initialize parameters $\theta$
Split $D \rightarrow D_{\text{train}}, D_{\text{test}}$ (70/30)
ResNet Branch
Load ResNet18, remove FC, add FC 512→128
CNN Branch
for $l = 1:3$ do
    Conv2D (3×3, $f_l$), BatchNorm, LeakyReLU
    MaxPool (2×2), Dropout $p = 0.1l + 0.1$
end for
GlobalAvgPool, FC 128
Fusion
Concatenate: $\mathbf{h} = [\mathbf{h}_R; \mathbf{h}_C]$
FC 256→$C$ + Softmax
Training
for epoch = 1 : 5 do
    for batch $(X, y)$ in $D_{\text{train}}$ do
        $\hat{y} \leftarrow M(X)$
        $\mathcal{L} \leftarrow \text{CE}(\hat{y}, y)$
        $\theta \leftarrow \theta - \eta \nabla \mathcal{L}$
    end for
    Validate every 30 batches
end for
$A \leftarrow \text{Accuracy}(M(D_{\text{test}}))$
```

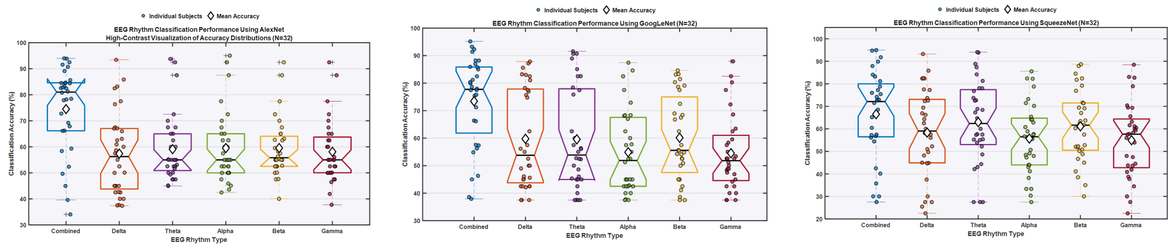
32, and 3 consistently exceed 70% accuracy, while subjects 31 and 1 show lower performance (Fig. 4.14). Gamma and delta rhythms perform better across more subjects, though notable inter-subject variability reflects distinct EEG characteristics.

(b) GoogleNet Results:

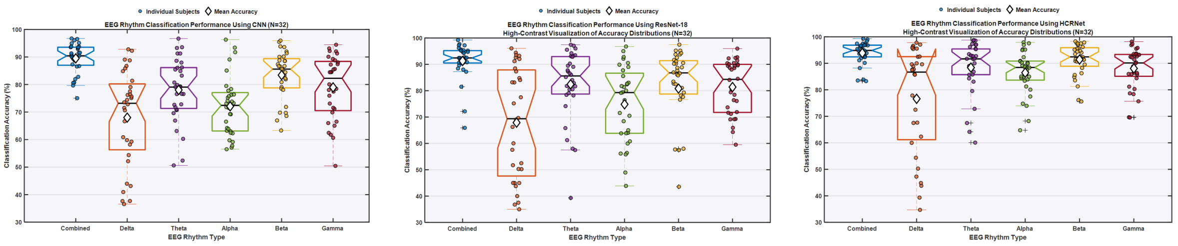
GoogleNet exhibits promising but inconsistent classification results, achieving an average accuracy across rhythms of 60.39%, higher than AlexNet on average. Subjects 27 and 32 reach the highest accuracies (over 82%), while subjects 1 and 31 perform lowest (37%) (Fig. 4.14). Individual frequency bands generally underperform compared to combined rhythms, with beta and delta slightly better than others (Fig. 5.18). Some subjects (25, 22, and 12) achieve exceptional accuracy across multiple rhythms. The high standard deviation (12.84%) reflects substantial inter-subject variability.

(c) SqueezeNet Results:

SqueezeNet achieved moderate performance, slightly less consistent than GoogleNet, with an average accuracy of 58.62% (Fig. 4.14). Combined rhythms perform best (66.51%), while



(a) AlexNet (b) GoogleNet (c) SqueezeNet



(d) CNN (e) ResNet-18 (f) HCRNet

Figure 4.13: Subject-wise EEG rhythm classification performance using different deep neural networks.

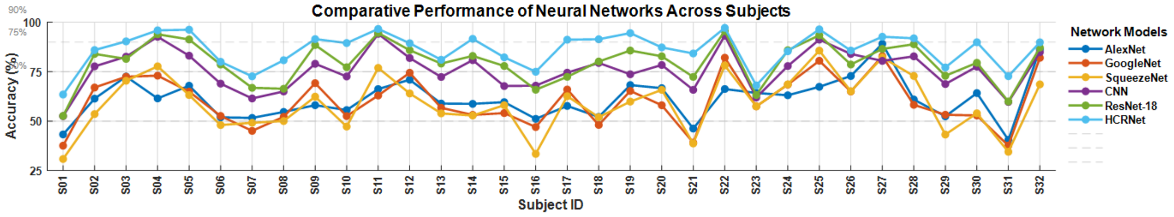


Figure 4.14: Comparative analysis of subject-wise mean EEG rhythm classification accuracy using different deep neural network architectures.

theta rhythms show relatively high accuracy (62.99%), followed by beta and delta (Fig. 5.19). A standard deviation of 14.37% indicates substantial inter-subject variability and limited generalization.

(d) CNN Results:

CNN outperformed AlexNet, SqueezeNet, and GoogleNet, achieving an average accuracy of 76.10%. Combined rhythms showed the highest performance (89.51%), followed by beta (83.42%) and gamma (78.97%) bands (Fig. 5.20). Accuracy was lowest for subjects 1 (52.55%) and 31 (59.72%), while subjects 22 (93.24%) and 25 (91.24%) achieved exceptional results (Fig. 4.14). A standard deviation of 10.23% indicates greater consistency, making CNN a reliable model for EEG-based emotion recognition across subjects and rhythms.

(e) ResNet-18 Results:

ResNet-18 outperformed AlexNet, SqueezeNet, GoogleNet, and CNN, achieving higher accuracy and consistency across all rhythms. Combined rhythms yielded the highest average accuracy of 91.30%, highlighting the benefit of multi-scale feature integration. Among individual bands, theta (82.43%) and gamma (81.39%) achieved the highest accuracies, while

Table 4.7: Mean classification accuracy (%) across 32 subjects in the DEAP dataset using different EEG rhythms.

Model	**Delta (δ)**	**Theta (θ)**	**Alpha (α)**	**Beta (β)**	**Gamma (γ)**	**Combined ($\delta + \theta + \alpha + \beta + \gamma$)**
AlexNet	57.40 ± 15.88	59.08 ± 12.78	59.51 ± 13.47	59.54 ± 11.14	58.12 ± 12.06	74.49 ± 16.37
GoogleNet	59.77 ± 17.88	59.57 ± 18.43	54.88 ± 14.96	60.15 ± 15.28	54.52 ± 13.71	73.43 ± 16.27
SqueezeNet	58.34 ± 19.44	62.99 ± 18.53	55.69 ± 15.67	61.11 ± 15.68	54.95 ± 16.58	66.51 ± 19.39
CNN	66.99 ± 5.40	78.03 ± 11.37	72.08 ± 10.78	83.42 ± 8.74	78.97 ± 11.57	89.51 ± 5.40
ResNet-18	67.87 ± 21.02	82.42 ± 13.40	74.84 ± 14.40	80.81 ± 19.32	83.46 ± 09.70	91.30 ± 4.32
Proposed HCRNet	**74.11 ± 19.67**	**88.39 ± 9.92**	**86.45 ± 8.04**	**91.43 ± 6.05**	**88.08 ± 7.58**	**93.89 ± 4.45**

delta showed the lowest average (67.88%) and highest standard deviation (21.02%), indicating high inter-subject variability. Beta and alpha bands showed moderate performance at 80.81% and 74.84%, respectively.

(f) Results HCRNet:

Table 4.7 compares EEG-based emotion recognition performance of AlexNet, GoogleNet, SqueezeNet, CNN, ResNet-18, and HCRNet across five frequency bands. HCRNet achieved the highest accuracy, with combined rhythms reaching 93.89% across 32 individuals. Among individual bands, beta (91.43%) and theta (88.39%) performed best, while delta showed the lowest accuracy (74.11%) and highest variability (19.67%). Alpha (86.45%) and gamma (88.08%) performed competitively. The combined rhythms also exhibited a low standard deviation (4.25%), indicating consistent and stable performance across subjects.

Table 4.8: HCRNet performance metrics for EEG rhythms on the DEAP dataset.

Rhythm	**Prec.**	**Rec.**	**F1**	**Kappa**	**MCC**	**AUC**
δ	66.51	67.82	65.17	0.5267	0.5312	0.7493
θ	90.90	89.94	90.35	0.8558	0.8535	0.9239
α	85.68	84.68	84.76	0.7660	0.7664	0.8780
β	92.20	91.12	91.61	0.8698	0.8708	0.9318
γ	88.12	86.35	87.12	0.8034	0.8035	0.8962
$\delta + \theta + \alpha + \beta + \gamma$	**93.90**	**92.80**	**93.30**	**0.8953**	**0.8963**	**0.9444**

4.4.3.2 Ablation Study

Tables 4.8 and 4.9 show that the ablation study improves performance across both datasets. For DEAP [29], combined rhythms (δ+θ+α+β+γ) achieve 93.9% precision, 92.8% recall, and 0.9444 AUC, confirming the method's effectiveness. On DREAMER [30], combined rhythms reach 97.11% accuracy, 97.08% F1-score, and 0.976 AUC, demonstrating reliability across datasets. Beta rhythms consistently show high individual performance (92.20% F1 on DEAP, 96.68% on DREAMER), highlighting their importance. Multi-rhythm fusion improves F1 scores by 3.7–5.9% over individual rhythms, confirming the effectiveness of optimized signal processing and hybrid architecture across EEG-based affective computing tasks.

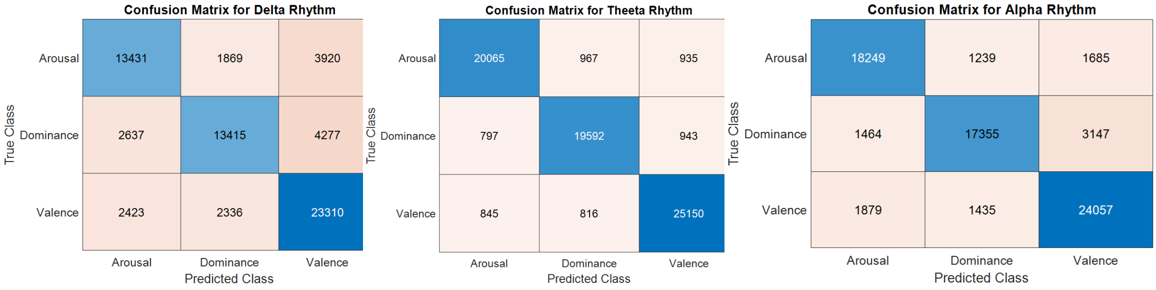
4.4.3.3 HCRNet Performance Parameters

Figure 5.24b shows subject-wise emotion classification accuracy across EEG rhythms (Delta, Theta, Alpha, Beta, Gamma) for valence, arousal, and dominance. Beta, theta, and gamma rhythms achieved the highest accuracies, alpha showed moderate performance, and delta was

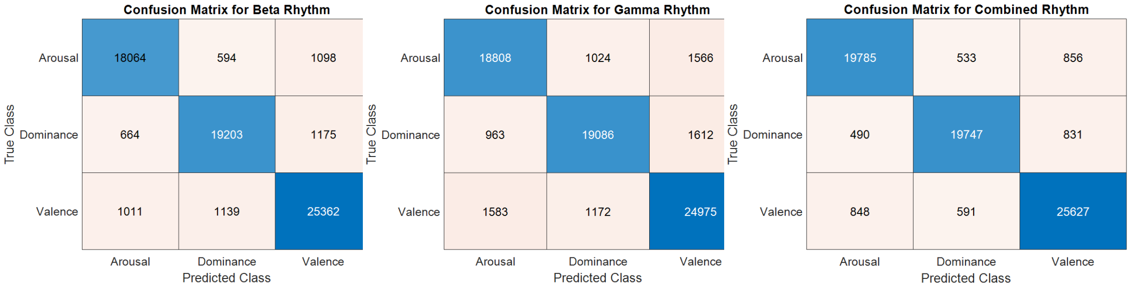
Table 4.9: HCRNet Performance Metrics for EEG Rhythms on the DREAMER Dataset.

Rhythm	**Acc.**	**Prec.**	**Rec.**	**F1**	**Kappa**	**MCC**	**AUC**
δ	73.50	71.69	69.14	69.92	0.624	0.560	0.772
θ	95.91	96.59	95.65	95.87	0.940	0.938	0.967
α	96.69	97.26	94.74	95.70	0.955	0.945	0.963
β	96.71	96.95	96.62	96.68	0.951	0.948	0.974
γ	91.97	93.06	91.80	91.94	0.881	0.882	0.938
δ+θ+α+β+γ	**97.11**	**97.53**	**96.87**	**97.08**	**0.955**	**0.956**	**0.976**

generally lower, especially for valence and arousal. Combined rhythms yielded promising results, with most participants exceeding 90% accuracy. Figure 4.15 indicates valence classification outperforms arousal and dominance. Table 4.8 reports HCRNet performance: beta (92.20% precision, 91.12% recall, 91.61% F1) and theta (90.90%, 89.94%, 90.35%) performed best, alpha and gamma were strong (F1: 84.76%, 87.12%), and delta lagged (F1: 65.17%, precision: 66.51%). MCC and kappa peaked at 0.87 (beta) and 0.86 (theta). AUC ranged from 0.75 (delta) to 0.93 (beta). The combined rhythms ($\delta + \theta + \alpha + \beta + \gamma$) achieved the best overall performance: 93.90% precision, 92.80% recall, 93.30% F1, 0.8963 kappa, and 0.9444 AUC, highlighting the benefit of multi-rhythm fusion.



(a) δ Rhythm (b) θ Rhythm (c) α Rhythm



(d) β Rhythm (e) γ Rhythm (f) Combined (δ+θ+α+β+γ)

Figure 4.15: Confusion matrices for individual and combined EEG rhythms using HCRNet on the DEAP dataset.

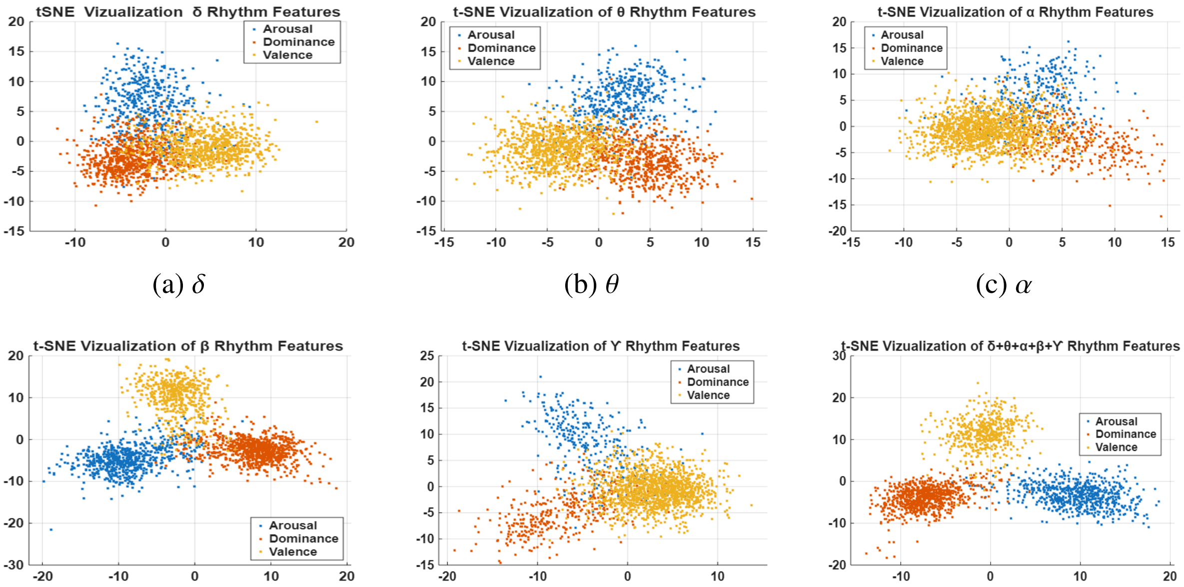
(a) Confusion Matrix

Figure 4.15 presents three-class emotion recognition results across six EEG rhythm bands (δ, θ, α, β, γ, and combined $\delta + \theta + \alpha + \beta + \gamma$) using confusion matrices. True positives are on the diagonals ($C11$, $C22$, $C33$), and misclassifications are off-diagonal. Delta shows moderate accuracy with notable confusion between valence and dominance/arousal (Fig. 4.15a).

Theta demonstrates strong classification with high diagonal values ($C11$=20065, $C22$=19592, $C33$=25150) (Fig. 4.15b). Alpha accuracy is slightly lower than theta but reliably identifies valence ($C33$=24057) (Fig. 4.15c). Beta effectively distinguishes classes ($C11$=18064, $C22$=19203, $C33$=25362) (Fig. 4.15d). Gamma shows slightly higher misclassifications, particularly for dominance and valence (Fig. 4.15e). The combined rhythm ($\delta + \theta + \alpha + \beta + \gamma$) achieves the best performance with the highest diagonal values ($C11$=19785, $C22$=19747, $C33$=25627) (Fig. 4.15f), highlighting the advantage of integrating all EEG bands.

4.4.3.4 t-SNE Visualizations

The Fig. 4.16 displays EEG rhythms with t-SNE visualization. Compared to alpha and delta rhythm, beta, theta, and gamma rhythm characteristics exhibit superior discrimination. The best emotional discrimination is displayed by the combined rhythm.



(a) δ (b) θ (c) α

(d) β (e) γ (f) Combined (δ+θ+α+ β+ γ)

Figure 4.16: t-SNE visualizations of EEG rhythms: δ, θ, α, β, γ, and Combined(δ+θ+α+ β+ γ).

4.4.3.5 Computational Complexity Analysis

The worst-case time complexity of PSO is

$$O(N_{\text{Swarm}} \times I_{\text{Max}} \times C_{\text{eval}}),$$

where each particle (N_{Swarm}) evaluates the fitness function over a maximum of I_{Max} iterations. The evaluation cost C_{eval} includes filter design, signal filtering and reconstruction, and MSE computation. Runtime scales linearly with swarm size, number of iterations, or evaluation cost. The ASLT computational complexity is $O(L \log L)$, with L denoting signal length. HCRNet has an inference cost of 3.4 GFLOPs per forward pass and 12.6 million parameters, defining its complexity. Offline training, optimization, and near-real-time inference are achievable on modern hardware, consistent with observed runtimes (Table 4.10).

Table 4.10: Summary of computational metrics.

Component	Metric	Value
PSO	Time complexity	$O(N_s \cdot I_m \cdot C_e)$
ASLT	Time complexity	$O(L \log L)$
HRCNet	Params / FLOPs	12.6 M / 3.4 GFLOPs
PSO opt.	Runtime (GPU / CPU)	42 min / 1.8 h
ASLT preproc.	Runtime per signal (GPU / CPU)	121 ms / 335 ms
HRCNet inf.	Latency per sample (GPU / CPU)	16 ms / 42 ms

4.4.4 Comparison With Previous Study

(a) DEAP Dataset Previous Study:

The TMLP with SRDANN [113] focuses on transferable source samples to extract robust EEG features, achieving 61.8% and 57.07% accuracy for valence and arousal. AP-CapsNet [114] combines coordinate attention with a pre-trained convolutional capsule network, reaching 62.71% (valence), 63.51% (arousal), and 64.00% (dominance). DT-EEGNet [62], a dual-task model using multiscale attention and shared feature extraction, attains 63.44% (valence) and 63.87% (arousal). The lightweight 1DCNN-BiLSTM [115] achieves over 90% accuracy with low computational cost, making it suitable for wearable EEG-based emotion recognition.

Table 4.11: Comparison of models on the DEAP and DREAMER datasets in terms of accuracy and standard deviation (%).

Model (Year)	Valence(%)	Arousal(%)	Dominance(%)
TMLP+SRDANN (2023) [113]	61.88/5.55	57.07/7.23	–
AP-CapsNet (2023) [114]	62.71	63.51	64.00
DT-EEGNet (2024) [62]	63.44/4.59	63.87/2.38	–
1DCNN-BiLSTM (2024) [115]	95.28/1.91	95.57/2.25	–
Proposed HCRNet (2025)	**94.14/4.98**	**92.26/5.23**	**92.53/5.51**
DREAMER Dataset			
ECNN-C (2022)[116]	96.89	97.03	97.04
1DCNN-BiLSTM (2024) [115]	91.5/6.21	91.64/6.07	–
Proposed HCRNet (2025)	**98.45/3.53**	**95.27/5.55**	**97.09/3.46**

(b) DREAMER Dataset Previous Study

The ECNN-C model [116] combines supervised contrastive learning with efficient CNN blocks, achieving 96.89%, 97.03%, and 97.04% accuracy for valence, arousal, and dominance while reducing computational complexity. The lightweight 1DCNN-BiLSTM [115] lowers memory and computation through preprocessing, hybrid feature extraction, and quantization, achieving 91.5% (valence) and 91.6% (arousal), suitable for resource-constrained environments. The proposed lightweight HCRNet delivers the highest classification accuracy for three emotions across five EEG rhythms.

4.5 Summary

EEG signals are inherently prone to contamination from physiological and external sources, such as muscle activity, eye movements, and environmental interference. These artifacts significantly reduce the reliability of emotion recognition systems, making optimized preprocessing a critical step. To address this challenge, advanced frameworks combining filter optimization, adaptive decomposition, and noise-resilient transformations are employed.

First, EEG rhythms (δ, θ, α, β, and γ) are extracted through Butterworth bandpass filtering, with parameters optimized using the Crayfish Optimization Algorithm. By minimizing the mean squared error between original and reconstructed signals, this method improves frequency selectivity and reduces distortion. Similarly, a particle swarm optimization-based IIR filter adaptively selects cutoff frequencies and filter order, ensuring robust rhythm decomposition even under heavy artifacts conditions. The ASLT further refines time–frequency localization, effectively suppressing noise leakage and strengthening feature reliability. The study also introduces HCRNet, which combines customized CNN-based feature learning with transfer learning to improve feature extraction, robustness, and classification. The following chapter concentrates on adaptive mode filtering to increase EEG noise robustness and signal reliability after the filter parameters have been optimized.

Chapter 5

An Optimal Mode Filtering Framework for EEG Denoising

5.1 Introduction

EEG-ConvNet employs Mel and STFT spectrograms with optimized CNN architectures for high accuracy [117], while CRNN models combining CNN and Bi-RNN effectively capture spatial–temporal EEG features [61]. The superlet transform enhances time–frequency resolution for emotional valence detection [118], and wavelet scattering transform integrated with SVM achieves robust performance on the DEAP dataset [55, 119].

Despite these advancements, traditional time–frequency methods, such as wavelet, scalogram, and Mel spectrogram, suffer from redundancy, lack of invariance, and poor frequency–time interaction modeling. The joint time–frequency scattering (JTFS) technique [56] addresses these issues by providing stable, compact, and discriminative representations that capture both temporal and spectral modulations. Consequently, JTFS emerges as a powerful tool for accurate and efficient EEG-based emotion recognition. Emotions play a vital role in human cognition and decision-making, influencing mental and physical health. EEG-based emotion recognition has gained prominence over facial, speech, and gesture-based methods due to its direct measurement of brain activity [40, 120, 121, 2]. EEG signals, being non-stationary, are effectively analyzed through time–frequency techniques, including STFT, CWT, and WVD, which convert signals into image representations for deep learning models.

Deep learning methods like CNNs, CRNNs, and hybrid architectures (e.g., CNN-GRU-BiLSTM, DT-EEGNet, PTSFNN) have demonstrated strong performance, achieving accuracies above 90% on benchmark datasets such as DEAP and SEED [52, 62]. EEG denoising in emotion identification requires simultaneous processing of non-stationarity, deformation sensitivity, and artifact contamination, which are challenging to address with a single filtering process. Therefore, a hybrid mode-filtering framework was built using complementary signal improvement techniques.

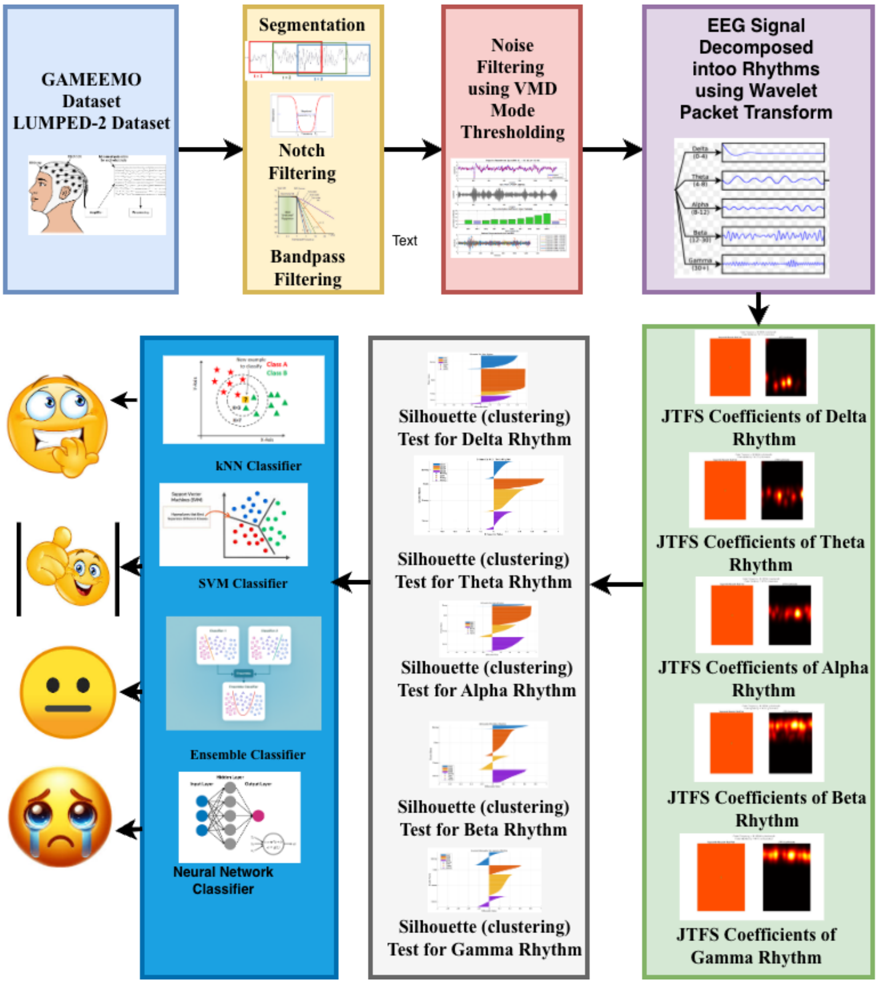


Figure 5.1: Block diagram of the proposed methodology for emotion recognition using VMD-based mode filtering and joint time–frequency scattering feature classification.

The first method uses Joint Time-Frequency Scattering (JTFS) to produce deformation-resistant signal representations, allowing for robust neural rhythm extraction while maintaining temporal-spectral consistency across emotional states. Although JTFS improves representation stability, it provides just limited adaptive suppression of noise components. To address this limitation, a Group Sparse Mode Decomposition (GSMD)-based method is used to selectively reconstruct physiologically significant oscillatory modes while suppressing artifact-dominated components. The Superlet transform is subsequently utilized to improve temporal-frequency localization and decrease spectral leakage.

5.2 Joint Time Frequency Scattering-Based Framework for Emotion Recognition

5.2.1 Methodology

The block diagram for the proposed methodology is displayed in Figure 9.1

5.2.1.1 Preprocessing

EEG signals from both datasets are segmented into 10-second intervals using a 5-second window with 50% overlap, corresponding to 1280 samples with a sliding step of 640 samples.

Each segment is labeled with its emotion class, and data from all 8 or 14 channels are stored independently as row-wise segments with labels in a structured format. A Butterworth band-pass filter (0.5–48 Hz) and notch filter are applied to each segment.

5.2.1.2 Joint Time-Frequency Scattering

The joint time-frequency scattering transform [56] is a non-learned deep convolutional representation that ensures time-shift invariance through modulus nonlinearities and cascading wavelet convolutions. Using analytically defined complex wavelet filters, it decomposes signals across time and frequency, applying modulus and averaging for stability.

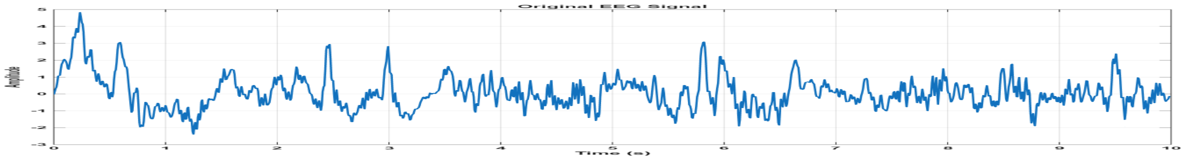
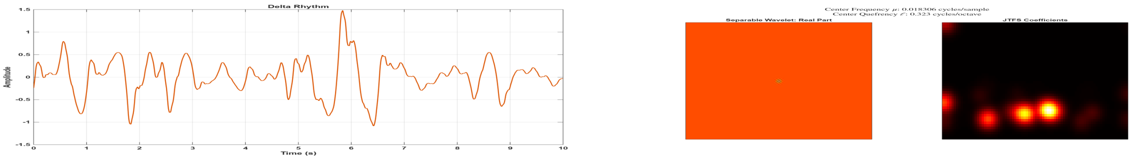


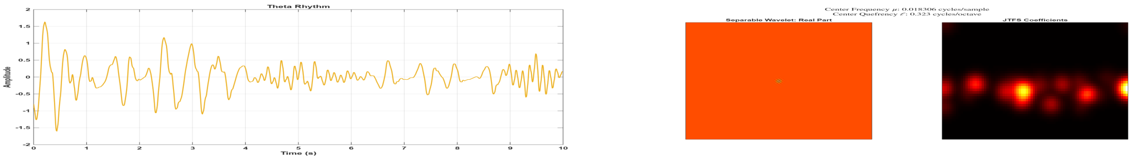
Figure 5.2: Original EEG signal.



(a) δ Rhythm.

(b) JTFS Coefficients (δ).

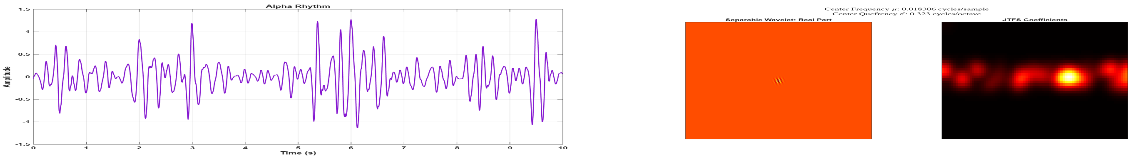
Figure 5.3: JTFS analysis of δ rhythm.



(a) θ Rhythm.

(b) JTFS Coefficients (θ).

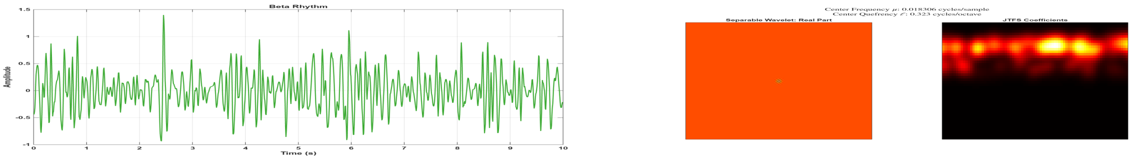
Figure 5.4: JTFS analysis of θ rhythm.



(a) α Rhythm.

(b) JTFS Coefficients (α).

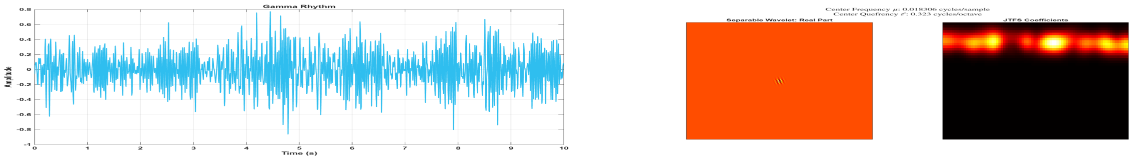
Figure 5.5: JTFS analysis of α rhythm.



(a) β Rhythm.

(b) JTFS Coefficients (β).

Figure 5.6: JTFS analysis of β rhythm.



(a) γ Rhythm.

(b) JTFS Coefficients (γ).

Figure 5.7: JTFS analysis of γ rhythm.

JTFS improves upon mel-spectrograms and time scattering by capturing joint time-frequency structures, effectively representing non-stationary patterns such as frequency modulations and transients. It produces stable, hierarchical, and interpretable representations without requiring learned parameters.

(a) Wavelet Time Scattering:

Wavelet time scattering extracts translation-invariant, low-variance features from time series, with time-shift invariance adjustable by scale. The method convolves the input with wavelet filters, applies a pointwise modulus to obtain a scalogram capturing time-frequency structure, and smooths it with a low-pass scaling filter to preserve local time-shift invariance.

Given an input signal $m(t)$, the network is composed of two successive filter banks and is structured as a tree [22a].

$$\begin{aligned}\left\{m *^T \psi_\lambda^{(1)} *^T \phi_T^T\right\} &\longleftarrow \left\{m *^T \psi_\lambda^{(1)}\right\} \\ &\longrightarrow \left\{\left|m *^T \psi_\lambda^{(1)}\right| *^T \psi_\mu^{(2)}\right\} \\ &\longrightarrow \left\{\left|\left|m *^T \psi_\lambda^{(1)}\right| *^T \psi_\mu^{(2)}\right| *^T \phi_T^T\right\} \end{aligned} \tag{5.1}$$

Where:

- $*_T$ represents convolution in time.
- $\psi_\lambda^{(1)}$ and $\psi_\mu^{(2)}$ are the time-domain wavelets in the first- and second-order filter banks, respectively, where the subscripts λ and μ represent the corresponding wavelet center frequencies.
- $\phi_T(t)$ represent scaling function (lowpass filter).

The first-order scalogram represents the absolute magnitude of the continuous wavelet transform (CWT) of the input data [22b].

$$U_1 m(t, \lambda) = \left|m *_T \psi_\lambda^{(1)}(t)\right| \tag{5.2}$$

The first-order scattering coefficients are obtained by low-pass filtering [22b]:

$$S_1 m(t, \lambda) = U_1 m(t, \lambda) *_T \phi_T(t) \tag{5.3}$$

Second-order Scattering Coefficients

The second-order coefficients capture temporal modulations [122]:

$$U_2 m(t, \lambda, \mu) = \left|U_1 m(t, \lambda) *_T \psi_\mu^{(2)}(t)\right| \tag{5.4}$$

where $\psi_\mu^{(2)}(t)$ are second-order wavelets with modulation rate μ in Hz. The smoothed coefficients are [122]:

$$S_2 m(t, \lambda, \mu) = U_2 m(t, \lambda, \mu) *_T \phi_T(t) \tag{5.5}$$

(b) JTFS Mathematical Analysis:

The joint time-frequency transform is inspired by auditory processing, where neurons in the primary auditory cortex respond to spectro-temporal receptive fields (STRFs), i.e., patterns across time and frequency. These STRFs exhibit ripple-like activity in the time-frequency space, characterized by two modulation rates including frequency modulation (ℓ, cycles per octave) and temporal modulation (μ, Hz). Thus, cortical responses are modeled using four parameters: time (t), frequency (λ), temporal rate (μ), and frequency rate (ℓ).

The separable wavelet is defined as [56]:

$$\Psi_{\mu,\ell,s}(t, \lambda) = \psi_\mu^{(2)}(t) \cdot \psi_{\ell,s}(s\lambda) \tag{5.6}$$

where:

- $\psi_\mu^{(2)}(t)$: Temporal wavelet (modulation rate μ)
- $\psi_{\ell,s}(s\lambda)$: Frequential wavelet (modulation rate ℓ in cycles/octave)
- $s \in \{+1, -1\}$: Spin parameter

$$U_1 m(t, \lambda) = \left| m *_T \psi_\lambda^{(1)}(t) \right| \tag{5.7}$$

The scatteringTransform method of the waveletScattering object generates first- and second-order coefficients, with the latter representing smoothed second-order scalogram values. Joint time–frequency scattering builds on this by adding a feature extractor that captures the full cortical response [56].

$$U_{\text{JTFS}} m(t, \lambda, \mu, \ell, s) = \left| U_1 m(t, \lambda) *_{T,F} \Psi_{\mu,\ell,s}(t, \lambda) \right| \tag{5.8}$$

$$\begin{aligned} S_{\text{JTFS}} m(t, \lambda, \mu, \ell, s) = {} & U_{\text{JTFS}} m(t, \lambda, \mu, \ell, s) \\ & *_{T,F} \phi_{T,F}(t, \lambda) \end{aligned} \tag{5.9}$$

Where $*_{T,F}$ denotes convolution in both time and frequency and $\phi_{T,F}(t, \lambda)$ is a joint time-frequency low-pass filter.

The second-order coefficients depend on μ, the center frequencies of the second-order time wavelets $\psi_\mu^{(2)}(t)$. To capture the ripple-like patterns observed in time-frequency responses, JTFS applies a two-dimensional separable wavelet.

(c) JTFS and Sensitivity to Time-Frequency Geometry:

Joint Time-Frequency Scattering produces various feature groups or coefficients, each emphasizing unique patterns within the time-frequency structure of a signal. The EEG signal displayed in Figure 5.2 is decomposed into δ, θ, α, β, and γ using algorithm 5.3. The 1D representations of δ, θ, α, β, and γ are illustrated in Figures 5.3a, 5.4a, 5.5a, 5.6a, and 5.7a. The corresponding 2D representations, utilizing the real part coefficients from JTFS separable wavelets, are shown Figures 5.3b, 5.4b, 5.5b, 5.6b,ands 5.7b. Figures 5.3b, 5.4b, 5.5b, 5.6b,and 5.7b

Algorithm 5.1 : Joint Time-Frequency Scattering Feature Extraction for Classification

```
Require: EEG segments $\mathbf{X} = \{x_i(t)\}_{i=1}^M, x_i \in \mathbb{R}^N$, Labels $\mathbf{y} \in \{1,2,3,4\}^M$, JTFS params: $J_t, Q_t, J_f, Q_f, T, F$
Ensure: Feature matrix $\mathbf{F} \in \mathbb{R}^{M\times D}$, Label vector $\mathbf{y}$
Construct time wavelets $\psi_\lambda^{(1)}(t) = 2^{-j/Q_t}\psi(2^{-j/Q_t}t)$
Construct freq. wavelets $\psi_{\ell,s}(\log_2\lambda) = e^{i2\pi s\ell\log_2\lambda}\psi(\log_2\lambda)$
Define $\Psi_{\mu,\ell,s}(t,\lambda) = \psi_\mu^{(2)}(t)\psi_{\ell,s}(\log_2\lambda)$
Define $\phi_{T,F}(t,\lambda) = \exp(-t^2/2T^2)\exp(-\log_2^2\lambda/2F^2)$
for $i = 1$ to $M$ do
    $x \leftarrow x_i(t)$
    $U_1x(t,\lambda) = |x *_t \psi_\lambda^{(1)}(t)|$, $S_1x = U_1x *_t \phi_T$
    $U_2x = |U_1x *_t \psi_\mu^{(2)}(t)|$, $S_2x = U_2x *_t \phi_T$
    for each $(\mu,\ell,s)$ do
        $U_{\text{JTFS}} = |U_1x *_{t,\lambda} \Psi_{\mu,\ell,s}|$
        $S_{\text{JTFS}} = U_{\text{JTFS}} *_{t,\lambda} \phi_{T,F}$
    end for
    $\mathbf{f}_i = [\text{vec}(S_1x), \text{vec}(S_2x), \text{vec}(S_{\text{JTFS}})]^T$
    $\mathbf{F}[i,:] \leftarrow \mathbf{f}_i$
end for
return $\mathbf{F}, \mathbf{y}$
```

demonstrate that the JTFS coefficients for δ, θ, and α are distinct and separable from those of β and γ. The algorithm 5.1 outlines a step-by-step procedure for extracting the JTFS features.

5.2.1.3 Silhouette Clustering Test

Silhouette clustering [123] assesses feature consistency by comparing a data point's similarity within its own cluster against other clusters, yielding values from -1 to $+1$. Higher scores indicate better clustering, with ≥ 0.7 considered strong, ≥ 0.5 decent, and ≥ 0.25 weak. In high-dimensional data, the curse of dimensionality often reduces silhouette scores due to more uniform point distances.

The average distance between point i and all other points in its cluster C_I, excluding itself, is given by [123]:

$$a(i) = \frac{1}{|C_I| - 1} \sum_{\substack{j\in C_I \\ j\neq i}} d(i,j) \tag{5.10}$$

The minimum average distance from point i to all points in any other cluster C_J (where $J \neq I$) is given by [123]:

$$b(i) = \min_{J\neq I}\left(\frac{1}{|C_J|}\sum_{j\in C_J} d(i,j)\right) \tag{5.11}$$

The silhouette value $s(i)$ computes how well point i is clustered, computed as the normalized difference between the nearest-cluster distance $b(i)$ and intra-cluster distance $a(i)$ [123]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5.12}$$

$$s(i) = \begin{cases} 1 - \dfrac{a(i)}{b(i)}, & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ \dfrac{b(i)}{a(i)} - 1, & \text{if } a(i) > b(i) \end{cases} \tag{5.13}$$

$$-1 \leq s(i) \leq 1 \tag{5.14}$$

$$s(i) = 0 \quad \text{if } |C_I| = 1 \tag{5.15}$$

5.2.2 Results and Discussion

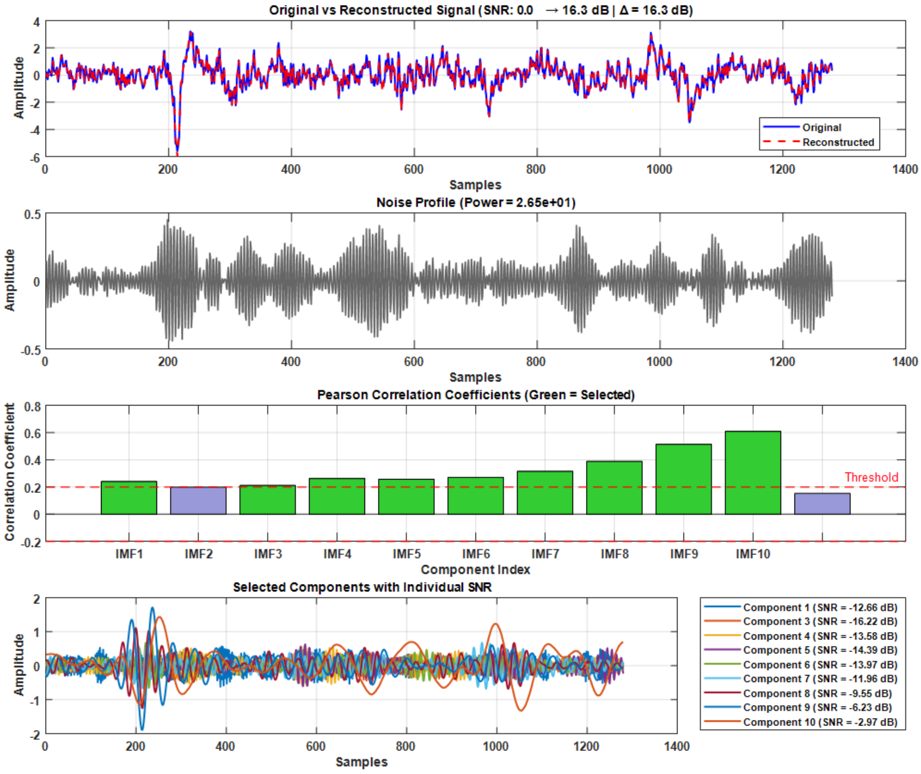


Figure 5.8: EEG Signal Denoising using VMD.

5.2.2.1 Filtering of Noise

Algorithm 5.2 employs variational mode decomposition (VMD) [124] to denoise EEG signals. Using preset parameters including number of modes (K), absolute tolerance (ϵ), relative tolerance (τ), penalty factor (α), and LMUpdate rate (η)—VMD decomposes the signal into a residual component and several intrinsic mode functions (IMFs). The algorithm computes the Pearson correlation between each IMF and the original signal, retaining only modes with significant correlation $|r| > 0.2$ for reconstruction. The filtered EEG signal is obtained by summing these relevant modes, effectively removing noise while preserving important patterns. All EEG signals from Dataset 1 (GAMEEMO) and Dataset 2 (LUMED-2) are processed using this method to provide clean inputs for further analysis.

5.2.2.2 EEG Signal Decomposition into Rhythms

Wavelet Packet Decomposition (WPD) is applied to the filtered EEG signals from Datasets 1 and 2 to extract standard brain rhythms (δ, θ, α, β, γ) by decomposing the signals into frequency-aligned subbands using a complete binary tree. Terminal nodes are reconstructed

Algorithm 5.2 : Correlation-Thresholded VMD Framework for Signal Reconstruction

```
Require: EEG signal $x \in \mathbb{R}^N$
Ensure: Reconstructed $x_{\text{rec}}$, correlations $\rho$, mask $\mathbf{m}$
Decompose signal:
IMF, res, _ ← VMD($x$,
    {$\alpha$=1000, $\epsilon$=5×10$^{-6}$, $\tau$=0.005,
    $K$=10, $\eta$=0.01, init='peaks'})
Combine: $\mathbf{C} \leftarrow [\text{IMF}_1, \ldots, \text{IMF}_{10}, \text{res}] \in \mathbb{R}^{N\times 11}$
Correlations:
for $k = 1$ to 11 do
    $\rho_k \leftarrow \dfrac{\sum(x-\bar{x})(c_k-\bar{c}_k)}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(c_k-\bar{c}_k)^2}}$
end for
Thresholding: $\mathbf{m} \leftarrow [|\rho_k| > 0.2]$
if $\sum m_k = 0$ then
    $j \leftarrow \arg\max |\rho_k|$, $m_j \leftarrow$ true
end if
Reconstruct: $x_{\text{rec}} \leftarrow \sum m_k \cdot \mathbf{C}_{:,k}$
return $x_{\text{rec}}, \rho, \mathbf{m}$
```

and mapped to the corresponding EEG rhythms, enabling temporal-domain reconstruction for further analysis. Algorithm 5.3 details this step-by-step decomposition process.

5.2.2.3 Feature Extraction using JTFS

EEG rhythms are processed to extract features that improve emotion recognition. After loading the EEG data and corresponding labels, the JTFS transform is configured to capture detailed signal characteristics in both time and frequency domains. The JTFS parameters used for feature extraction are listed in Table 5.1.

Table 5.1: Empirical values of JTFS parameters.

PARAMETERS	VALUES
SignalLength	1280
NumTimeOctaves	3
TimeQualityFactors	4
TimeInvarianceScale	40
NumFrequencyOctaves	2
FrequencyQualityFactor	1
FrequencyMaxPaddingFactor	2
FilterDatatype	Single
EnergyCorrectFilters	TRUE

JTFS processes each EEG segment independently to generate features that capture the signal dynamics. These features are combined with their corresponding labels to form a feature-label matrix, which is used to train machine learning algorithms for emotion classifi-

Algorithm 5.3 : EEG Signal Decomposition using Wavelet Packet Transform

```
Require: EEG signal $x$, sampling rate $F_s$, level $L$, wavelet $\psi$
Ensure: Rhythms $\{r_b\}$ for $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$
Initialize $r_b \leftarrow 0$ for all $b$
$\mathcal{B} \leftarrow \{\delta : [0.5, 4], \theta : [4, 8], \alpha : [8, 13], \beta : [13, 30], \gamma : [30, F_s/2]\}$
$T \leftarrow \text{WPDEC}(x, L, \psi)$, $N \leftarrow \text{LEAVES}(T)$
for each node $k \in N$ do
    $d_k \leftarrow \text{WPRCOEF}(T, k)$
    $[f_l, f_h] \leftarrow \text{NODE\_FREQ}(k, L, F_s)$
    $f_c \leftarrow (f_l + f_h)/2$
    for each $b \in \mathcal{B}$ do
        if $f_c \in [f_b^{\min}, f_b^{\max}]$ then
            $r_b \leftarrow r_b + d_k$
        end if
    end for
end for
return $\{r_b\}$
function NODE_FREQ($k, L, F_s$)
    $\Delta f \leftarrow \frac{F_s}{2^{L+1}}$
    return $[k \cdot \Delta f, (k + 1) \cdot \Delta f]$
end function
```

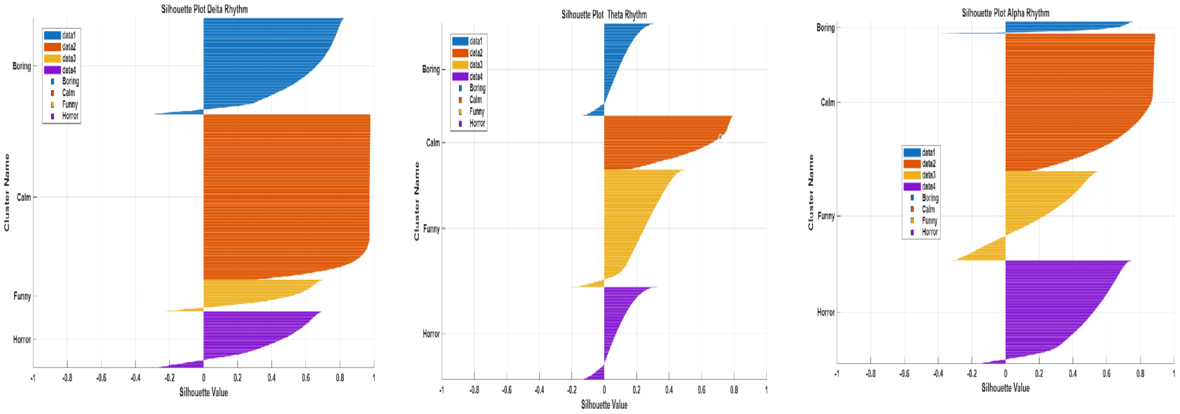
cation. Algorithm 5.1 details the step-by-step feature extraction process. The JTFS transform produces feature matrices of size 5040×2280 for Dataset I (GAMEEMO [35]) (5040 signals) and 5223×2280 for Dataset II (LUMPED-2[55]) (5223 signals), with a total of 2280 time-frequency scattering features. Due to the large number of features, a clustering test is performed to evaluate their relevance.

5.2.2.4 Silhouette Feature Test for Dataset-I

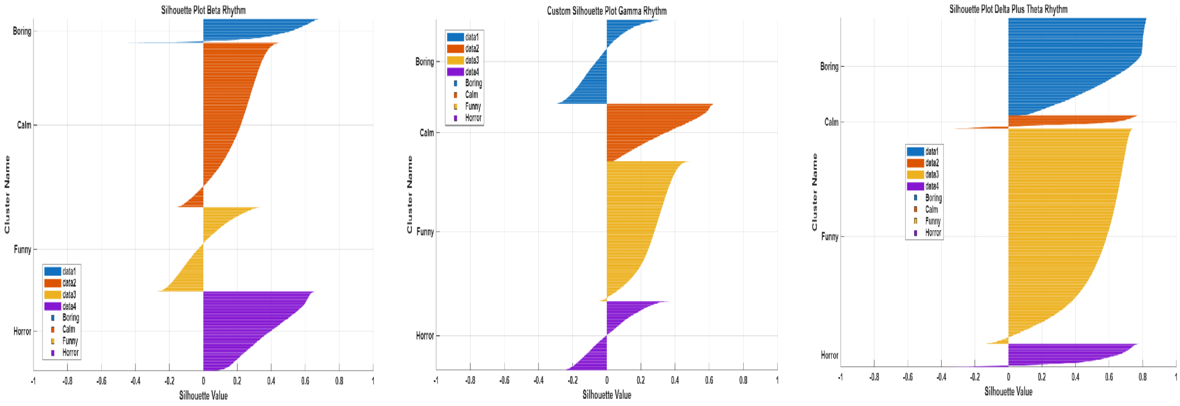
The silhouette method evaluates the optimal number of data clusters by measuring how well a data point fits within its cluster compared to others, with values ranging from -1 to 1. Scores ≥ 0.7 indicate strong clustering, ≥ 0.5 are decent, and ≥ 0.25 are weak. Clusters identified via the Silhouette Test are linked to emotions. Average silhouette values for the δ, α, and $\delta + \theta$ rhythms exceed 0.5 (Figures 5.9a, 5.9c, 5.9f), indicating these rhythms provide significant discriminative features for emotion classification. The θ rhythm shows moderate separation with a silhouette value above 0.3 and minimal negative values (Figure 5.9b), whereas β and γ rhythms exhibit clusters with highly negative values (Figures 5.9d, 5.9e).

5.2.2.5 Dataset-I Classification Results using Multiple Algorithms

JTFS features (subject-independent) are used as input for classifiers, including k-NN, SVM, ensemble methods, and neural networks. The features are split 70:30 for training and testing, respectively. Validation accuracy is assessed using ten-fold cross-validation, and testing is performed on randomly sequenced features.



(a) δ Rhythm (b) θ Rhythm (c) α Rhythm



(d) β Rhythm (e) γ Rhythm (f) $\delta + \theta$ Rhythm

Figure 5.9: Silhouette feature test results of EEG rhythms for classifying four emotions.

(a) Classifier Performance

The Fine KNN classifier achieves the highest average accuracy, outperforming other models, particularly in occipital and posterior channels (O1, O2, P8) and low-frequency rhythms (δ and θ) (Figures 5.10). Ensemble subspace KNN performs robustly in coupled δ and θ rhythms with 76.44% accuracy, ranking second overall at 65.19%. Fine Gaussian SVM shows moderate performance (62.72%) with reduced accuracy in high-frequency rhythms (β and γ), while the Wide Neural Network underperforms across all channels and rhythms, achieving 51.47%. The results suggest that simple distance-based methods like kNN handle the nonlinear and non-stationary nature of EEG rhythms more effectively than complex models such as neural networks.

(b) Rhythm-Specific Performance

Figure 5.10 show that the $\delta + \theta$ rhythms achieve the highest accuracy, reaching 80.13% with Fine kNN, consistent with neuroscience findings that slow-wave oscillations encode task-relevant properties like attention and sleep stages. The γ rhythm is the most challenging, with a maximum accuracy of 49.31%, due to its low amplitude and susceptibility to noise and artifacts. α and β rhythms show moderate performance, with accuracy decreasing as frequency increases. Low-frequency rhythms provide more discriminative features, whereas high-frequency rhythms require advanced noise reduction or feature engineering.

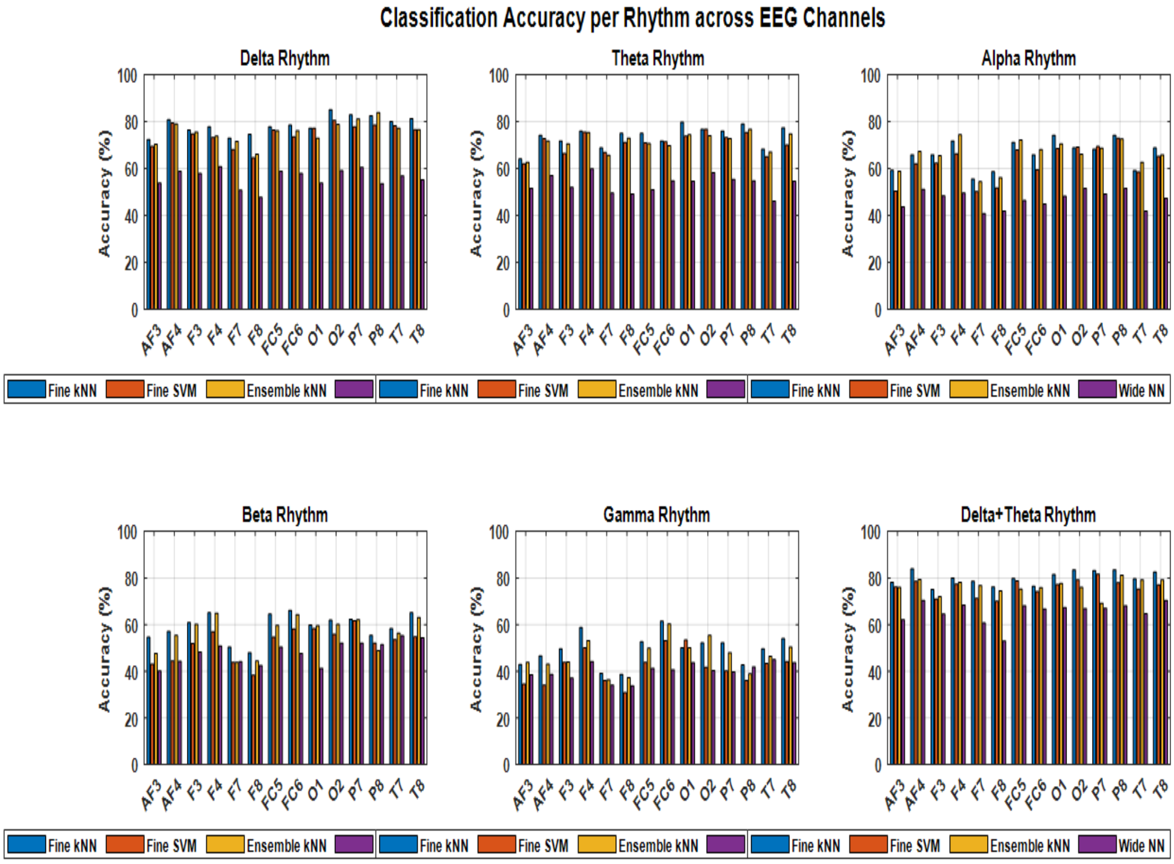


Figure 5.10: Channel-wise classification accuracy across EEG rhythms using Fine kNN, Fine Gaussian SVM, Ensemble kNN, and Wide Neural Network classifiers.

(c) Performance across EEG Channels

Figure 5.10 shows that posterior and occipital channels (O1, O2, P7, P8) achieve the highest classification accuracy, with O2 reaching 85.1% for δ rhythms using Fine kNN, reflecting strong low-frequency activity linked to visual processing. Frontal and central channels (F4, FC5, FC6) also perform well, particularly in δ and θ bands, consistent with their role in executive and cognitive regulation. Frontopolar channels (AF3, F7, F8) exhibit the lowest accuracy, with F7 capturing only 39.1% of γ rhythms, likely due to susceptibility to physiological artifacts such as eye movements and muscle activity. Overall, posterior channels are most effective for rhythm-based classification, while frontal channels may benefit from artifact removal techniques.

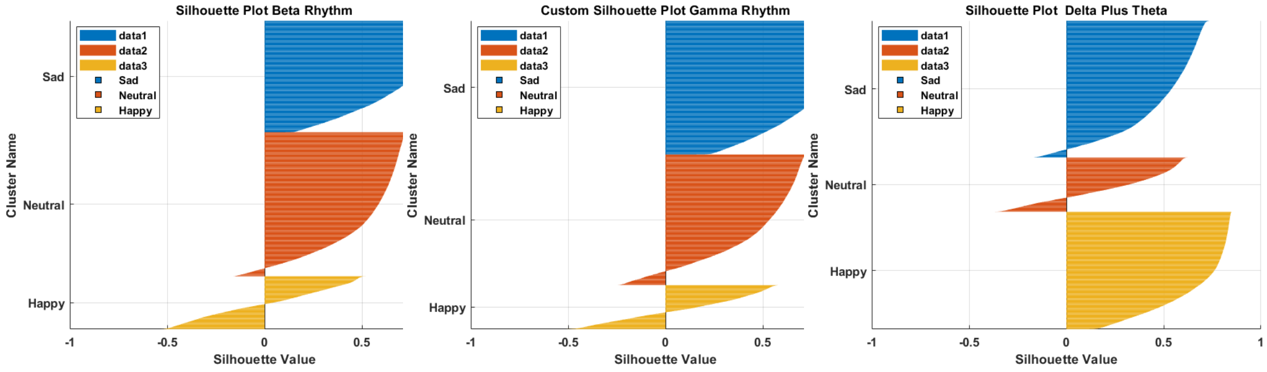
(d) Emotion-Specific Performance

Figure 5.10 shows that the highest classification accuracy is achieved for the boring emotion, followed by horror and funny, with horror and funny exhibiting similar accuracy. Figure 5.10 indicates that frontal, parietal, and occipital electrodes contribute most to emotion recognition.

5.2.2.6 Silhouette Feature Test for Dataset-II

The Silhouette Feature Test on JTFS features from different EEG rhythms (Figure 5.11) shows that δ, θ, α, and $\delta + \theta$ clusters perform well with average silhouette values above 0.5, whereas β and γ feature clusters exhibit more negative values.

δ Rhythm θ Rhythm α Rhythm



β Rhythm γ Rhythm $\delta + \theta$ Rhythms

Figure 5.11: Silhouette test results for EEG rhythms (δ, θ, α, β, γ, and δ+θ) in classifying four emotions for Dataset-II.

5.2.2.7 Dataset-II Classification Results using Multiple Algorithms

LUMPED-2, the second dataset, is smaller than the first, containing eight-channel EEG recordings from 13 individuals across three emotion classes. In contrast, the first dataset comprises 14 channels from 28 subjects, each with four emotion classes. LUMPED-2 achieves higher accuracy due to its smaller size, more consistent data, and fewer classes.

(a) Classifier Performance

Figure 5.12 shows that Fine kNN achieves the highest overall accuracy, excelling in posterior regions and low-frequency bands (δ/θ). Ensemble Subspace kNN performs well in frontal areas and combined rhythms. Neural networks show competitive performance for δ and θ but moderate overall accuracy, while Fine Gaussian SVM performs the lowest, especially in high-frequency bands.

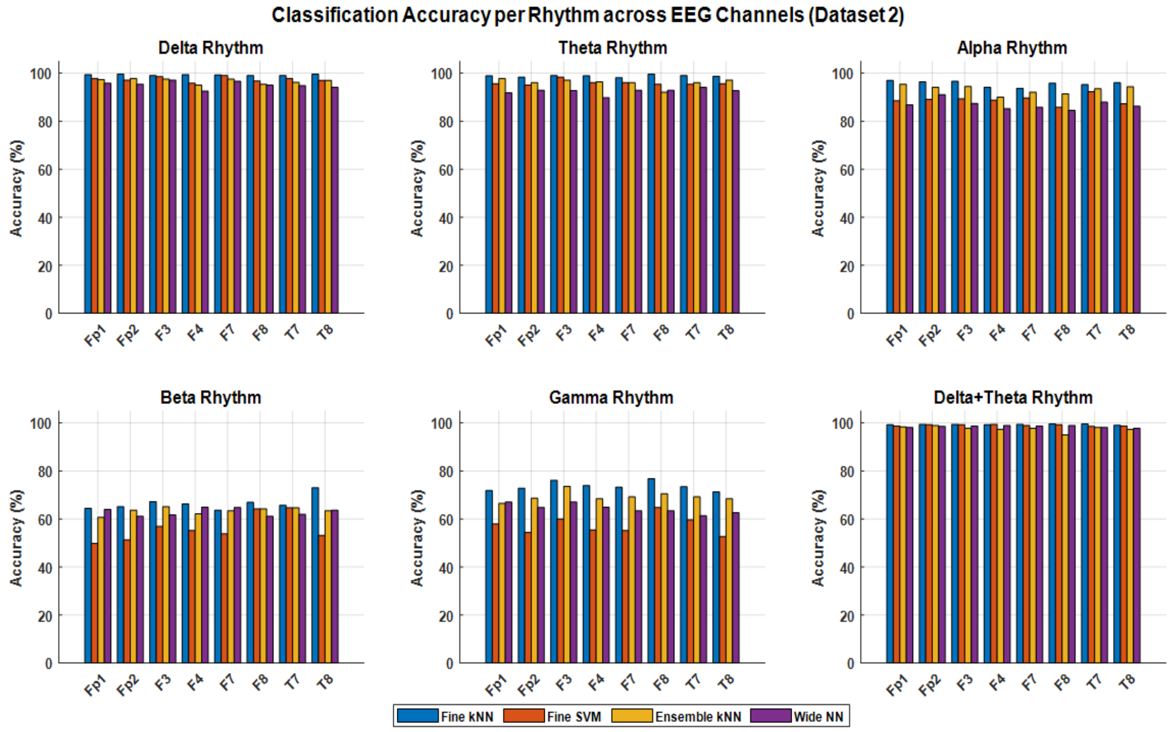


Figure 5.12: Classification accuracy across EEG channels for different rhythms using four classifiers: Fine kNN, Fine Gaussian SVM, Ensemble kNN, and Wide Neural Network for Dataset II.

(b) Rhythms Specific Performance

Figure 5.12 shows that Fine kNN performs well across all rhythms, demonstrating reliable classification. All classifiers achieve optimal accuracy, often exceeding 92%, for low-frequency rhythms (δ and θ). High-frequency rhythms (β and γ) are more challenging, with overall accuracy dropping by 20–30%. Combining δ and θ rhythms further improves performance, yielding near-perfect accuracy and highlighting the effectiveness of rhythm combinations for emotion recognition.

(c) Performance across EEG Channels

Figure 5.12 shows that Fine kNN achieves the highest classification accuracy (89.6%) in frontal and temporal channels (F3, F8, T8), underscoring their importance in emotion-related EEG analysis. Low-frequency rhythms (δ, θ, and $\delta + \theta$) consistently perform best across all channels, with accuracies between 98% and 99.6%. In contrast, the γ rhythm shows variable performance, peaking at 76.8% for F8 with Fine kNN and dropping to 52.7% at Fp2 with SVM, indicating sensitivity to both classifier choice and channel location.

(d) Emotion-Specific Performance

Figure 5.12 shows that sad emotion achieves the highest accuracy, followed by happy and neutral. Figure 5.12 indicates that frontal and temporal electrodes play a key role in emotion recognition.

5.2.2.8 Performance Parameters for Dataset I and II

The δ and θ rhythms achieve the highest performance (accuracy: 85–90%, F1: 76–82%, AUC: 0.84–0.88), strongly linking them to emotional processing. α and β rhythms show moderate discriminative ability (accuracy: 74–82%, F1: 57–70%, AUC: 0.71–0.81), while γ rhythms perform poorest (accuracy: 68–71%, F1: 48–53%, AUC: 0.65–0.69). The combined $\delta + \theta$ rhythms outperform individual rhythms, reaching accuracy of 87–90%, F1 of 77–82%, and AUC of 0.85–0.88. Among emotions, “boring” shows the best metrics, “horror” and “funny” perform similarly, and “calm” is slightly less distinct. Overall, δ and θ are the most effective rhythms, with results validated by MCC and Cohen’s Kappa, confirming performance beyond chance.

The δ, θ, and $\delta + \theta$ rhythms consistently achieve outstanding performance across happy, neutral, and sad states, with the $\delta + \theta$ combination performing best—exceeding 99% on all metrics, especially for sadness (accuracy: 99.73%, F1: 99.66%, MCC: 99.44, AUC: 0.9945). θ and $\delta+\theta$ rhythms classify happy emotions effectively (accuracy up to 99.54%, MCC ≈ 0.99), while neutral emotions also show F1 scores above 99% across these rhythms. Sad emotion yields exceptionally high recall and F1, particularly with $\delta + \theta$ (recall: 99.68%, F1: 99.66%). The α rhythm performs well but slightly lower (accuracy and F1: 95–97%), whereas β shows the weakest results, especially for happy and sad emotions, with accuracy around 76%, F1: 63–69%, and MCC: 46–49%.

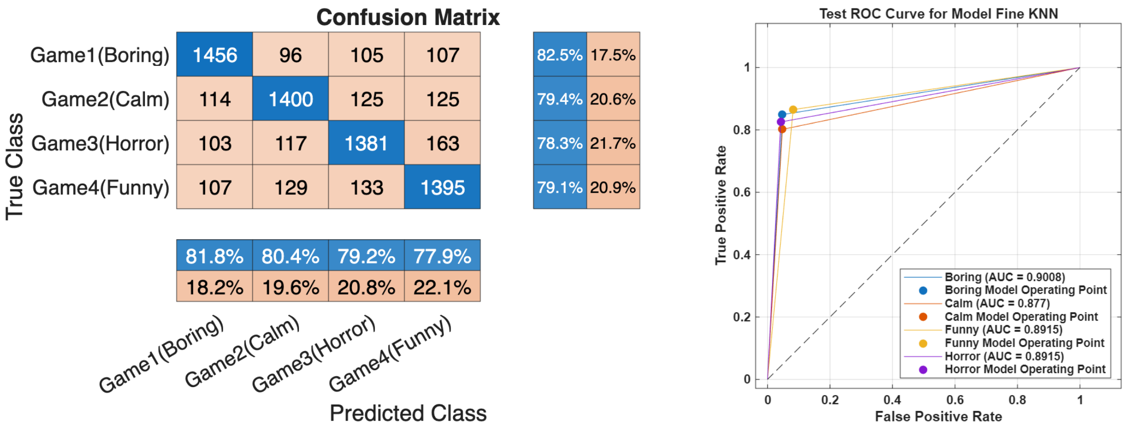
5.2.2.9 Confusion Matrix and ROC Curve

A confusion matrix evaluates classification performance by comparing actual and predicted labels, showing true positives, false positives, true negatives, and false negatives. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at varying thresholds to assess class separability. The area under the ROC curve (AUC) quantifies the probability that a positive sample ranks higher than a negative one: 0.5 indicates random guessing, 0.7–0.8 acceptable discrimination, and 0.8–0.9 outstanding discrimination. A higher AUC reflects stronger classification ability.

(a) Dataset-I Performance

The confusion matrix for dataset I shows dominant diagonal values, indicating effective recognition of all four emotional states with nearly 80% accuracy and consistent performance. Misclassifications occur between similar classes, such as ”boring” vs. ”calm” (96 and 114) and horror vs. funny, likely due to overlapping features and emotional similarities in the δ and θ bands. Overall, the matrix confirms high correct prediction rates and supports that combining δ and θ rhythms is most effective for emotion recognition.

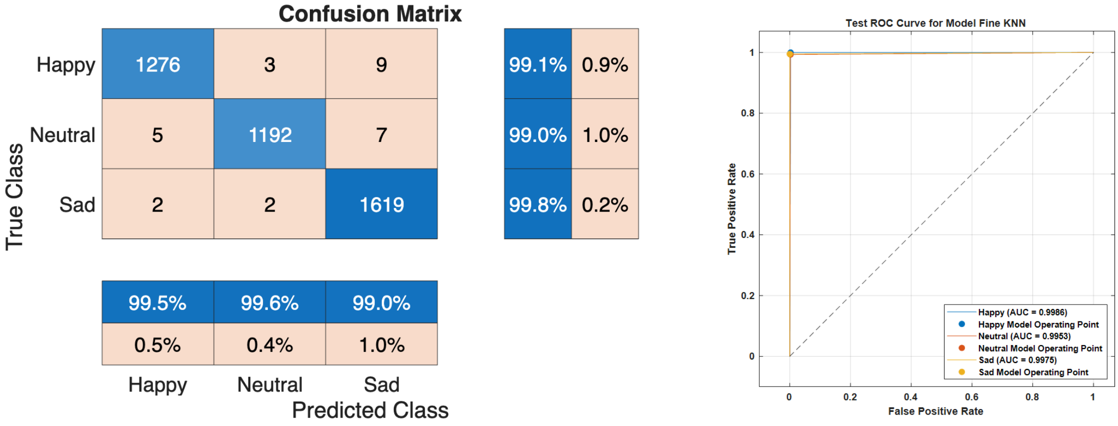
The AUC for boring is 0.9008, showing strong separability from other states. Calm has a slightly lower AUC of 0.877, likely due to overlap with boring. Funny and horror both achieve an AUC of 0.8915, indicating outstanding separability but also suggesting overlapping or similar brain response patterns.



	Game1(Boring)	Game2(Calm)	Game3(Horror)	Game4(Funny)		
Game1(Boring)	1456	96	105	107	82.5%	17.5%
Game2(Calm)	114	1400	125	125	79.4%	20.6%
Game3(Horror)	103	117	1381	163	78.3%	21.7%
Game4(Funny)	107	129	133	1395	79.1%	20.9%
	81.8%	80.4%	79.2%	77.9%		
	18.2%	19.6%	20.8%	22.1%		

(a) DS1: Confusion Matrix

(b) DS1: ROC Curve



	Happy	Neutral	Sad		
Happy	1276	3	9	99.1%	0.9%
Neutral	5	1192	7	99.0%	1.0%
Sad	2	2	1619	99.8%	0.2%
	99.5%	99.6%	99.0%		
	0.5%	0.4%	1.0%		

(c) DS2: Confusion Matrix

(d) DS2: ROC Curve

Figure 5.13: Performance metrics of emotion recognition using VMD-based mode filtering and JTFS features: confusion matrices and ROC curves for Datasets I and II.

Table 5.2: Comparison of previous studies using Dataset I and Dataset II.

Authors	Methods Used	N.C.	Accuracy (%)
Dataset I (4-Class Emotion Recognition)			
Alakus et al.[35], 2020	Hjorth, Entropy, SD, VAR, ZC, and MLPNN	4	73.21
Alakus et al. [35], 2020	Hjorth, Entropy, SD, VAR, ZC, and MLPNN	2	81.35
Alakus and Turkoglu [125], 2020	Spectral Entropy and Bi-LSTM	2	76.93
Aslan et al. [126], 2022	Scalogram, GoogleNet, and ELM	2	98.41
Tuncer et al. [127], 2022	Fused features by LEDPatNet19 and Cubic SVM	2	99.29
Su et al. [128], 2023	GA-PDPL	4	49.01
Su et al. [128], 2023	GA-PDPL	2	64.34
M. Aslan et al. [129], 2024	Energy, R.E., S.D., VAR, SK, and CNN	4	95.30
Proposed work (Dataset I)	**δ+θ, JTFS, Silhouette, kNN (SI)**	**4**	**80.00**
Dataset II (3-Class Emotion Recognition)			
Alak et al. [36], 2021	Convolutional Neural Network and Decision Tree	3	74.20
Alam et al. [130], 2023	Statistical Features and SVM, KNN, and Decision Tree	3	99.00
Proposed work (Dataset II)	**δ+θ, JTFS, Silhouette, kNN (SI)**	**3**	**99.35**

(b) Dataset-II Performace

The confusion matrix shows large diagonal values, indicating strong classifier performance across all emotional classes, with minimal misclassifications: 3 happy as neutral, 4 sadness misclassified, and 7 neutral as sad. The high accuracy confirms that emotional states produce distinct patterns in the $\delta+\theta$ bands. The Fine kNN classifier effectively distinguishes emotional states using $\delta + \theta$ EEG rhythms, with all classes achieving AUC values above 0.995. This indicates outstanding discriminative power and confirms the reliability and relevance of the extracted features for multi-class emotion classification.

5.2.3 Comparison With Previous Studies

5.2.3.1 Dataset-I Previous Studies

Table 5.2 summarizes studies using Dataset I (GAMEEMO). Alakus et al.[35] achieved 73.21% accuracy with MLPNN on features like Hjorth, entropy, SD, VAR, and ZC, while kNN and SVM performed lower at 42.85% and 55%. For two-class classification, they reported 81.3% with MLPNN, 72.5% with SVM, and 66% with kNN. Alakus and Turkoglu [125] proposed a bidirectional LSTM, reaching 76.91% accuracy using spectral entropy across all channels. Aslan et al. (2022) [126] applied GoogleNet with scalogram-based EEG image conversion, extracting features for classification via kNN, SVM, and ELM.

Tuncer et al.[127] (2022) proposed LEDPatNet 19, combining TQWT-derived 18 sub-bands, statistical and nonlinear textural features, and S-box integration with Chi2 and Relief feature selection, achieving 99.29% accuracy for channel-wise two-class classification. Su et al. [128] (2023) introduced projection dictionary pair learning (PDPL) with genetic algorithm optimization, reaching 64.34% accuracy for two classes and 49.01% for four classes. Aslan et al. [129] used a CNN with statistical features and discrete wavelet transform to improve classification accuracy.

5.2.3.2 Dataset-II Previous Studies:

Table 5.2 presents studies using Dataset II (LUMPED-2). Cimtay et al. [55] applied CNNs to fuse EEG-GSR data for high/low valence-arousal classification, improving accuracy, and used a decision tree to map emotional distances into seven emotions with weighted outputs. Alam et al. [130] employed datasets (LUMED-2, SWELL, YAAD) containing EEG, ECG, and GSR from emotional video sessions, extracting features to train Random Forest, SVM, KNN, and Decision Tree models for emotion classification.

5.3 Group Sparse and Super Resolution Time-Frequency-Based Method for Emotion Recognition

5.3.1 Methodology

The block diagram for the proposed methodology is displayed in the Figure 5.16.

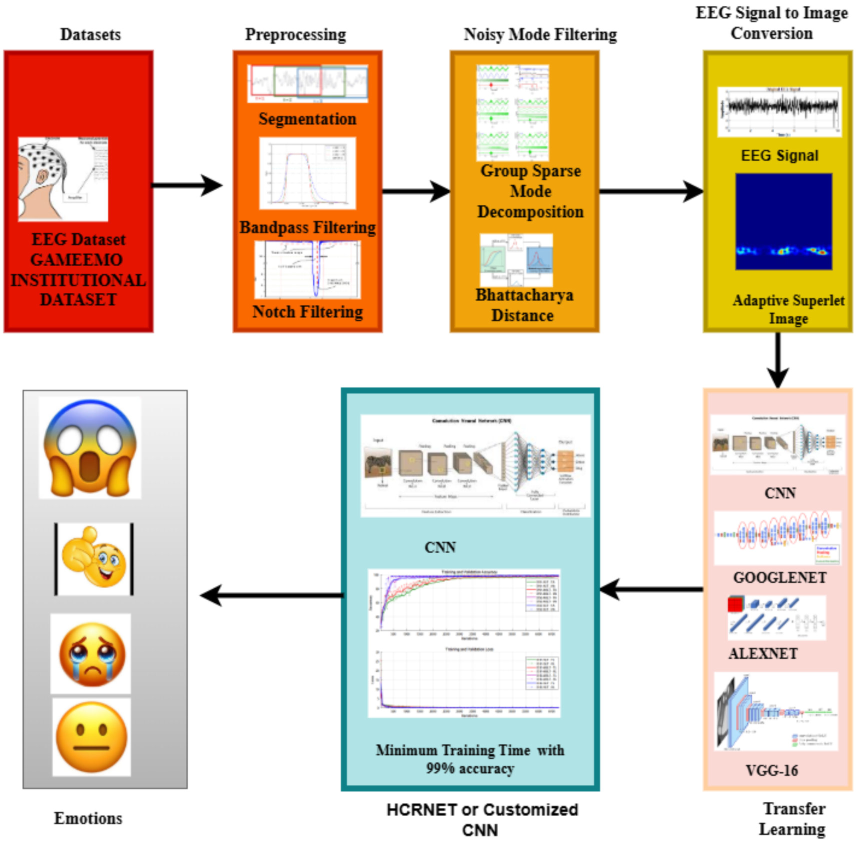


Figure 5.14: Block diagram of GSMD decomposition with Bhattacharyya filtering, superlet representation, and SRNet for emotion classification.

5.3.1.1 Group Sparse Mode Decomposition (GSMD)

Group sparse mode decomposition [131] decomposes a non-stationary signal into intrinsic mode functions (IMFs) with reduced and mutually independent bandwidths. The IMFs are computed using the ideal filter, with regularization enforced through a weighted penalty ℓ_{10}-norm. The weighting and regularization parameters are adaptively determined via short-period energy detection. Thus, GSMD decomposes the EEG signal $s(t)$ into IMFs and noise, expressed mathematically as [131]:

$$s(t) = \sum_{i=1}^{I} h_i(t) + n(t), \quad t = 1, \ldots, N \tag{5.16}$$

Where $h_i(t)$ represents the i_{th} component of signal $s(t)$, I represents the total number of decomposed components, N represents the number of sampling points, and $n(t)$ denotes the noise. The GSMD is based on the principle that each intrinsic mode function (IMF) has limited bandwidth in the discrete cosine transform (DCT) domain. The DCT, an orthogonal real transform using cosine functions, offers superior energy concentration compared to the discrete Fourier transform (DFT), enabling both coarse and fine sparsity. By applying the DCT, equation (5.16) is transformed from the time to the frequency domain as [131].

$$c_s = \sum_{i=1}^{I} c_{h_i} + c_v \tag{5.17}$$

where c_{h_i}, c_s, and c_v the DCT coefficient vectors of $h_i(t)$, $s(t)$, and $n(t)$, respectively.

In the DCT domain, an ideal filter is constructed to compute the IMFs. The zero-phase filter banks, defined for each frequency segment, are represented as sparse vectors of zeros and ones. The optimal filter bank is then obtained using weighted norm regularization, expressed as [131] :

$$\arg\min_{u} G(u) = \|c_s - c_s \circ u\|_2^2 + \alpha\|u\|_{0,\varphi} \tag{5.18}$$

Where $\|c_s - c_s \circ u\|_2^2$ represents the fidelity term, that maintains the data fidelity in the frequency domain and $\circ$ denotes the elementwise product of matrices. The term $\alpha\|u\|_{0,\varphi}$ in equation (5.18) causes group sparsity cutting the frequency spectrum into narrow-band signals with the optimal filter bank. Here, u represents the ideal filter bank and α represents the weighting parameter balancing the fidelity and regularization terms. Choosing the acceptable value of α is significant for designing the ideal filter bank. This equation can be represented as follows [131]:

$$\arg\min_{u} G(u) = (c_s[i])^2(1-u)^2 + \alpha\varphi_i I\{u\} \tag{5.19}$$

Where $I\{u\}$ represents the indicator function and φ_i represents a series of positive weighting coefficients. The condition of minimization of function $G(u)$ is as follows [131] :

$$u = \begin{cases} 1, & \text{if } (c_s[i])^2 \geq \alpha\varphi_i \\ 0, & \text{otherwise} \end{cases} \tag{5.20}$$

The ideal filter bank enables the recovery of every component by using the following equation [131]:

$$\hat{c}_i = c_s \circ u \tag{5.21}$$

The decomposed components are retrieved using the inverse DCT as follows [131]:

$$y(t) = \sum_{i=1}^{I} s_i(t) = \sum_{i=1}^{I} D^{-1}[\hat{c}_i] \tag{5.22}$$

Where $s_i(t)$ represents the decomposed components and $y(t)$ represents the recovered signals.

5.3.1.2 Bhattachrya Distance (BD)

The Bhattacharyya distance quantifies similarity between two probability distributions, where greater overlap indicates higher similarity and hence a smaller distance. It can be computed using variance between distributions. Let M and N denote the probability density variables; their variance is evaluated using equations (6.7) and (6.8) as [132]:

$$D(M) = E(M^2) - [E(M)]^2 \tag{5.23}$$

$$D(N) = E(N^2) - [E(N)]^2 \tag{5.24}$$

The variances of $D(M)$ and $D(N)$'s Bhattacharyya distances are defined as follows [132]:

$$BD(D(M), D(N)) = -\ln[BC(D(M), D(N))] \tag{5.25}$$

where BC denotes the discrete probability distributions' Bhattacharyya coefficient [132]:

$$BC(D(M), D(N)) = \sum_{m\in M, n\in N} D(M)D(N) \tag{5.26}$$

5.3.1.3 Filtering EEG Signals using GSMD Based on Improved BD

This study employs group sparse mode decomposition with Bhattacharyya distance to evaluate similarity or variance between the original signal and its decomposed IMFs. Significant and insignificant IMFs are identified, and the improved BD components are used for signal reconstruction. The detailed procedure is outlined in Algorithm 5.4.

Algorithm 5.4 : GSMD-Based EEG Signal Filtering using Bhattacharyya Distance

```
Load the EEG data from 'Dataset'.
Initialize the filtered signals matrix to zeros with the same size as the EEG data.
Get the number of signals and samples from the EEG data.
for k = 1 to numSignals do
    Get the current signal $s(t)$ from the EEG data.
    Perform the decomposition using GSMD with specified parameters to obtain IMFs.
    Construct MRA matrix by appending IMFs.
    Initialize Bhattacharyya distance array.
    for $i = 1$ to number of IMFs do
        Compute Bhattacharyya distance between each IMF and the original signal.
    end for
    Sort IMFs based on Bhattacharyya distance.
    Define levelForReconstruction to select components based on threshold Bhattacharyya distance.
    Set the number of components to select.
    Select the components with the minimum Bhattacharyya distance.
    Sum the selected multiresolution signals to obtain the filtered signal $y(t)$.
    Store the filtered signal in the filtered signals matrix.
end for
```

5.3.1.4 Deeplearning Models For Classification

A deep learning model consists of multiple layers of neurons that sequentially process data, with each neuron performing computations based on activations from the previous layer. A basic network includes an input layer, hidden layers, and an output layer. Models learn by training, adjusting synaptic weights. Key parameters include the learning rate, which controls weight updates, and dropout, which prevents overfitting. Activation functions like ReLU (activates positive values) and Softmax (for class predictions) enhance network performance.

5.3.2 Results and Discussion

5.3.2.1 Noisy Modes Filtering

This section compares signal decomposition and noise filtering using multiple algorithms, producing intrinsic mode functions (IMFs) representing high, moderate, and low frequencies. Group sparse mode decomposition [131] is compared with variational mode decomposition [132], tunable Q wavelet transform [133], and maximum overlap discrete wavelet transform (MODWT) [134].

Algorithm 5.5 : EEG-Based Emotion Detection using SLT and ASLT

```
Input: Filtered EEG signals from datasets I and II.
Output: Classification of EEG signals [4 Emotions]
Step 1: Generation of Spectrogram using SLT and ASLT
    1. Apply the filtered EEG signal.
    2. Assign mother wavelet parameters such as central frequency ($f_0$), time spread parameter ($W_n$):
       $$\Psi_{f_0,n}(t) = \frac{1}{W_n\sqrt{2\pi}} e^{-\frac{t^2}{2W_n^2}} e^{j2\pi f_0 t}$$
    3. Selection of SL by taking multiple wavelets:
       $$SL_{f_0,w} = \left\{\Psi_{f_0,n} \mid n = n_1, n_2, \dots, n_w\right\}$$
    4. Determine the number of cycles in the wavelet using the multiplicative rule:
       $$n_i = n \times i$$
    5. Generation of spectrogram by taking the geometric mean of each wavelet response:
       $$R\left[\Psi_{f_0,n_i}\right] = \sqrt{2} \cdot s(t) * \Psi_{f_0,n_i}$$
    6. Generation of ASLT spectrogram using equation(4.15)
Step 2: Apply various deeplearning models to the 2-D spectrogram picture data as input.
Step 3: Extract characteristics from the EEG signals and categorize them into four emotional groups using the deep-learning models (SRNET, VGG-16, GoogleNet, and AlexNet).
```

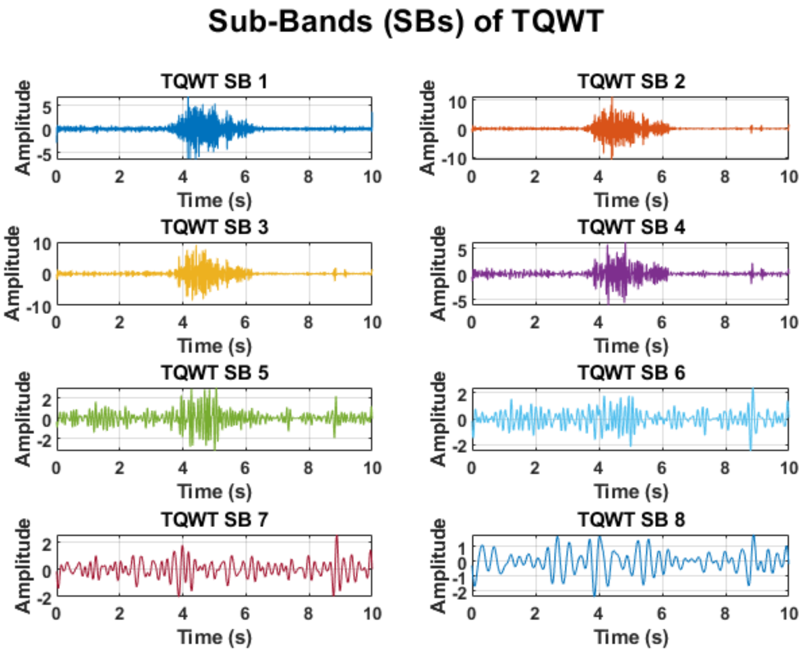


Figure 5.15: The EEG signal is decomposed into sub-bands using TQWT.

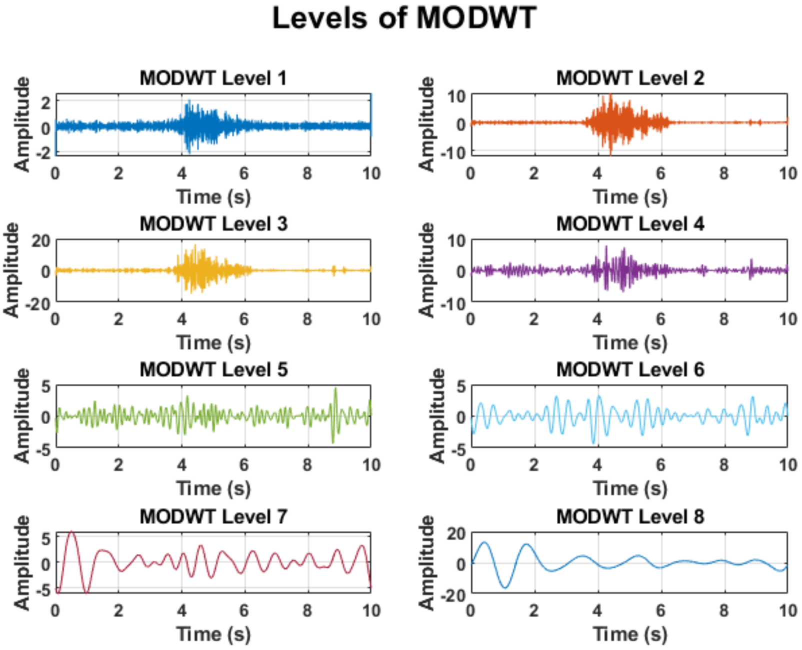


Figure 5.16: The EEG signal is decomposed into levels using MODWT.

TQWT uses eight levels with a quality factor of 1, MODWT uses a ’sym-4’ wavelet with eight levels, and VMD parameters are absolute tolerance 5×10^{-6}, relative tolerance 0.005, maximum iterations 500, number of IMFs 8, penalty factor 1000, LMUpdate rate 0.01, and peak initialization. GSMD parameters are reference SNR 10 dB, $\alpha = 1 \times 10^{-4}$, and maximum and minimum iterations set to 8 and 1, respectively, ensuring stable convergence and sensitive detection of relevant frequency components while avoiding spurious modes and excessive computation.

The four signal-decomposition methods decompose the EEG signal, as shown in Figures 5.15–5.18 using the specified tuning parameters. Mode tuning and selection remove noise, with lower Bhattacharyya distance indicating higher similarity; the improved Bhattacharyya distance determines the number of selected modes (Figure 5.19). Figures 5.15–5.17 show that early modes have high frequency and later modes have low frequency, whereas GSMD (Figure 5.18) produces low-frequency signals in early modes and high-frequency signals in later modes. GSMD modes are smoother than those of other methods, demonstrating effective noise filtering during decomposition.

Figures 5.19 and 5.20 show the Bhattacharyya distance between individual modes and the original signal, while Figure 5.21 presents the original and reconstructed EEG signals using the improved BD. In TQWT, the last five modes have higher BD than the first three. For MODWT and VMD, the first mode exhibits the highest BD, and the last mode the lowest. In GSMD, the first mode has lower BD, whereas the other modes have higher BD. Average BD values are 0.8 for TQWT, 0.6 for MODWT, slightly below 0.6 for VMD, and 0.5 for GSMD, indicating that GSMD’s recombined sub-bands are most similar to the original signal.

In mode analysis, a mode’s central frequency indicates where its spectral energy is concentrated, with higher-frequency components potentially containing noise. Modes with higher BD, corresponding to these frequencies, are removed during EEG reconstruction. Figure 5.20 shows that after removing high-BD modes, the average BD between the modes and the original

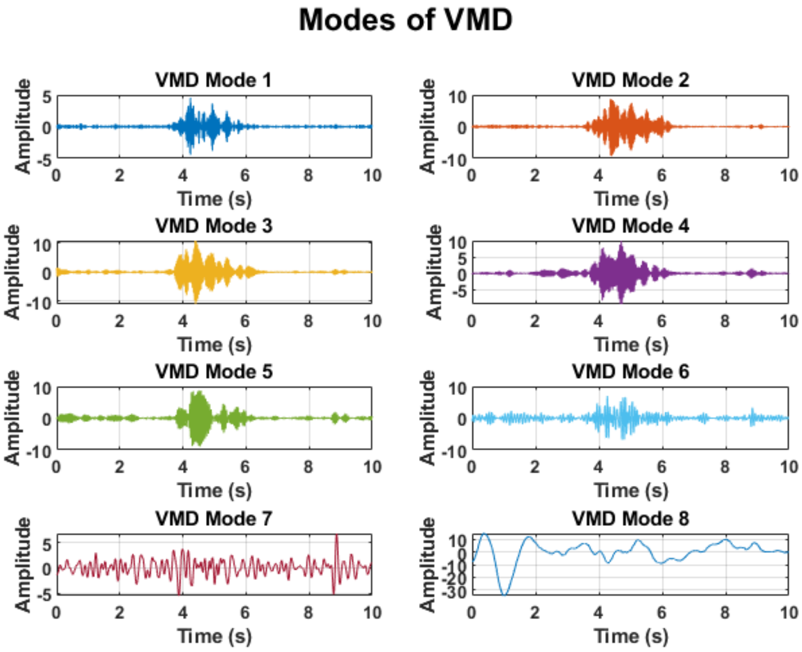


Figure 5.17: The EEG signal is decomposed into modes using VMD.

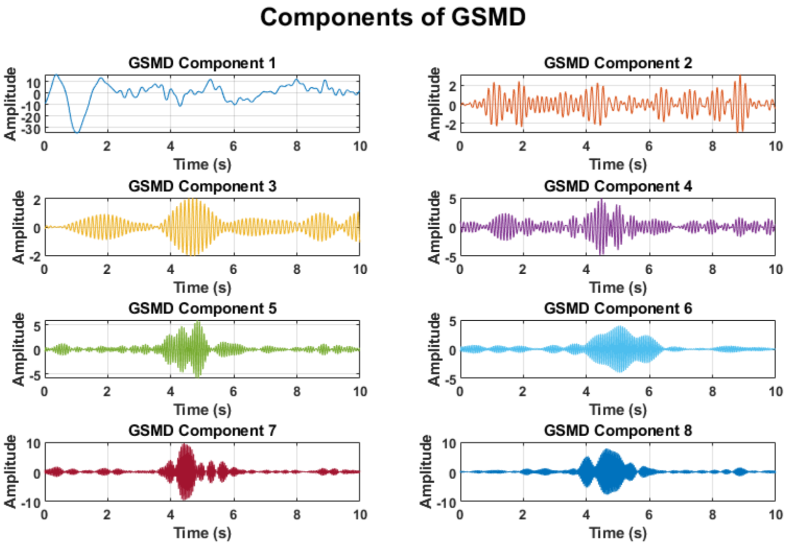


Figure 5.18: The EEG signal is decomposed into components using GSMD.

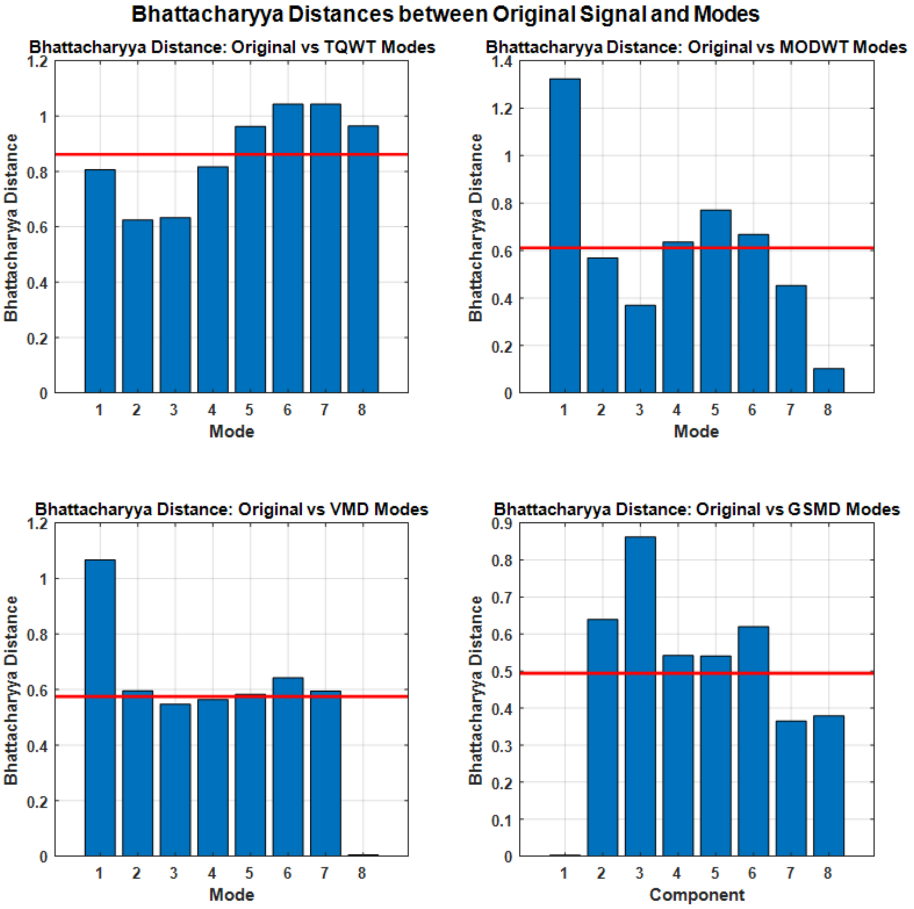


Figure 5.19: Bhattacharyya distances between the original EEG signal and its decomposed modes.

signal improves when selected modes are used for reconstruction.

The signal-to-noise ratio (SNR) quantifies the extent of noise distortion in a signal. In signal decomposition, the original signal is split into component modes, and the reconstructed signal is obtained by summing the selected modes based on the improved BD. SNR is defined as follows:

$$\text{SNR (dB)} = 10 \cdot \log_{10}\left(\frac{\text{var}(s(t))}{\text{var}(s(t) - s(t)_{\text{reconstructed}})}\right) \tag{5.27}$$

Let $s(t)$ and $s(t)_{\text{reconstructed}}$ denote the original and reconstructed signals. The SNRs for TQWT, MODWT, VMD, and GSMD are 0.909, 5.507, 9.492, and 12.980dB, respectively, with GSMD achieving the highest SNR. Figure 5.21 illustrates that GSMD reconstruction is more robust to noise than other methods, making it suitable for EEG filtering. Algorithm 5.4 details the step-by-step procedure for filtering signals from datasets I and II.

5.3.2.2 Filtered EEG Signal to Image Conversion

The noise-filtered EEG data are transformed into images using time-frequency resolution (TFR) functions, specifically superlet and adaptive superlet transforms, which capture the dynamic frequency content of EEG signals to analyze brain activity. For datasets I and II, the TFR functions use sampling frequencies of 128 Hz and 256 Hz, respectively.

Within the TFR function, a frequency range of 0.4–48 Hz and six cycles are defined,

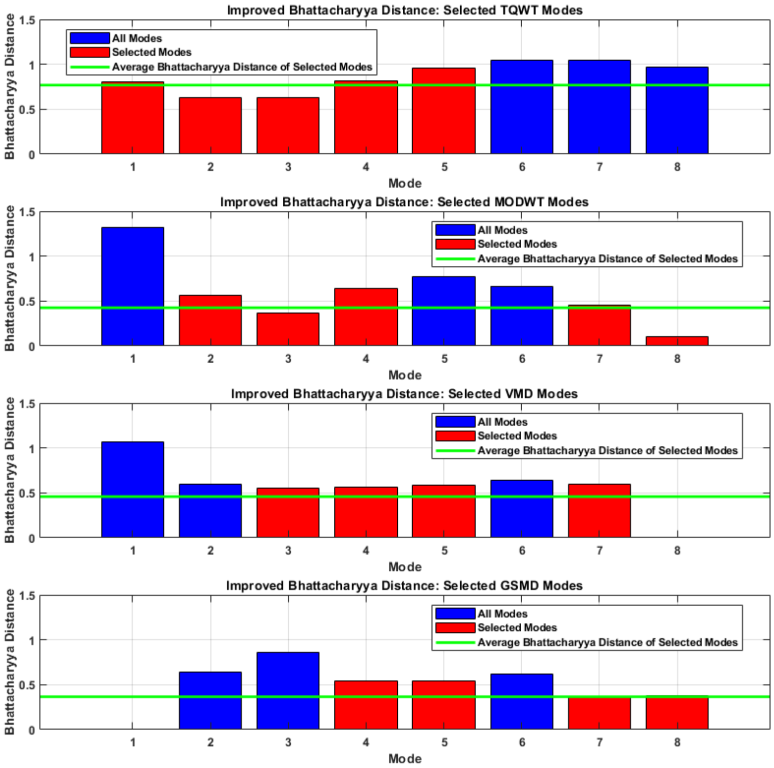


Figure 5.20: Bhattacharyya distance improvement after selecting optimal modes.

affecting time-frequency resolution, while the order parameter uses default values. The time range is fixed at 10 s based on sample size and frequency. Time-frequency images (224×224 pixels) are saved for analysis. Randomly selected filtered EEG signals converted into ASLT images, with the conversion process detailed in Algorithm 5.5. These images are classified using GoogLeNet, AlexNet, VGG-16, and SRNET.

5.3.2.3 SRNET Architecture

The proposed SRNET is a lightweight deep learning model for classifying emotional states from superlet and adaptive superlet time-frequency EEG images. Its architecture is optimized to balance high classification performance with low computational complexity. A detailed explanation of SRNET follows.

(a) Input Layer

SRNET's input layer accepts EEG signals transformed into SLT and ASLT images of size 224×224×3 (RGB, single image per batch) with zero-center normalization, allowing the network to utilize both spectral and spatial features encoded in the images.

(b) Convolutional Block 1

The convolutional layer uses eight 3×3 filters with "same" padding, producing feature maps of size 224×224×8 with learnable weights (3×3×3×8) and biases (8). Batch normalization normalizes each channel using learnable offsets and scales to reduce internal covariate shift.

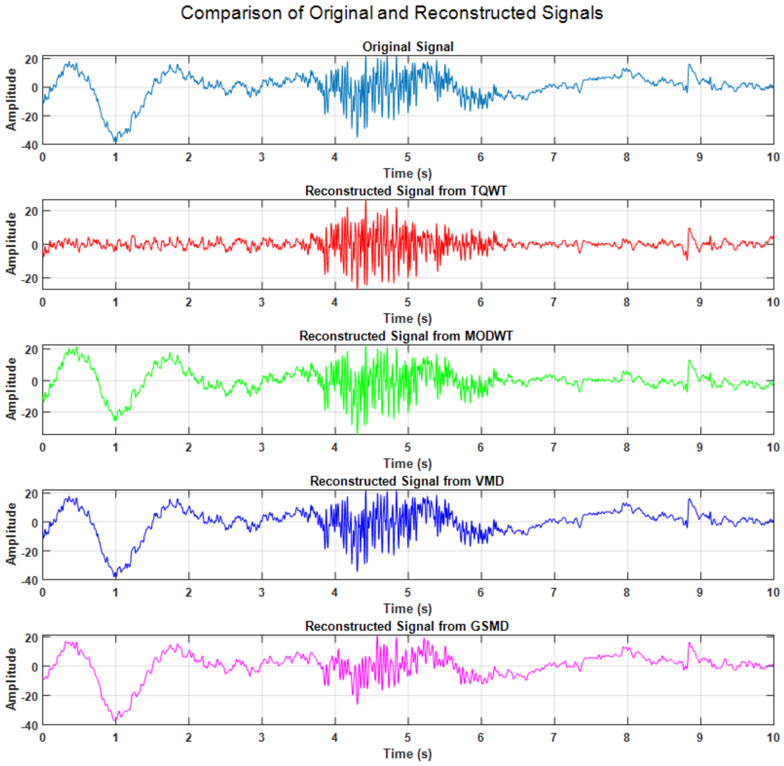


Figure 5.21: Comparison between the original signal and its reconstructed modes.

ReLU introduces non-linearity, while max pooling (2×2, stride 2) halves spatial dimensions to 112×112×8, enhancing translation invariance and reducing computation. A dropout layer with a 0.2 rate prevents overfitting by randomly deactivating neurons during training.

(c) Convolutional Block 2

The convolutional layer uses sixteen 3×3 filters with ”same” padding, producing feature maps of size 112×112×16. Batch normalization stabilizes learning with learnable scale and offset for each channel. ReLU maintains non-linearity, and max pooling (2×2, stride 2) reduces dimensions to 56×56×16 while preserving key features. A dropout layer with a 0.2 rate improves generalization.

(d) Fully Connected and Output Layers

The fully connected layer outputs a 4-element vector for the four emotional classes. The Softmax layer produces a normalized probability distribution across classes, and the classification layer uses cross-entropy loss to predict the final emotion class.

5.3.2.4 Training Configuration and Optimization Strategy

EEG signals are converted into superlet and adaptive superlet spectrograms and organized in a labeled image datastore by emotion. Data is split into 70% training, 15% validation, and 15% testing (subject-independent). The model is trained using the Adam optimizer with an initial learning rate of 0.001 and piecewise scheduling (reducing by 0.1 every 10 epochs) to improve convergence. Training uses a batch size of 64 for memory efficiency and gradient stability, with a maximum of 50 epochs. Dropout layers (rate 0.2) follow each convolutional block to prevent overfitting, and early stopping is applied if validation accuracy does not improve after ten epochs.

5.3.2.5 Proposed SRNET Numerical Parameters

The proposed SRNET is a custom-built, end-to-end trainable network for classifying high-resolution superlet-based EEG spectrograms. Its architecture supports residual learning and requires low computational power, making it suitable for real-time applications. The feature extraction pipeline includes convolution, normalization, dropout, and pooling to capture emotion-specific temporal-spatial-frequency features. Training hyperparameters and layer details are provided in Table 5.3 and Table 5.4. The network's input layer accepts images of size 224×224×3, with the first layer extracting features from these inputs.

Table 5.3: Training hyperparameters used for deep neural network model optimization.

Hyperparameter	**Value**
Optimizer	Adam
Max Epochs	50
Mini-Batch Size	64
Initial Learning Rate	10^{-3}
Learning Rate Schedule	Piecewise
Learning Rate Drop Factor	0.1
Learning Rate Drop Period	20
L2 Regularization	0.0005
Execution Environment	Auto

Table 5.4: Layer-wise details of the proposed SRNet.

Layer Number	**Layer Name**	**Activations**	**Parameters**	**Total Learnable**
1	imageInput	224×224×3	—	0
2	Conv1	224×224×8	Weights: 3×3×3×8, Bias: 1×1×8, Stride: [1 1], Padding: 'same'	224
3	BN1	224×224×8	Offset: 1×1×8, Scale: 1×1×8	16
4	ReLU1	224×224×8	—	0
5	MaxPool1	112×112×8	Pool Size: [2 2], Stride: [2 2]	0
6	Dropout1	112×112×8	Dropout Rate: 0.2	0
7	Conv2	112×112×16	Weights: 3×3×8×16, Bias: 1×1×16, Stride: [1 1], Padding: 'same'	1168
8	BN2	112×112×16	Offset: 1×1×16, Scale: 1×1×16	32
9	ReLU2	112×112×16	—	0
10	MaxPool2	56×56×16	Pool Size: [2 2], Stride: [2 2]	0
11	Dropout2	56×56×16	Dropout Rate: 0.2	0
12	FC1	1×1×4	Weights: 4×1568, Bias: 1×4	6288
13	Softmax	1×1×4	—	0
14	Classification	—	—	0

The proposed architecture efficiently extracts and classifies the features.

5.3.2.6 Performance Comparison of GoogleNet, AlexNet, VGG-16, and SRNET

The study classifies four emotional classes for Dataset-I (Game1–Game4) and Dataset-II (fear, happy, relax, sad), using 28,224 and 39,520 images, respectively, split 70:5:15 for training, validation, and testing. Pre-trained CNNs (GoogleNet, AlexNet, VGG16) and SRNET are trained with the same parameters on superlet and adaptive superlet images. Performance metrics, including accuracy, sensitivity, specificity, precision, and F1-score, range from 97% to 99.97%, showing robustness. To prevent overfitting, the SRNET model included various regularization techniques, including Adam optimization, L2 regularization (0.0005), early stopping, and adaptive learning-rate scheduling (initial learning rate of 0.001, with a drop factor of 0.1 every 20 epochs). The training was constrained to 50 epochs with a mini-batch size of 64 to ensure consistent convergence and prevent overfitting. Early stopping was used to terminate training when validation performance stopped enhancing, thereby avoiding the recollection of training information. Additionally, the model consistently performed well across several datasets (around 99%), suggesting high generalization as compared to overfitting to a specific dataset. ASLT images improve performance for all classifiers; AlexNet and VGG-16 achieve higher metrics but with longer training, while SRNET has the lowest training time after 20 epochs.

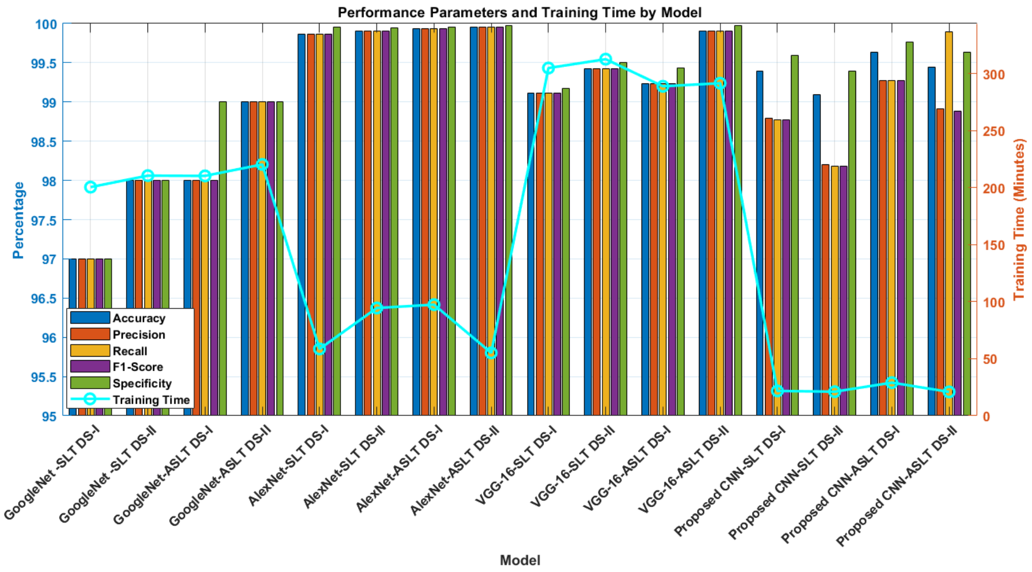


Figure 5.22: Comparison of performance parameters and training time by different deep-learning models.

5.3.2.7 Deep Learning Models Training Time Comparison

Table 5.5 shows that SRNET greatly outperforms other models in training efficiency. For DS-I, VGG-16 takes 304 min 12 s, whereas SRNET completes training in 21 min 47 s, achieving roughly a 93% reduction in time. Similar trends are observed across all datasets.

5.3.2.8 Performance Measurement of Proposed SRNET

The proposed SRNET achieves the lowest training time due to its fewer layers and parameters. Its validation (VA) and testing accuracy (TA) for Datasets I and II are shown in

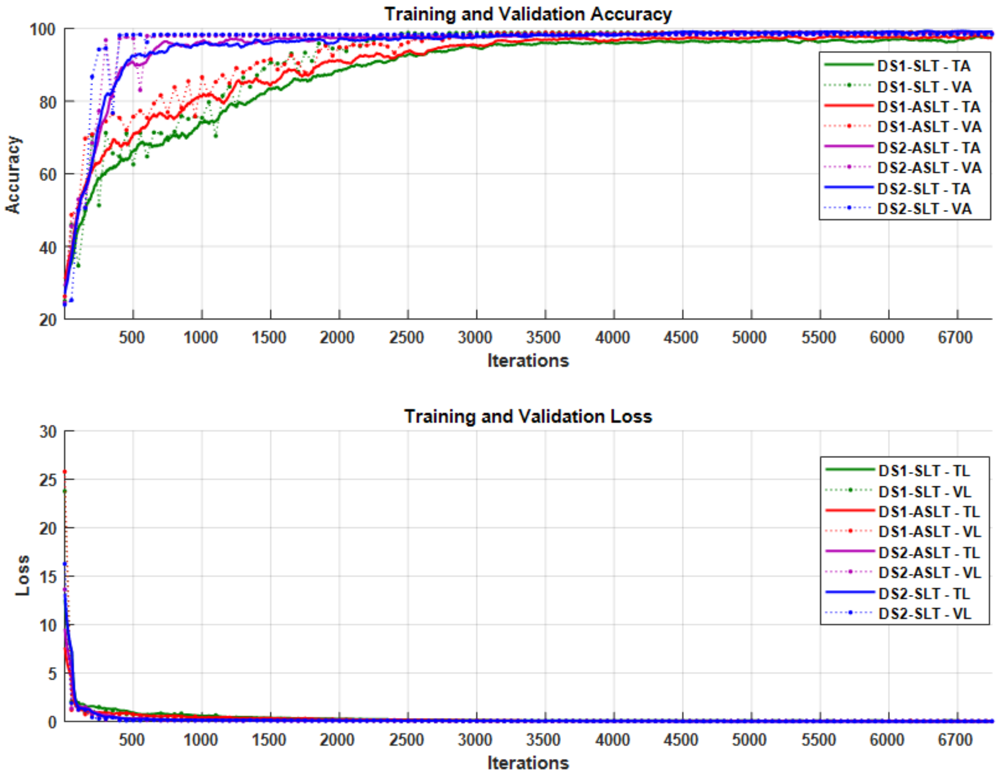


Figure 5.23: Comparison of validation and training accuracy and validation and training loss.

Table 5.5: Comparison of the training time of different deep neural networks.

Model	**SLT Training Time**		**ASLT Training Time**	
	Dataset I	**Dataset II**	**Dataset I**	**Dataset II**
GoogleNet	204 Min 23 Sec	211 Min 12 Sec	211 Min 14 Sec	220 Min 12 Sec
AlexNet	58 Min 58 Sec	94 Min 49 Sec	97 Min 43 Sec	55 Min 14 Sec
VGG-16	304 Min 12 Sec	311 Min 26 Sec	278 Min 51 Sec	280 Min 31 Sec
Proposed SRNET	21 Min 47 Sec	23 Min 51 Sec	28 Min 54 Sec	22 Min 34 Sec

Table 5.6 and Figure 5.22, demonstrating reliable emotion classification. Performance metrics—including accuracy, sensitivity, specificity, precision, F1-score, Cohen's kappa, MCC, and AUC—fall within the ranges: 99.09–99.63, 98.18–99.27, 99.39–99.76, 98.20–99.27, 0.9757–0.9902, 0.9758–0.9903, and 0.9865–0.9951, respectively. Cohen's kappa values above 0.97 indicate perfect agreement, MCC values between 0.9758 and 0.9903 confirm accurate predictions, and AUC values over 0.9865 demonstrate excellent class discrimination.

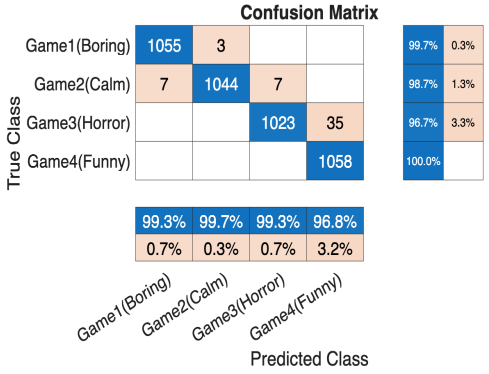
5.3.2.9 Confusion Matrix

The confusion matrices for test set images of Datasets I and II demonstrate the effectiveness of the proposed method in emotion classification. SLT image classifications are shown in Figures 5.24a and 5.24c, while ASLT image classifications are in Figures 5.24b and 5.24d. All individual classes achieve over 96% accuracy, with SLT images in Game4 (funny) and Happy classes reaching 100%. ASLT images exhibit higher classification accuracy than SLT images.

Table 5.6: Performance metrics of the proposed SRNET model across both datasets (DS-I and DS-II) for SLT and ASLT methods.

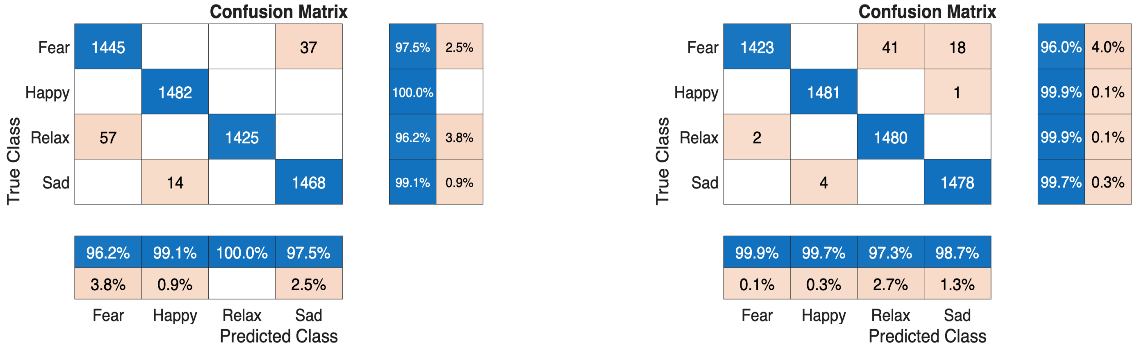
Metric	Dataset-I				Dataset-II			
	SLT		ASLT		SLT		ASLT	
	VA	TA	VA	TA	VA	TA	VA	TA
Accuracy (%)	99.43	99.39	99.56	99.63	99.17	99.09	99.39	99.44
Sensitivity (%)	98.87	98.77	99.13	99.27	98.35	98.18	98.68	98.89
Specificity (%)	99.62	99.59	99.71	99.76	99.45	99.39	99.56	99.63
Precision (%)	98.88	98.79	99.13	99.27	98.36	98.20	98.70	98.91
F1-Score (%)	98.87	98.77	99.13	99.27	98.35	98.18	98.68	98.88
Cohen's Kappa	0.9849	0.9836	0.9884	0.9902	0.9780	0.9757	0.9825	0.9852
MCC	0.9845	0.9837	0.9884	0.9903	0.9780	0.9758	0.9825	0.9852
AUC	0.9865	0.9918	0.9942	0.9951	0.9879	0.9879	0.9912	0.9926

VA = Validation Accuracy, TA = Testing Accuracy, SLT = Superlet Transform, ASLT = Adaptive Superlet Transform, MCC = Matthews Correlation Coefficient, AUC = Area Under Curve



(a) CM of SLT (Dataset-I)

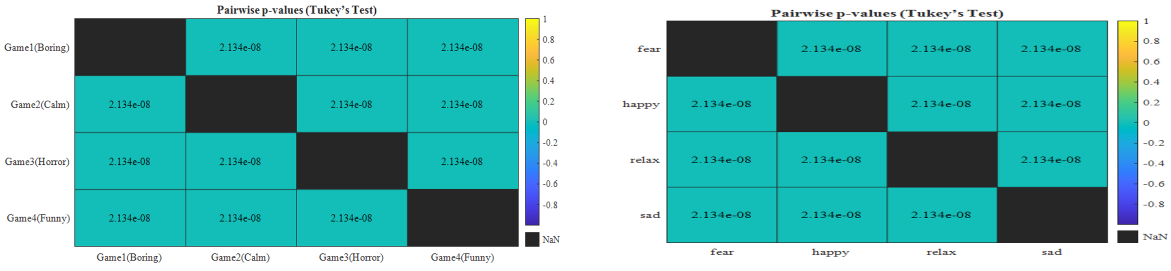
(b) CM of ASLT (Dataset-I)



(c) CM of SLT (Dataset-II)

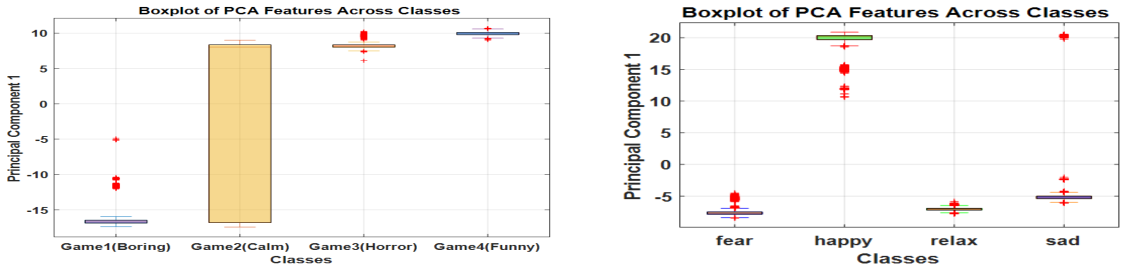
(d) CM of ASLT (Dataset-II)

Figure 5.24: Confusion matrices of GSMD plus Bhattacharyya plus Superlet, and SRNet based emotion classification (Datasets I and II).



(a) P-Values between classes in DS-I

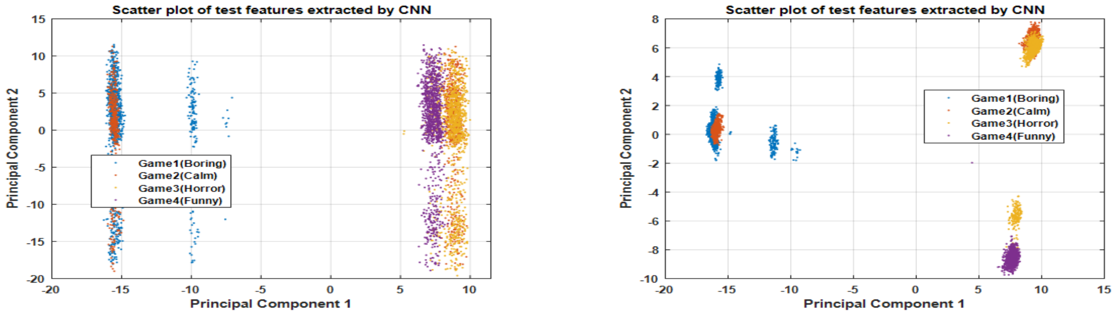
(b) P-Values between classes in DS-II



(c) Box-Plot of features in DS-I

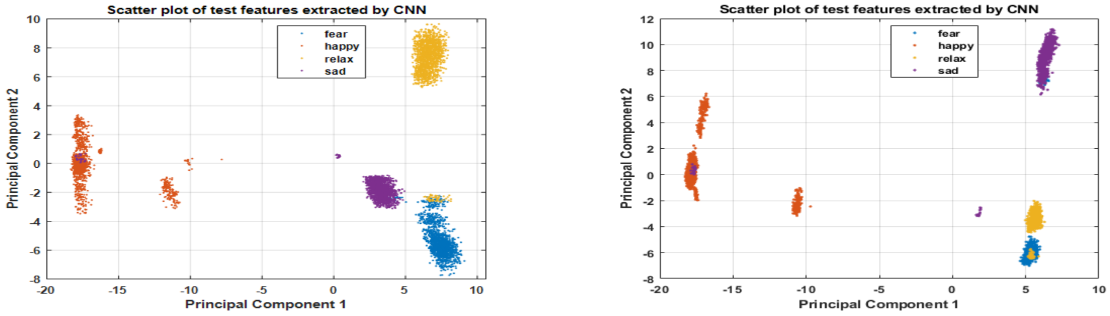
(d) Box-Plot of features in DS-II

Figure 5.25: P-Values and Box-Plots of features from Dataset I and II.



(a) SLT in DS-I

(b) ASLT in DS-I



(c) SLT in DS-II

(d) ASLT in DS-II

Figure 5.26: t-SNE visualizations of SLT and ASLT features showing separation of four emotion clusters in Datasets I and II

Table 5.7: Computational complexity of the proposed emotion classification framework.

Stage	**Computational Complexity**	**Time (s) DS-I**	**Time (s) DS-II**
GSMD Filtering	$O(S \cdot M \cdot T)$	0.6129 Sec.	0.7536 Sec.
Superlet Image	$O(F \cdot T)$	0.6266 Sec.	0.7055 Sec.
DNN Training	$O(E \cdot \frac{N}{B} \cdot P)$	1307 Sec	1431 Sec
DNN Inference	$O(P)$	14.7578 Sec	10.5576 Sec
Evaluation Metrics	$O(N + C)$	0.0376 Sec	0.1506 Sec
Total		1323.0349 Sec	1443.1673 Sec

5.3.2.10 Features Plots

Figures 5.25, 5.25a, and 5.25b show that the p-values of features from four classes are below 0.05, rejecting the null hypothesis, while Figures 5.25c and 5.25d display distinct boxplot medians, indicating class discrimination. t-SNE is used to assess EEG feature quality and class separability. High-dimensional features are extracted from training and testing images, and PCA identifies the most significant features by variance. Figures 5.26a, 5.26b (DS-I) and Figures 5.26c, 5.26d (DS-II) show t-SNE plots, demonstrating clear discrimination among all four classes.

5.3.2.11 Computational Complexity

The proposed EEG-based emotion recognition pipeline—comprising GSMD filtering, Superlet image generation, and SRNET classification—is computationally efficient. GSMD filtering has complexity $O(S \cdot M \cdot T)$, with efficiency improved by small signal lengths and well-defined decomposition parameters. Superlet transform generates time-frequency images with complexity $O(F \cdot T)$, while image display and storage incur minimal overhead. SRNET training has complexity $O(E \cdot \frac{N}{B} \cdot P)$, accelerated by GPU execution and early stopping, and inference requires $O(P)$ per image. Evaluation metrics, including confusion matrix and F-score, are computed in linear time $O(N + C)$. Overall, the pipeline is suitable for large-scale, real-time applications.

5.3.3 Comparison of Previous Studies

(a) Previous studies with dataset-I

Table 5.8 summarizes prior studies on Dataset-I. Alakus et al. (2020) [35] used MLPNN to classify four classes with 73.21% accuracy, while k-NN and SVM achieved 42.85% and 55%, respectively; for two-class classification, accuracies were 81.3% (MLPNN), 72.5% (SVM), and 66% (k-NN). Alakus and Turkoglu [125] (2020) employed bidirectional LSTM on spectral entropy features, achieving 76.91% accuracy for two-class classification. Aslan et al. [126] (2022) converted EEG signals to scalogram images using CWT and applied GoogLeNet; k-NN, SVM, and ELM classifiers achieved two-class accuracies of 98.53%, 98.41%, and 98.78%, respectively. Tuncer et al. [127] (2022) utilized LEDPatNet 19 with 18 TQWT sub-bands and iterative Chi2 and ReliefF feature selection, achieving 99.29% accuracy for channel-wise two-class classification. Su et al. [128] (2023) proposed PDPL optimized via evolutionary algorithms, achieving 64.34% and 49.01% accuracies for two- and four-class

Table 5.8: Performance comparison with prior studies on both datasets.

Study (Year)	**Method / Features**	**NC**	**Classifier**	**Accuracy (%)**
Dataset-I Studies				
Alakus et al. [35] (2020)	Hjorth, Entropy, Variance, Zero Crossing	4	k-NN, SVM, MLPNN	42.85, 55, 73.21
Alakus et al. [35] (2020)	Hjorth, Entropy, Variance, Zero Crossing	2	k-NN, SVM, MLPNN	66, 72.5, 81.35
Alakus & Turkoglu [125] (2020)	Spectral Entropy	2	Bi-LSTM	76.93
Aslan et al. [126] (2022)	EEG scalogram + GoogleNet features	2	k-NN, SVM, ELM	98.53, 98.87, 98.41
Tuncer et al. [127] (2022)	LEDPatNet19 fused features	2	Cubic SVM	99.29
Su et al. [128] (2023)	GA-PDPL	4, 2	GA-PDPL	49.01, 64.34
M. Aslan et al. [129] (2024)	Energy, Variance, Skewness	4	CNN	95.30
Proposed (Dataset-I)	**GSMD, SLT, ASLT**	**4**	**Proposed SR-NET**	**SLT: 99.39, ASLT: 99.63**
Dataset-II Studies				
Bajaj et al. [32] (2018)	FAWT	4	Weighted k-NN	86.1
Krishna et al. [33] (2019)	TQWT	4	ELM	87.1
Bajaj et al. [34] (2019)	EMD-VMD	4	MC LS-SVM	90.63
Khare & Bajaj (2021a) [135]	SPWVD	4	CNN	93.01
Khare et al.[57] (2021a)	ATQWT	4	MC LS-SVM	95.70
Khare & Bajaj et al. [136] (2021b)	OVMD	4	ELM	97.24
Proposed (Dataset-II)	**GSMD, SLT, ASLT**	**4**	**Proposed SR-NET**	**SLT: 99.09, ASLT: 99.39**

classification. Aslan et al. [129] applied DWT and statistical features with CNN, achieving 95% accuracy.

(b) Previous studies with dataset-II

Table 5.8 summarizes key studies using Dataset II. Bajaj et al. [32] (2018) applied weighted k-NN with flexible analytic wavelet transform, achieving 86.10% accuracy. Krishna et al. [33] (2019) employed ELM classifiers with TQWT, achieving an accuracy of 87.1%. Taran and Bajaj [34] (2019) employed MC-LS SVM with VMD and EMD, reaching 90.63%. Khare and Bajaj [137] (2021a) achieved 93.01% accuracy using CNN with SPWVD, while Khare et al. [57] (2020) obtained 95.70% using MCLS-SVM with adaptive TQWT. Khare and Bajaj [136] (2021b) reported 96.33%, 96.43%, and 97.24% using different ELM kernels, EVCM, and O-VMD. The proposed framework employing GSMD, SLT, ASLT, and SRNET outperforms these methods, achieving 99.63% accuracy on both datasets.

5.4 Summary

The study used two specialized methods to develop an optimal mode filter framework with the objective of eliminating EEG noise. The first utilizes joint time–frequency scattering to create a deformation-resistant filter that extracts clean neural rhythms and region-specific activations. The second approach directly employs a group sparse mode decomposition framework, which acts as an optimal mode filter by using selective reconstruction and superlet processing to suppress artifacts while preserving salient neural dynamics. Combined, these methods demonstrate the core principle of adaptive mode filtering, producing cleaned signal representations that enable highly discriminative feature extraction for subsequent analysis. Building upon this foundation, the next research objective focuses on optimization in wavelet parameter selection using nature-inspired algorithms.

Chapter 6

Optimal Wavelet Design using Nature-Inspired Algorithms

6.1 Introduction

Emotion plays a vital role in human communication, decision-making, and daily life, and its recognition has become increasingly significant in domains such as brain–computer interfaces, healthcare, and entertainment systems [127, 138]. Among various modalities, EEG has emerged as a reliable tool for emotion recognition because it provides a non-invasive, cost-effective, and temporally precise measure of brain activity. Nevertheless, the low amplitude of EEG signals and their susceptibility to noise introduce challenges for machine learning-based classification, making effective preprocessing and feature extraction essential.

A wide range of signal decomposition techniques has been explored for extracting discriminative features from EEG. Conventional approaches such as the discrete wavelet transform (DWT), wavelet packet transform (WPT), stationary wavelet transform (SWT), tunable Q factor wavelet transform, flexible analytic wavelet transform (FAWT), and variational mode decomposition have been widely applied [51, 139, 140, 33, 32, 141, 142, 136]. These methods decompose EEG signals into frequency bands or sub-bands from which statistical, entropy, or energy-based features can be derived and subsequently classified using machine learning models such as k-NN, SVM, Random Forests, and ELM. While effective, most wavelet-based methods rely heavily on empirically chosen decomposition parameters, which may cause redundancy, mode mixing, or suboptimal signal representations [53, 54].

To overcome these limitations, recent research has focused on integrating wavelet-based methods with optimization algorithms to achieve adaptive parameter selection.

Collectively, these advances illustrate the importance of combining decomposition methods with nature-inspired optimization algorithms. By addressing the challenges of empirical parameter selection, redundancy, and emotion-specific channel dependence, approaches such as FAWT–PSO, TQWT–COA, and WST–PSO offer more reliable, accurate, and interpretable EEG-based emotion recognition systems. Such optimization-driven frameworks represent a promising direction for developing robust methodologies capable of generalizing across diverse datasets and emotional states.

6.2 Adaptive Flexible Analytic Wavelet Transform for EEG-Based Emotion Recognition

6.2.1 Methodology

The methodology of the proposed is displayed in Figure 6.1 using a graphical extract.

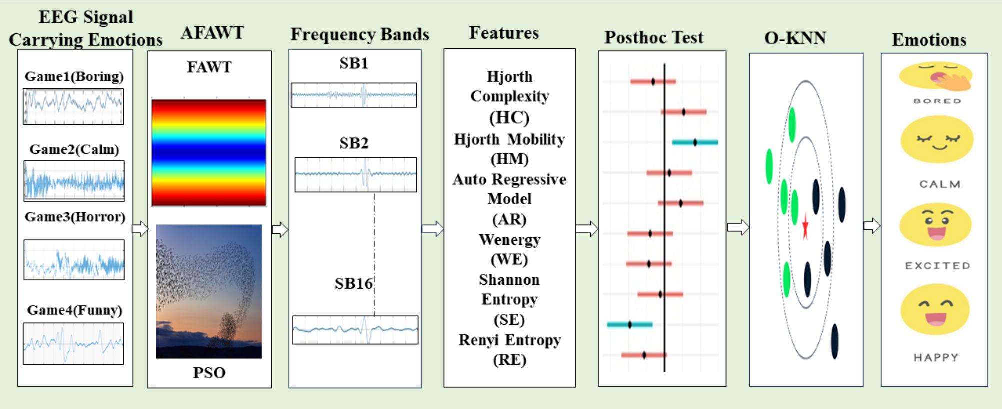


Figure 6.1: Block diagram of the proposed methodology for EEG emotion classification using AFAWT.

6.2.1.1 Flexible Analytic Wavelet Transform (FAWT)

FAWT [143] is a signal processing algorithm that decomposes the signal into different frequency sub-bands with minimal information loss. FAWT consists of one low-pass (LP) and two high-pass (HP) channels at each stage. The high-pass channel uses pairs of atoms in the Hilbert transform for negative and positive frequency analysis. The complex structure of FAWT allows more flexibility in selecting the parameters quality factor (Q), redundancy factor, and dilation factor. The HP and LP channels filter frequency responses, which are represented by the symbols M($\hat{\omega}$) and N($\hat{\omega}$). The following expression is given as the frequency response for a low-pass filter [143] :

$$N(\hat{\omega}) = \left\{ \begin{array}{c} \sqrt{s_1 s_2}, \mid \hat{\omega} \mid < \hat{\omega}_m \\ \sqrt{s_1 s_2}\theta\frac{(\hat{\omega}-\hat{\omega}_m)}{(\hat{\omega}_n-\hat{\omega}_m)}, \hat{\omega}_m \leq\mid \hat{\omega} \mid\leq \hat{\omega}_n \\ \sqrt{s_1 s_2}\theta\frac{(\pi-\hat{\omega}+\hat{\omega}_m)}{(\hat{\omega}_n-\hat{\omega}_m)}, -\hat{\omega}_m \leq\mid \hat{\omega} \mid\leq -\hat{\omega}_n \\ 0, \mid \hat{\omega} \mid\geq \hat{\omega}_n \end{array} \right\} \tag{6.1}$$

where,

$$\hat{\omega}_m = \frac{(1-\beta)\pi + \epsilon}{s_1}; \quad \hat{\omega}_n = \frac{\pi}{s_2}$$

Similarly, for high-pass filter, the frequency response is given by [143] :

$$M(\hat{\omega}) = \left\{ \begin{array}{c} \sqrt{2 s_3 s_4}\theta\frac{(\pi-\hat{\omega}-\hat{\omega}_a)}{(\hat{\omega}_b-\hat{\omega}_a)}, \hat{\omega}_a \leq\mid \hat{\omega} \mid\leq -\hat{\omega}_b \\ \sqrt{2 s_3 s_4}, \hat{\omega}_b \leq\mid \hat{\omega} \mid\leq -\hat{\omega}_c \\ \sqrt{2 s_3 s_4}\theta\frac{(\hat{\omega}-\hat{\omega}_c)}{(\hat{\omega}_d-\hat{\omega}_c)}, \hat{\omega}_c \leq\mid \hat{\omega} \mid\leq -\hat{\omega}_d \\ 0, \hat{\omega} \in [0, \hat{\omega}_a] \cup (\hat{\omega}_d, 2\pi) \end{array} \right\} \tag{6.2}$$

where,

$$\hat{\omega}_a = \frac{(1-\beta)\pi + \epsilon}{s_3}; \quad \hat{\omega}_b = \frac{s_1 \pi}{s_2 s_3}$$
$$\hat{\omega}_c = \frac{\pi - \epsilon}{s_3}; \quad \hat{\omega}_d = \frac{\pi + \epsilon}{s_3}$$

Where the LP channel's up-sampling and down-sampling factors are represented by s_1 and s_2, respectively. The HP channel's up-sampling and down-sampling factors are represented by s_3 and s_4, respectively. The characteristics of the FAWT wavelet transform are determined by tuning parameters s_1, s_2, s_3, s_4, and β. The parameter s_1 in FAWT determines the scaling factor. The FAWT's ability to extract features from the signal at different scales is affected by changes in s_1, which change the wavelet's size. The parameter s_2 determines the wavelet function's shift. It explains the wavelet's location along the signal axis. A varying s_2 can be used to determine the transient phenomena within the signal. The parameter s_3 determines the wavelet function's complex plane rotation angle. The parameter s_4 determines the wavelet function's orientation and skewness. The parameter β determines the smoothness or regularity of the wavelet function. The parameter β determines sudden or slow changes in the signal by balancing the tradeoff between temporal and frequency localization. The optimum values of tuning parameters s_1, s_2, s_3, s_4, and β are required for the efficient analysis of EEG signals. Standard FAWT uses empirical parameter selection, which may result in increased decomposition error.

6.2.1.2 Proposed Adaptive FAWT Algorithm

The wavelet transform is a mathematical tool employed to analyze a non-stationary signal using wavelet frames (WFs). WFs represent a set of fundamental functions that cover many aspects of the time-frequency domain, including analytical complexity, adaptive window, and improved resolution. FAWT differs from conventional wavelet transforms with its improved flexible frequency focus and complex coefficients [143]. Overall, FAWT's adaptive characteristics distinguish it as a valuable tool for oscillatory signal analysis. However, FAWT dyadic wavelet transforms (dual-tree wavelet transforms and discrete wavelet transforms) have some significant limitations, including rigid time frequency response, limited frequency resolution, less adaptability, and inefficiency for event-related analysis, which has the limitation of requiring a precise selection of tuning parameters. The proposed AFAWT overcomes this limitation. Figure 6.2 shows the general approach to emotion classification using AFAWT. The suggested AFAWT is a combination of FAWT and PSO. An EEG signal $s[n]$ can be decomposed by FAWT into a J number of SBs as follows:

$$\hat{s}[n] = \sum_{k=1}^{J} d_k[n] \tag{6.3}$$

The number of decompositions depends on the sampling rate and the number of samples per epoch while increasing FAWT decomposition levels also leads to increase in classification accuracy [53]. This work tested several AFAWT decomposition levels, and the best accuracy is obtained at 16 levels. Where $d_k[n]$ stands for the k_{th} FAWT sub-band, and $\hat{s}[n]$ is the decomposed version of the EEG signal. The decomposition error $D_e[n]$ can be defined as:

$$D_e[n] = s[n] - \hat{s}[n] \tag{6.4}$$

The decomposition error has been minimized using constraint optimization. Constraint-based optimization considers specific parameter values that follow predefined restrictions. These restrictions limit parameter values, either through binding restrictions or as linear or nonlinear inequalities. The constraints define the set of viable solutions and direct the selection of the best parameter values. The inequality for the constrained optimization is defined by the following equations:

$$1 - \frac{s_1}{s_2} \leq \beta \leq \frac{s_3}{s_4} \tag{6.5}$$

$$1 \leq \frac{s_4}{s_3} \leq 3 \tag{6.6}$$

$$\frac{1}{2} < \frac{s_1}{s_2} < 1 \tag{6.7}$$

The first constraint is defining the perfect reconstruction condition for FAWT filter banks.

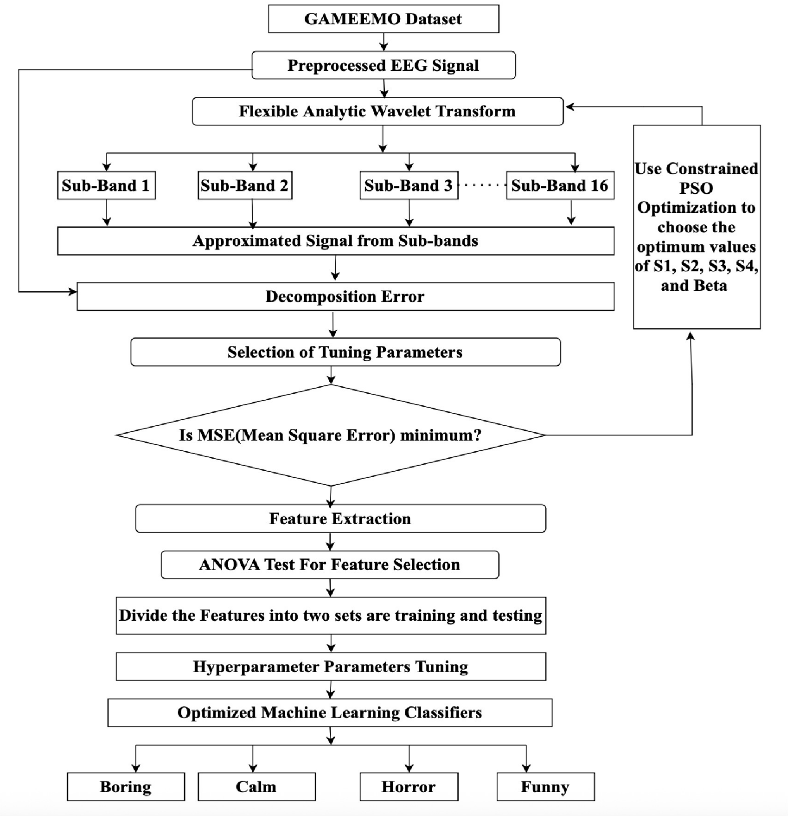


Figure 6.2: Flowchart for developing the AFAWT algorithm.

The second constraint applies to avoid ringing effects and choose the proper adjustable basis functions for the decomposition. The third constraint applies to get a finer frequency resolution than the dyadic wavelet transforms. To improve compatibility between $\hat{s}[n]$ and s[n], an optimized version of FAWT called adaptive-FAWT has been used. In AFAWT, the best tuning parameters are selected by making the MSE of the decomposition error as small as possible while adhering to specific constraints. The parameter search space (Ψ) for the adaptive flexible analytic wavelet transform is defined for particle swarm optimization.

$$\Psi = \{min(\varnothing_e) \mid s_1, s_2, s_3, s_4 \text{ and } \beta\} \tag{6.8}$$

The objective function of PSO is to minimize MSE, allowing the adaptive selection of the best values of s_1, s_2, s_3, s_4 and β an equation designed specifically for this purpose as follows:

$$minMSE = \frac{1}{N}\sum_{n=1}^{N} D^2{}_e[n] \tag{6.9}$$

The mean squared error between the original and reconstructed signals has been minimized to achieve the best values for the tuning parameters. Algorithm 6.1 describes the step-by-step approach of the proposed AFAWT algorithm. The AFAWT algorithm increases flexibility in the selection of the Q-factor. Flexibility in changing the Q factor more effectively can result in enhanced performance and superior decomposition of EEG signals.

Algorithm 6.1 : Adaptive Flexible Analytic Wavelet Transform

```
1: Initialize the tuning parameters $s_1$, $s_2$, $s_3$, $s_4$, and $\beta$.
2: Evaluate the decomposition levels $j$.
3: Decompose the EEG signal.
4: Reconstruct the signal using inverse FAWT.
5: Calculate decomposition error $D_e$ from steps 3 and 4.
6: Define the objective function to minimize MSE of $D_e$ using equation (6.9) with constraints.
7: while the stop criteria are not satisfied do
8:     for each particle do
9:         Calculate the objective function in equation (6.9).
10:        Update $P_{\text{best}}$ if required.
11:        Update $G_{\text{best}}$ if required.
12:    end for
13:    Update inertia.
14:    for each particle do
15:        Update position $x$ according to equation (4.10).
16:        Update velocity $v$ according to equation (4.11).
17:    end for
18: end while
19: return $G_{\text{best}}$: the best estimation of global optimum $s_{1\text{opt}}$, $s_{2\text{opt}}$, $s_{3\text{opt}}$, $s_{4\text{opt}}$, and $\beta_{\text{opt}}$.
```

6.2.1.3 ANOVA Test

Analysis of variance, or ANOVA, is a statistical method for comparing the means of two or more groups. Its main objective is to determine whether these various groups' means differ noticeably. In an ANOVA, the p-value determines whether the differences observed in the groups are simply due to random variation or statistically significant. Real differences are suggested by a low p-value (< 0.05), which leads to the rejection of the null hypothesis. A high p-value (> 0.05) means that there is not enough data to conclude that there is a significant group difference.

6.2.1.4 Feature Extraction and Selection

AFAWT-based sub-bands are used to evaluate various features. A one-way ANOVA test is run to select the most important features. The features with p-values< 0.05 are selected from a large number of features. The most relevant features after the ANOVA test. Where N represents the EEG signal's length and $y(t)_i$ represents the SBs of the EEG signal in segment i. The constant term c denotes the time series' average value, and the time series' lag values are represented by $y(t–i)$.

(a) Hjorth Mobility (HM):

HM [72] is the square root of the variance of the signal's first derivative divided by the signal's variance. HM is defined as

$$HM = \sqrt{\frac{\operatorname{var}\left(\frac{dy(t)}{dt}\right)}{\operatorname{var}(y(t))}} \tag{6.10}$$

(b) Hjorth Complexity (HC):

HC [72] compares the signal's complexity to a sine wave; value closer to 1 means more similarity. HC is mathematically defined as

$$HC = \frac{\operatorname{Mob}\left(\frac{dy(t)}{dt}\right)}{\operatorname{Mob}(y(t))} \tag{6.11}$$

(c) Shannon Entropy (SE):

SE [73] measures average information or uncertainty in a signal. SE is mathematically defined as

$$H(x) = -\sum_{i=1}^{N} p(y_i) \log_2 p(y_i) \tag{6.12}$$

(d) Renyi Entropy (RE):

RE [144] is a measurement of the variability or uncertainty in the probability distribution. RE is mathematically defined ad

$$H_\alpha(x) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{N} p(y(t)_i)^2 \right) \tag{6.13}$$

(e) Auto-Regressive Model (AM) :

AM [145] predicts future values based on the linear relationship with previous values and has been used in time series analysis and prediction. AM is mathematically defined as

$$y(t) = c + \sum a(i) \times y(t-i) + \epsilon(t)$$

(f) Wenergy (WE):

The function wenergy [146] computes and returns the energy percentages related to the wavelet packet tree's terminal nodes. These percentages of energy show the proportionate contributions of each terminal node to the overall signal energy.

6.2.2 Results and Discussion

This section discusses the effectiveness of the suggested AFAWT algorithm for categorizing emotions. The AFAWT algorithm is a hybrid of the FAWT and PSO algorithms. The best values of FAWT tuning parameters are found by minimizing the objective function in PSO. The PSO algorithm parameter values are defined in Table 6.1. In PSO, the number of variables represents the number of parameters to optimize. The maximum iteration represents stopping criteria, and population size represents the number of particles in the swarm because each particle represents a potential solution in search space. The inertia weights w_{Max} and w_{Min} control the tradeoff between exploration and exploitation during optimization.

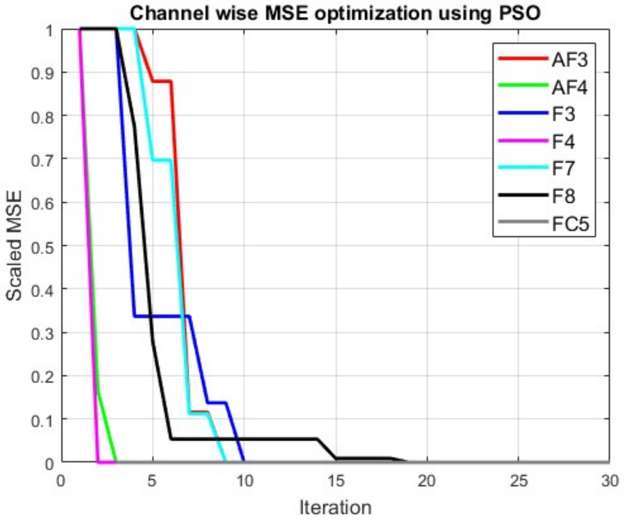


Figure 6.3: MSE-based optimization of EEG signals from the AF3, AF4, F3, F4, F7, F8, and FC5 channels.

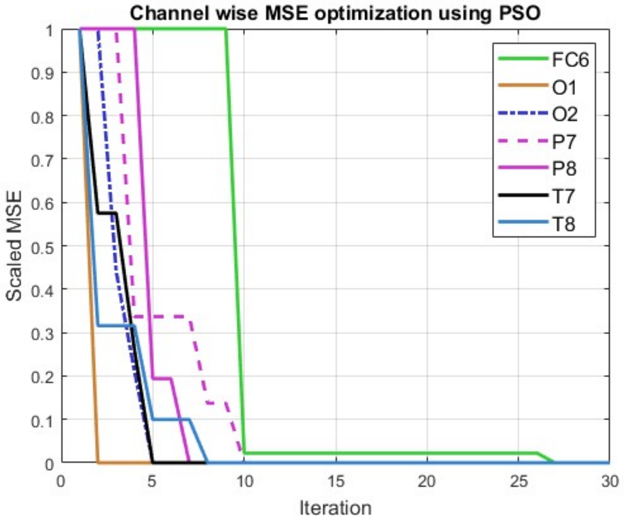


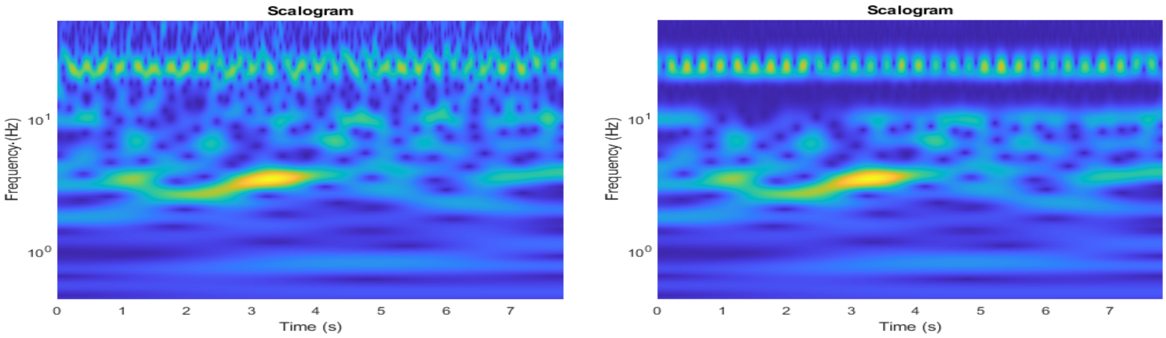
Figure 6.4: MSE-based optimization of EEG signals from the FC6, O1, O2, P7, P8, T7, and T8 channels.

The acceleration coefficients c_1 and c_2 determine the impact of personnel and global best on particle movement. The upper (lb) and lower (lb) bounds show the allowable range within which parameters can change during optimization. Maximum v_{Max} and minimum v_{Min} velocities are constraints that limit the range of particle velocities. The mean square error (MSE) measures the difference between the original and reconstructed signals. A lower MSE indicates that the decomposed signal resembles the original signal. Figures 6.3 and 6.4 illustrate the minimization of the MSE convergence curve for all 14 channels with a maximum of 30 iterations. It demonstrates that the MSE of channels for FC6, P7, F3, and F8 saturates after 10-15 iterations. The channels F4, AF4, FC5, T7, O1, and O2 MSE saturate after 3-5 iterations. The channels F7, F8, AF3, P8, T8, and F7 MSE saturate after 5–10 iterations. Table 6.2 displays the average MSE before and after optimization. The average MSE for all fourteen channels has significantly decreased after optimization, as seen in Table 6.2. This means that most of the information from the original signal is still present in the reconstruction signal. Therefore, a low MSE indicates there is minimal information loss because the decomposition process retained the key elements of the original signal [147]. The optimal values for FAWT tuning parameters for all channels are also displayed in Table 6.2. MSE derived from commonly used signal decomposition algorithms has been evaluated.

Table 6.1: Parameter configuration for the PSO algorithm.

Parameter	**Value/Description**
Number of variables	5
Maximum iterations	30
Population size	50
Inertia weight, W	Linearly decreasing from $W_{\max} = 0.8$ to $W_{\min} = 0.6$
Acceleration coefficients	$c_1 = 2, c_2 = 1$
Lower bound, **lb**	$[1,\ 1,\ 1,\ 1,\ 0.1]$
Upper bound, **ub**	$[30,\ 30,\ 30,\ 30,\ 0.99]$
Velocity bounds	$v_{\max} = 0.2 \times (\mathbf{ub} - \mathbf{lb})$, $v_{\min} = -v_{\max}$

Figure 6.5 (a) displays the time-frequency representation (TFR), scalogram, of the reconstructed EEG signal after applying FAWT. Figure 6.5 (b) displays the TFR of EEG signals after applying AFAWT. This shows that AFAWT improves time-frequency localization by eliminating irrelevant frequency components. The TFR representation is the absolute value of the continuous wavelet transform coefficient of the EEG signal. It has been observed that the average MSE provided by the proposed algorithm is minimal. A lower MSE in the context of AFAWT shows that the decomposition effectively captures the underlying structures and properties of the non-stationary signal. It is an indication of how well the decomposition is done. The decomposed SB signals are further analyzed for feature extraction.



(a) FAWT

(b) AFAWT

Figure 6.5: Time–frequency representation of EEG signals using (a) FAWT and (b) AFAWT.

Table 6.2: MSE of all EEG channels before and after optimization with optimum values of tuning parameters.

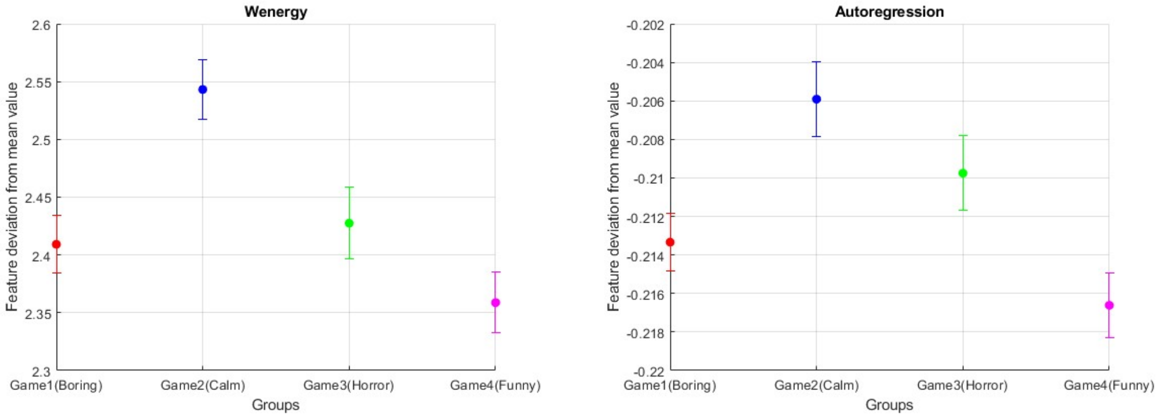
Channel	Avg. MSE Before FAWT Optimization	Avg. MSE After FAWT Optimization	Optimum Values of Tuning Parameters				
			s_1	s_2	s_3	s_4	β
AF3	8.42×10^{-20}	1.16×10^{-28}	1.00	1.38	6.75	8.18	0.31
AF4	0.13937	2.60×10^{-29}	9.26	13.39	5.90	6.99	0.35
F3	1.27×10^{-19}	1.10×10^{-29}	1.00	1.60	5.95	11.67	0.35
F4	1.40×10^{-19}	1.57×10^{-29}	4.68	6.80	1.00	1.00	0.27
F7	2.85×10^{-19}	7.58×10^{-28}	1.00	1.51	9.10	8.91	0.35
F8	2.68×10^{-28}	1.11×10^{-28}	1.74	2.39	3.79	5.69	0.40
O1	1.10×10^{-20}	1.20×10^{-29}	13.80	14.72	13.84	16.44	0.31
O2	1.16×10^{-19}	1.70×10^{-29}	15.41	24.81	8.52	17.68	0.34
P7	2.63×10^{-5}	2.54×10^{-29}	9.79	15.97	6.95	8.54	0.37
P8	0.02477	3.94×10^{-29}	16.28	24.87	7.81	13.45	0.43
T7	6.54×10^{-20}	1.11×10^{-29}	9.63	13.91	10.40	16.74	0.19
T8	7.28×10^{-19}	2.38×10^{-29}	14.32	12.57	3.16	9.20	0.25
FC5	5.81×10^{-29}	4.28×10^{-29}	1.00	1.18	5.23	3.31	0.32
FC6	1.14×10^{-29}	1.62×10^{-29}	26.11	26.35	7.22	22.11	0.11

6.2.2.1 Feature Statistical Test Analysis

To determine the feature set, an ANOVA and multi-comparison test are performed. The ANOVA compares the means of three or more groups and determines significant differences among them.

Table 6.3: Probabilistic value of selected features.

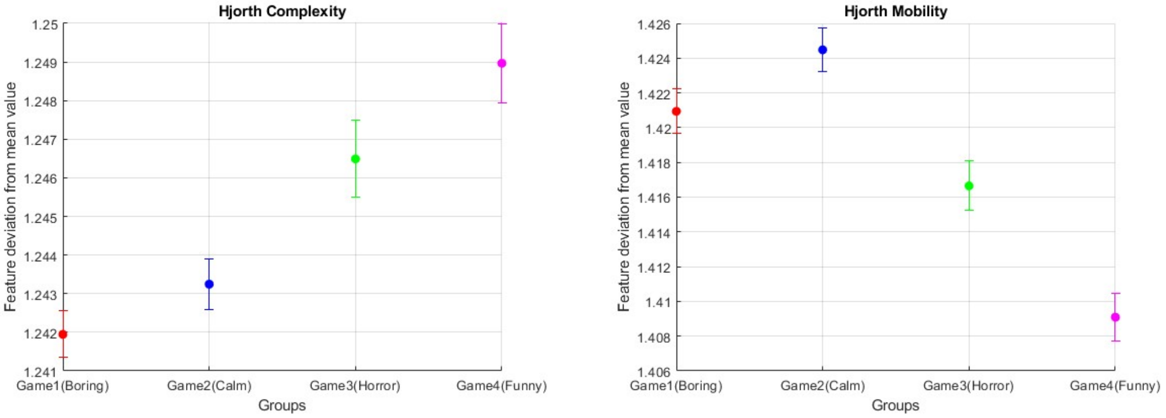
Statistical Feature	p-Value
Wenergy (WE)	1.485921×10^{-5}
Renyi Entropy (RE)	2.0×10^{-4}
Auto Regressive Model (AM)	3.14424×10^{-7}
Hjorth Mobility (HM)	1.30295×10^{-15}
Hjorth Complexity (HC)	3.69213×10^{-10}
Shannon Entropy (SE)	3.50949×10^{-8}



(a) Posthoc test of Wenergy

(b) Posthoc test of Autoregression

Figure 6.6: Posthoc test results of Wenergy and Autoregression features.

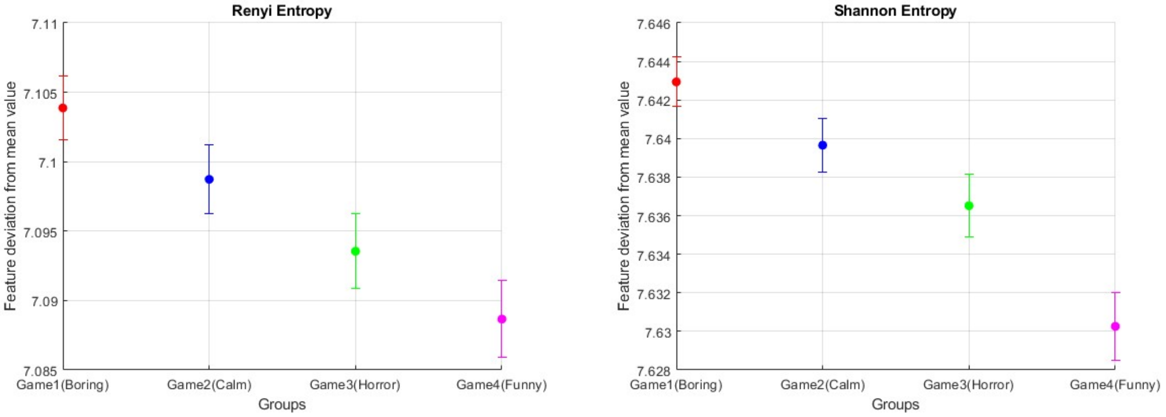


(a) Posthoc test of Hjorth Complexity

(b) Posthoc test of Hjorth Mobility

Figure 6.7: Posthoc test results of Hjorth features.

A multi-comparison test, also called a posthoc test, is performed after an ANOVA. The ANOVA identifies prominent features that have a strong relationship to the multiclass and, based on relative *p* values, ranks them. ANOVA is performed after all of the subsets and features are merged into one set. All samples of the EEG signals are used to extract twenty different features. ANOVA analysis shows that six exhibit a significant correlation for the four emotions under investigation, as shown in Table 6.3, After conducting the ANOVA test, a post hoc test is performed on the extracted features. Figures 6.6 to 6.8 show that all selected features demonstrate better discrimination. The Hjorth mobility in Figure 6.7 (a), Hjorth complexity in Figure 6.7(b), and Shannon entropy in Figure 6.8 (b) demonstrate zero overlapping, which highlights their better discrimination potential. Similarly, the auto-regressive model in Figure 6.6(b), Wenergy in Figure 6.6(a), and Renyi entropy in Figure 6.8 demonstrate minimal overlap. The post-hoc test results for all features demonstrate that they clearly distinguish each of the four emotions. The relevance of the selected features is confirmed by the probabilistic values in Table 6.3, which are relatively smaller than the standard threshold (< 0.05), serving as motivation to use this feature set. This feature set is used to form a feature matrix, which is used as input to various machine learning classifiers.



(a) Posthoc test of Renyi Entropy

(b) Posthoc test of Shannon Entropy

Figure 6.8: Posthoc test results of Renyi and Shannon energy features.

6.2.2.2 Effect of Hyperparameter Tuning on Emotion Classification Performance

In this work, different classification models are explored, including SVM, k-NN, Neural Network (NN), and Ensemble (EN) classifiers, which are used to assess the performance of the retrieved features. Subject-independent and cross-validation methods have been used to measure the performance classification models. Subject independence lowers the possibility of overfitting by ensuring that the model generalizes well to unobserved subjects or data samples [148]. Cross-validation provides a reliable estimate of performance on unknown data by evaluating the model's capacity for generalization across several validation sets, which helps avoid overfitting.

The final set of features is divided into training and testing sets (subject-independent). Ninety percent of the total features are used in the training process. Randomly sequenced training set validation performance is analyzed using 10-fold cross-validation. The 10-fold cross-validation method is used to assess each classification model. A ten percent randomly sequenced sample that was not used during training is used to test the trained model, which is then evaluated. Table 6.4 shows the results of the evaluated model for different channels, in which channels F4, P7, and T8 give more promising results than the other channels. Among all tested classification models, SVM, Ensemble, Neural Network, and k-NN classification models show promising results. The optimized version of the classifier is showing better accuracy than other versions. The k-NN classifier, along with its distinct version, exhibits superior accuracy compared to other classifiers. The optimized version results from finding the best combination of hyperparameters. This finding can also be explored in other datasets. In the present study, hyperparameters are optimized using a Bayesian optimizer and 10-fold cross-validation. The Bayesian optimizer uses a probabilistic approximate model to estimate the objective function (minimum classification error). It chooses which hyperparameters to test based on the acquisition function. By balancing exploration (trying new hyperparameters) and exploitation (selecting hyperparameters that are likely to improve performance). The 10-fold cross-validation provides a reliable estimate of the classifier's performance, which helps in avoiding overfitting to a particular group of data. Combining a Bayesian optimizer with 10-fold cross-validation simplifies the process of determining the best classifier hyperparameters. Details of optimized hyperparameters are displayed in Table 6.5.

Table 6.4: Classification accuracy (%) across EEG channels for different classifiers.

Classifier	Channel-wise Accuracy (%)													
	AF3	AF4	F3	F4	F7	F8	O1	O2	P7	P8	T7	T8	FC5	FC6
Medium NN	48	53	65	67	58	52	62	64	71	58	59	62	59	60
Wide NN	59	73	79	75	56	55	74	73	82	73	69	78	70	69
Optimizable NN	69	73	75	88	61	72	85	75	87	78	72	85	79	77
Bagged Trees EN	54	55	61	76	61	64	75	62	85	66	62	71	62	64
Subspace KNN EN	63	65	93	89	61	70	82	77	84	81	62	94	71	71
Optimizable EN	67	78	77	91	73	67	82	78	94	81	74	86	85	73
Quadratic SVM	57	66	72	70	57	57	71	62	69	69	57	79	61	63
Cubic SVM	67	76	74	84	66	72	81	77	91	80	71	84	78	79
Optimizable SVM	74	81	73	88	70	73	82	79	97	91	80	87	82	86
Fine k-NN	**86**	**89**	**92**	**97**	**80**	**87**	**89**	**88**	**97**	**93**	**86**	**95**	**89**	**89**
Weighted k-NN	64	76	73	86	64	65	80	76	87	84	68	80	76	74
Optimizable k-NN	84	**93**	**93**	**97**	83	**91**	90	90	**98**	**93**	**86**	**97**	89	**91**

Table 6.5: Classifier hyperparameters: Standard configurations vs. Bayesian optimization results.

Parameter	**Standard 1**	**Standard 2**	**Optimized (Bayesian)**
Support Vector Machine			
Kernel Function	Cubic	Quadratic	Gaussian
Kernel Scale	Auto	Auto	Auto
Box Constraint	1	1	**0.0012**
Multi-class Method	OvO	OvO	OvA
Standardize Data	Yes	Yes	Yes
k-Nearest Neighbors			
Number of Neighbors	10	10	**1**
Distance Metric	Euclidean	Euclidean	**Correlation**
Distance Weight	Equal	Sq. Inverse	**Equal**
Standardize Data	Yes	Yes	Yes
Ensemble Classifier			
Ensemble Method	Subspace	Bagged	**AdaBoost**
Learner Type	kNN	Decision Tree	Decision Tree
Number of Learners	30	907	**2206**
Subspace Dimension	72	30	**238**
Learning Rate	–	–	**0.858**
Neural Network			
Number of Layers	1	1	**2**
First Layer Size	25	100	**3.82e-4**
Second Layer Size	–	–	**300**
Third Layer Size	–	–	**23**
Activation Function	ReLU		
Iteration Limit	1000		
Regularization	None	None	**Yes**
Standardize Data	Yes		

6.2.2.3 Performance Metric Analysis

Performance metrics including accuracy (ACC), sensitivity (SEN), F1-score, specificity (SPE), Matthew's correlation coefficient (MCC), Cohen's Kappa coefficient (k), and area under the curve (AUC) are used to identify the robustness of the model. These metrics help in assessing the model's efficiency and dependability in correctly categorizing the data. Table 6.6 shows that the classifier's kappa values, O-SVM (optimized support vector machine) and O-Ensemble (optimized Ensemble), lie in the range 0.61 to 0.81, which means significant agreement, whereas O-kNN (optimized k Nearest Neighbors) kappa is 0.878, showing perfect agreement. The ACC, SEN, SPE, Precision, F1 score, MCC, Kappa, and AUC ranges are 76.85-91.07, 76.084-90.84, 76.02-91.94, 76.58-91.19, 75.99-90.74, 68-87.9, 0.6818-0.878, and 0.907-0.938 shown in Table 6.6 . It displays that the suggested method provides a dependable and consistent solution for the classification of four EEG-based emotions. O-kNN produces the least classification error compared to the other three classifiers. Table 6.6 compares the performance metrics of each of the four classifiers. It is easy to see that the O-NN, and O-Ensemble accuracy, precision sensitivity, and F1-score parameters are roughly the same. Compared to the other three classifiers, O-kNN performs better.

Table 6.6: Classifier performance ranking with O-kNN as top performer.

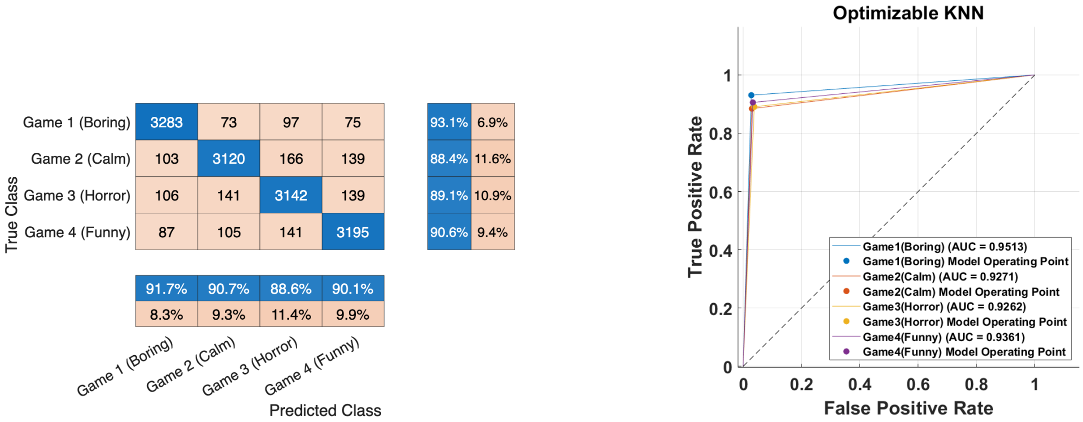
Model	**Acc**	**Sen**	**Spe**	**Pre**	**F1**	**MCC**	**Kap**	**AUC**
O-Neural Network	76.85	76.10	76.00	76.50	75.90	68.30	0.68	0.907
O-Ensemble	79.00	78.00	79.60	78.90	78.00	71.10	0.70	0.926
O-SVM	82.64	82.20	81.00	82.70	82.20	76.50	0.76	0.936
O-kNN	**91.07**	**90.80**	**91.90**	**91.10**	**90.70**	**87.90**	**0.87**	**0.938**

Acc = Accuracy, Sen = Sensitivity, Spe = Specificity, Pre = Precision,
MCC = Matthews Correlation Coefficient, Kap = Kappa, AUC = Area Under Curve

Table 6.6 compares the MCC, k, and AUC for each of the five classifiers. All classifiers have an AUC value of more than 0.9, MCC and kappa have almost equal values. O-Ensemble and O-Neural Network are approximately 0.7, O-SVM is approximately 0.8, and O-kNN is approximately 0.9, showing that O-kNN is a superior classifier. The AUC value in all classifiers is greater than 0.9, displaying a clear distinction in four classes.

6.2.2.4 Confusion Matrix

In multi-class classification, the confusion matrix is a popular performance matrix for assessing machine learning algorithms. Figure 6.9(a) shows that the true positive (TP) for the emotions of boredom in cell (1, 1), calmness in cell (2, 2), horror in cell (3, 3), and funny in cell (4, 4) are equal to 3283, 3120, 3142, and 3195, respectively. This means that the majority of the predicted and actual values are identical. False positives and false negatives show the incorrectly predicted values. The confusion matrix in Figure 6.9(a) shows that these values are relatively low. Precision and recall can be calculated using the above confusion matrix. Precision, calculated to be 0.90238, indicates how accurately positive predictions are made. On the other hand recall measures the proportion of correctly predicted positive cases and is calculated at 0.90235.



(a) Confusion Matrix of O-kNN (b) ROC Curve of O-kNN

Figure 6.9: Performance analysis of the O-kNN classifier using the confusion matrix and ROC curve.

6.2.2.5 Receiver Operating Characteristic

The performance of the four-class classification problem is widely assessed using the receiver operating characteristic (ROC) curve. The Boring, Calm, Horror, and Funny classes can be more clearly distinguished when the area under the curve (AUC) is greater. The model's effectiveness is shown by the ROC curve in Figure 6.9(b), which has a high AUC of 0.9513, 0.9271, 0.9262, and 0.9361, demonstrating the precise distinction between the four classes.

6.2.3 Comparative Analysis

Table 6.7 summarizes the findings of other relevant studies with similar statements and the same datasets. Alakus et al.[35] used a multi-layer perceptron neural network (MLPNN) for four class classifications and reported a maximum accuracy of 73.21%. The accuracy of the kNN and SVM classifiers was lower at 42.85% and 55%, respectively. Alakus et al. [35] reported an accuracy of 81.3% using MLPNN, 72.5% using SVM, and 66% using kNN for 2-class classifications. Alakus and Turkoglu et al. [125] suggested calculating the spectral entropy values of EEG signals across all channels.

A bidirectional long-short-term memory (LSTM) architecture was then used to classify these values into two classes, achieving an accuracy of 76.91%. Aslam et al.[126] suggest a deep learning technique based on GoogleNet for automatically identifying human emotions from EEG signals. Continuous Wavelet Transform is used to convert EEG signals into scalogram pictures, and pre-trained GoogLeNet is used to extract features. The k-NN, SVM, and ELM classifiers are then utilized to classify emotions using the acquired deep features. It achieved an accuracy of 98.78%, 98.53%, and 98.41% for 2-class using SVM, kNN, and ELM classifiers. Tuncer et al.[127] developed an emotion identification model using LEDPatNet19, which combines statistical feature extraction, nonlinear textural feature synthesis using an LED block cipher's S-box, and TQWT. It uses iterative Chi2 and ReliefF for feature selection and operates on the original EEG signal as well as the 18 TQWT sub-band. It reported an accuracy of 99.29% for channel-wise 2-class classification using a cubic SVM classifier. Su et al. [128] proposed projection dictionary pair learning (PDPL), which uses synthesis and analysis dictionaries together to improve feature representation. The genetic algorithm is used to optimize the PDPL parameters based on experience and produce the best possible model solution. It observed an accuracy of 64.34% for the two classes and 49.01% for the four classes using the PDPL. The proposed work suggests a robust signal decomposition algorithm that provides perfect multi-resolution for EEG signals. The proposed

Table 6.7: Performance comparison with prior EEG-based emotion recognition studies using the same dataset.

Study	**Feature Extraction Method**	**Classes**	**Classifier**	**Accuracy (%)**
Alakus et al. [35]	Hjorth, Entropy, Std. Dev., Variance, Zero Crossing	4	k-NN	42.85
			SVM	55.00
			MLPNN	73.21
		2	k-NN	66.00
			SVM	72.50
			MLPNN	81.35
Alakus et al. [125]	Spectral Entropy	2	Bi-LSTM	76.93
Aslam et al. [126]	EEG Scalogram (GoogleNet Features)	2	k-NN	98.53
			SVM	98.73
			ELM	98.41
Turners et al. [127]	Fused Features from LEDPatNet19	2	Cubic SVM	99.29
Su et al. [128]	GAPDPL (Genetic Algorithm Projection Dictionary Pair Learning)	4	GAPDPL	49.01
		2		64.34
Proposed Method	Hjorth Mobility, Hjorth Complexity, Shannon Entropy, Wenergy, AR Features	4	Optimized k-NN	**90.3**

work employs a wide range of features and uses statistical tests to select a specific set of features. Feature selection reduces dimensionality and captures important patterns. The quality of features has a direct impact on the model's performance. Classification models and their hyperparameters determine the model's ability to learn from the features. Hyperparameter tuning in the work makes adjustments to model complexity helps in capturing complex patterns. The Optimized models and selected features offer the classification accuracy of 90.3%.

6.3 Optimal Selection of TQWT Parameters using the Crayfish Optimization Algorithm

6.3.1 Methodology

Figure 6.10 shows the block diagram of the proposed methodology. This paper uses the publicly available datasets SEED IV [31] and DEAP[29].

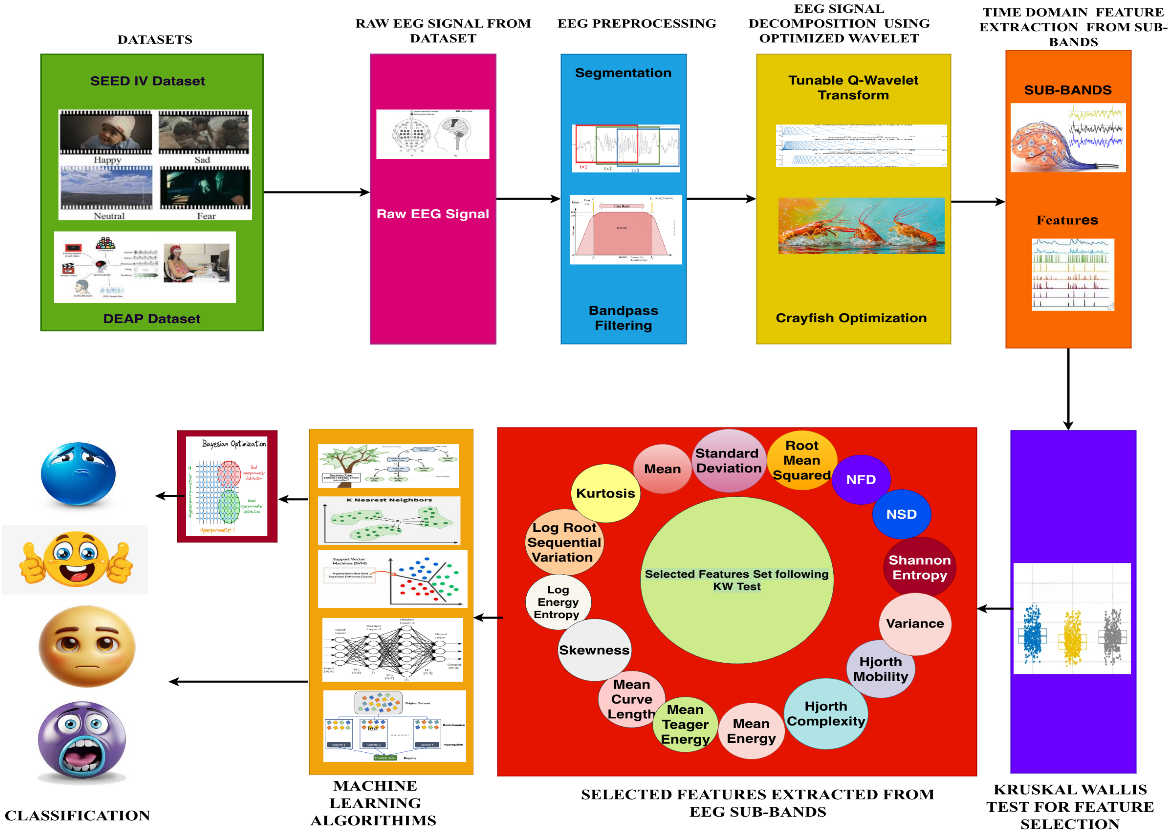


Figure 6.10: Block diagram of the proposed methodology for optimal TQWT parameter selection using the Crayfish Optimization Algorithm.

6.3.1.1 Tunable Q Wavelet Transform (TQWT)

The tunable-Q wavelet transform [133] decomposes a signal into sub-bands with a variable quality factor (Q), unlike traditional wavelet transforms with a fixed Q. It produces one low-pass sub-band (LPS) and multiple high-pass sub-bands (HPS), with the number determined by the decomposition levels (J). The algorithm is defined by three key parameters: oversampling rate (r), quality factor (Q), and decomposition levels (J), enabling flexible adaptation to different signal types.

$$\hat{M}_0^J(\hat{\omega}) = \begin{cases} \prod_{n=0}^{J-1} \hat{M}_0\left(\frac{\hat{\omega}}{\alpha^n}\right), & \text{if } |\hat{\omega}| \le \alpha^J \pi, \\ 0, & \text{if } \alpha^J \pi < |\hat{\omega}| \le \pi \end{cases} \tag{6.14}$$

$$\begin{gathered} \hat{M}_1^J(\hat{\omega}) = \hat{M}_1\left(\frac{\hat{\omega}}{\alpha^{J-1}}\right) \prod_{n=0}^{J-2} \hat{M}_0\left(\frac{\hat{\omega}}{\alpha^n}\right), \\ \text{if } (1-\beta)\alpha^{J-1}\pi \le |\hat{\omega}| \le \alpha^{J-1}\pi \\ 0, \ \text{for other } \hat{\omega} \in [-\pi, \pi] \end{gathered} \tag{6.15}$$

$$\beta = \frac{2}{Q+1} \qquad \alpha = 1 - \frac{\beta}{r} \tag{14}$$

In TQWT, the quality factor (Q) and redundancy (r) strongly affect wavelet performance. Optimal values are determined using the Crayfish Optimization Algorithm (COA) by minimizing the mean square error between original and reconstructed signals. High Q suits oscillatory signals, while low Q suits transient signals; redundancy controls decomposition resolution and scale overlap, with higher values enhancing reconstruction and noise resistance but increasing computation.

Algorithm 6.2 : COA for Optimizing Quality Factor (Q) and Redundancy (r) in TQWT

```
Input: Number of agents $N$, maximum iterations $T$, lower bounds $lb$, upper bounds $ub$, dimension $dim = 2$ (Q and r), objective function $f_{obj}$.
Initialize population $X$ with $N$ agents, each having random values for Q and r within $[lb, ub]$.
Evaluate the fitness of each agent using $f_{obj}(X(i, :))$ given in Eq.(15).
Set the initial best position and best fitness.
for each iteration $t = 1$ to $T$ do
    Calculate coefficient $C = 2 - (t/T)$ for updating positions.
    Set environmental factor $temp = rand \times 15 + 20$.
    Compute average position $xf = (best\_position + global\_position)/2$.
    for each agent $i$ do
        if $temp > 30$ then ▷ (Summer Resort Stage) Update position based on Eq. 4.5 or Eq. 4.6:
            if $rand < 0.5$ then
                $Xnew(i, :) = X(i, :) + C \times rand(1, dim) \times (xf - X(i, :))$.
            else
                $Xnew(i, j) = X(i, j) - X(z, j) + xf(j)$.
            end if
        else ▷ Foraging Stage Update position based on Eq. 4.3 or Eq. 4.4:
            Compute $P = 3 \times rand \times fitness_f(i)/f_{obj}(Xfood)$.
            if $P > 2$ then
                Update $Xfood = \exp(-1/P) \times Xfood$.
            else
                $Xnew(i, :) = (X(i, :) - Xfood) \times p\_obj(temp) + p\_obj(temp) \times rand(1, dim) \times X(i, :)$.
            end if
        end if
    end for
    Apply boundary conditions: Ensure $Q$ and $r$ are within $[lb, ub]$.
    for each agent $i$ do
        Evaluate updated positions and update global best position if necessary.
    end for
    Store the best solution and its fitness for convergence analysis.
end for
Output: The best values for Q and r, and minimized MSE.
```

6.3.1.2 Optimization of TQWT

The tunable Q wavelet transform decomposes and reconstructs EEG signals by adjusting the quality factor (Q) and redundancy (r), where manual selection often leads to information loss and higher mean square error (MSE). TQWT [133] is particularly effective for oscillatory biomedical signals such as EEG, EMG, and ECG, decomposing them into low-pass and high-pass regions across J stages. The frequency responses of the low-pass sub-band $\mathbf{M}_0^J(\hat{\omega})$ and high-pass sub-bands $\mathbf{M}_1^J(\hat{\omega})$ are given in (6.14) and (6.15), with each successive stage halving the input signal's frequency range (0.5–60 Hz).

Table 6.8: List of features used for classification with their mathematical expressions and definitions.

Feature	Mathematical Expression	Definition
Mean (μ)[2]	$\mu = \frac{1}{N}\sum_{i=1}^{N} \lvert s_i \rvert$	Arithmetic mean of absolute signal values.
Standard Deviation (σ_S)[2]	$\sigma_S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - \mu)^2}$	Spread of signal values around the mean.
Root Mean Square (RMS)[2]	$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} s_i^2}$	Represents signal magnitude using quadratic mean.
Normalized First Difference (NFD)[2]	$\delta_F = \frac{1}{N-1}\sum \lvert s_{i+1} - s_i \rvert,\ \text{NFD} = \frac{\delta_F}{\sigma_S}$	First-order signal difference normalized by standard deviation.
Normalized Second Difference (NSD)[2]	$\delta_S = \frac{1}{N-1}\sum \lvert s_{i+2} - s_i \rvert,\ \text{NSD} = \frac{\delta_S}{\sigma_S}$	Second-order signal difference normalized by standard deviation.
Shannon Entropy	$S_{\text{en}} = -\sum p(s_i)\log_2(p(s_i))$	Quantifies uncertainty or randomness in signal.
Variance [2]	$\text{Var}(X) = \mathbb{E}[(s-\mu)^2]$	Statistical dispersion equal to the square of the standard deviation.
Hjorth Mobility[72]	$\text{HM} = \sqrt{\frac{\text{Var}\left(\frac{ds}{dt}\right)}{\text{Var}(s)}}$	Indicates signal's mean frequency.
Hjorth Complexity [72]	$\text{HC} = \sqrt{\frac{\text{Mob}\left(\frac{ds}{dt}\right)}{\text{Mob}(s)}}$	Represents variation in frequency content.
Hjorth Activity[72]	$\text{HA} = \frac{1}{N}\sum(s_i - \mu)^2$	Indicates the power of the signal.
Mean Energy[149]	$\text{ME} = \frac{1}{N}\sum s_i^2$	Represents average energy of the signal.
Teager Energy (MTE)[149]	$\text{MTE} = \log\left(\frac{1}{N}\sum(s_{i-1}^2 - s_i s_{i-2})\right)$	Captures energy variations within short intervals.
Mean Curve Length (MCL)[149]	$\text{MCL} = \frac{1}{N-1}\sum \lvert s_{i+1} - s_i \rvert$	Measures signal complexity over time.
LSSV[150]	$\text{LRSSV} = \log\sqrt{\sum(s_i - s_{i-1})^2}$	Log root sum of squared signal variations.
Log Energy Entropy[151]	$H_{\text{logEn}} = -\sum \log_2(p(s_i))^2$	Entropy measure based on log energy.
Kurtosis[2]	$k_2 = \frac{\sum(s_i-\mu)^4}{\left(\frac{1}{N}\sum(s_i-\mu)^2\right)^2} - 3$	Measures the peakedness or flatness of a distribution.
Skewness [2]	$b_1 = \frac{\sum(s_i-\mu)^3}{\left(\frac{1}{N-1}\sum(s_i-\mu)^2\right)^{3/2}}$	Quantifies asymmetry of the signal distribution.

Optimizing these parameters is crucial for improved time–frequency localization and accurate emotion recognition, with optimal Q enhancing decomposition and optimal r improving sub-band transitions and overlap. To achieve this, the metaheuristic COA is employed, which balances exploration and exploitation by mimicking crayfish behavior to iteratively update positions. The optimum Q and r values are selected by minimizing MSE, computed using Eq. (15), where $s(n)$ is the original signal and $y(n) = s(n)_{\text{reconstructed}}(Q, r)$ is the reconstructed signal.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(s(n)_i - y(n)_i)^2 \tag{15}$$

6.3.1.3 Statistical Features

The proposed method uses TQWT to decompose non-stationary EEG signals into multiple sub-band wavelets, from which discriminative features for emotion recognition are extracted (Table 6.8). Each EEG segment is represented as $s(t)_i$, with N denoting its length.

6.3.2 Results and Discussion

This section explains the acquired results and the interpretation of the corresponding data. It also details the sequence of stages involved in producing results and the empirical values utilized in each process.

6.3.2.1 Preprocessing

The SEED IV dataset comprises EEG signals recorded from 15 participants, each providing 24 trials (six per emotion across four emotions). Recordings were acquired with 62 electrodes, with sample lengths ranging from 8,000 to 45,000. For preprocessing, six recordings of the same emotion were combined, yielding 372 signals per emotion and 1,488 signals (4×372) per subject. As emotional responses occur in the latter part of recordings, only the final 4.6 seconds (4,600 samples at 1,000 Hz) were retained for analysis. Each EEG signal was band-limited with a fourth-order Butterworth filter (0.5–60 Hz). In Dataset-II, segmentation was performed using 10-second windows with 50% overlap, followed by band-pass and notch filtering to suppress noise and power-line interference.

Table 6.9: Parameter values used in the Crayfish Optimization Algorithm (COA).

Parameter	**Value**
Number of variables (dim)	2
Maximum Iteration (T)	50
Lower Bound (lb)	[1, 1]
Upper Bound (ub)	[50, 50]
Number of search agents (N)	2
Objective Function (f_{obj})	MSE (Eq. (15))

6.3.2.2 Optimization Using Crayfish Optimization Algorithm

Wavelets are vital for EEG analysis due to their multi-resolution and time–frequency localization. In TQWT, the quality factor (Q) controls oscillatory behavior, while redundancy (r) governs sub-band overlap; their optimal values are determined using the COA by minimizing MSE between original and reconstructed signals. This study applies COA to optimize Q and r with constant decomposition levels, which balance temporal and frequency resolution (lower levels favor temporal detail, higher levels favor frequency detail). Multiple levels (5, 8, 12, 15) were tested, with minimum MSE obtained

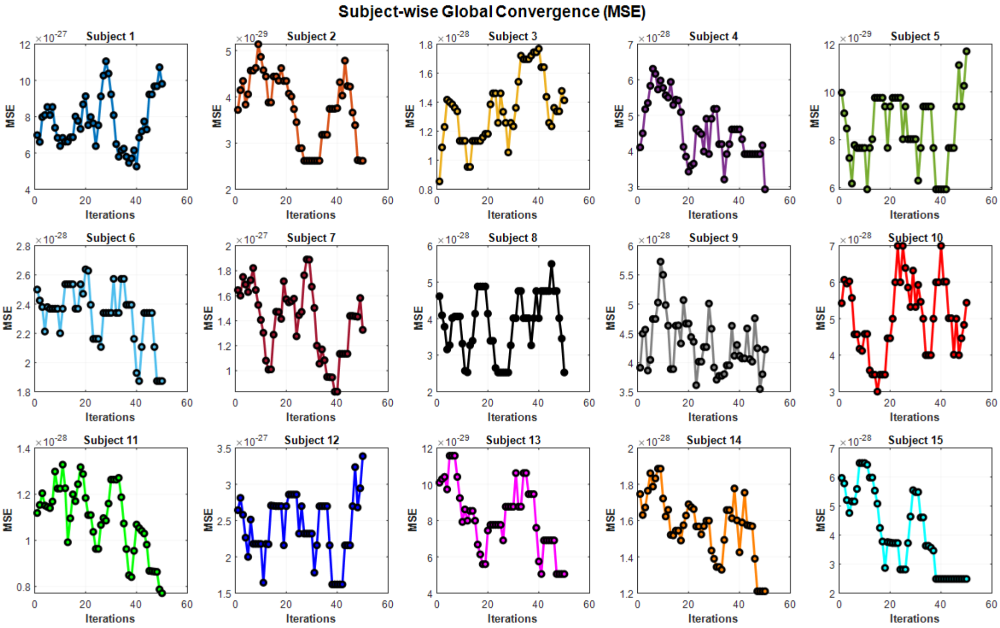


Figure 6.11: Convergence of optimal fitness during TQWT parameter optimization for EEG signal decomposition using MSE minimization (Dataset 1).

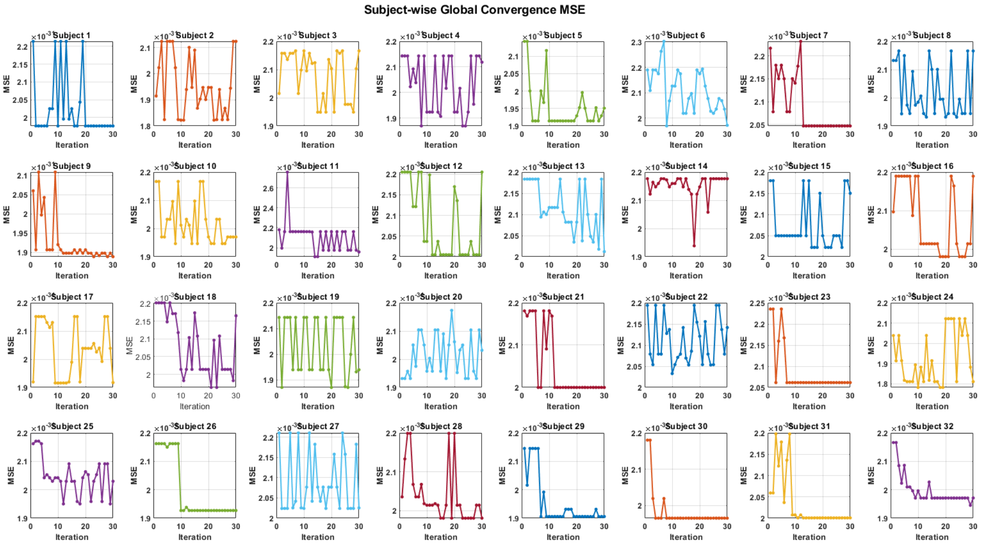


Figure 6.12: Convergence of optimal fitness during TQWT parameter optimization for EEG signal decomposition using MSE minimization (Dataset 2).

at 15 levels (16 sub-bands) for SEED IV and 7 levels (8 sub-bands) for DEAP. Algorithm 6.2 details the optimization of Q and r, while Table 6.9 provides COA parameters and objective functions. For Dataset-II, iterations were reduced to 30 to shorten optimization time, and the number of search agents was increased for better performance. Here, the number of variables (dim) denotes parameters to optimize, the number of agents (N) represents crayfish, and lower/upper bounds (lb/ub) define the search space range.

The COA determines the best tuning parameters that minimize the MSE in signal reconstruction, crucial for high-quality feature extraction and classification accuracy. The subject-wise mean MSE after optimization and the optimum value of the tuning parameters Q and r are displayed in Tables 6.10 and 6.11. The subject-wise EEG signals are decomposed into 16 sub-bands and 8 sub-bands for the SEEDIV and DEAP datasets using the tuning parameters given in Tables 6.10 and 6.11. Figures 6.11

Table 6.10: Average MSE before and after optimization, along with the optimized quality and redundancy factor values for Dataset I.

Subject	**MSE1**	**MSE2**	**Q**	**r**
Subject 1	5.50×10^{-27}	3.78×10^{-27}	2.42	19.57
Subject 2	5.40×10^{-29}	2.62×10^{-29}	9.38	9.85
Subject 3	8.55×10^{-29}	7.75×10^{-29}	9.97	13.78
Subject 4	3.28×10^{-28}	2.93×10^{-28}	8.66	18.71
Subject 5	6.64×10^{-29}	5.93×10^{-29}	8.63	33.18
Subject 6	2.20×10^{-28}	1.87×10^{-28}	1.22	14.38
Subject 7	1.43×10^{-27}	8.15×10^{-28}	23.35	9.20
Subject 8	2.58×10^{-28}	2.53×10^{-28}	9.50	12.57
Subject 9	5.20×10^{-28}	2.53×10^{-28}	9.38	9.85
Subject 10	3.00×10^{-28}	2.49×10^{-28}	2.58	36.27
Subject 11	9.17×10^{-29}	7.71×10^{-29}	12.94	5.94
Subject 12	1.66×10^{-27}	1.53×10^{-27}	9.21	13.67
Subject 13	1.09×10^{-28}	5.06×10^{-29}	2.66	25.94
Subject 14	1.39×10^{-28}	1.21×10^{-28}	2.37	7.85
Subject 15	5.15×10^{-28}	2.49×10^{-28}	2.60	23.29

Table 6.11: Average MSE before and after optimization, along with the optimized quality and redundancy factor values for Dataset II.

Subject	**Q**	**r**	**MSE**	**Subject**	**Q**	**r**	**MSE**
Subject 1	6.07	11.90	1.98×10^{-31}	Subject 17	6.08	19.15	2.15×10^{-31}
Subject 2	2.27	20.70	1.82×10^{-31}	Subject 18	6.44	19.23	2.01×10^{-31}
Subject 3	6.83	11.79	1.95×10^{-31}	Subject 19	2.22	15.00	1.87×10^{-31}
Subject 4	2.93	20.36	1.87×10^{-31}	Subject 20	2.15	15.78	1.93×10^{-31}
Subject 5	3.75	5.11	1.91×10^{-31}	Subject 21	3.18	1.20	2.00×10^{-31}
Subject 6	6.72	11.90	1.97×10^{-31}	Subject 22	5.69	26.64	2.03×10^{-31}
Subject 7	3.07	2.03	2.05×10^{-31}	Subject 23	1.05	2.20	2.06×10^{-31}
Subject 8	2.64	15.11	1.93×10^{-31}	Subject 24	2.46	15.55	1.78×10^{-31}
Subject 9	2.96	8.24	1.89×10^{-31}	Subject 25	6.78	11.14	1.95×10^{-31}
Subject 10	6.87	11.31	1.95×10^{-31}	Subject 26	3.21	1.20	1.93×10^{-31}
Subject 11	2.84	24.26	1.91×10^{-31}	Subject 27	3.92	5.51	2.02×10^{-31}
Subject 12	3.77	1.82	2.00×10^{-31}	Subject 28	3.64	5.60	1.98×10^{-31}
Subject 13	2.56	15.48	2.01×10^{-31}	Subject 29	3.47	5.46	1.91×10^{-31}
Subject 14	2.22	15.48	1.94×10^{-31}	Subject 30	3.16	1.32	1.97×10^{-31}
Subject 15	2.35	24.92	2.02×10^{-31}	Subject 31	2.92	1.09	2.00×10^{-31}
Subject 16	6.86	11.65	1.98×10^{-31}	Subject 32	3.22	5.58	1.95×10^{-31}

and 6.12 show that there is a significant decrease in MSE after optimization. This adaptability in COA helps achieve robust parameter optimization, contributing to the model's overall reliability and effectiveness. The decomposed signals are further used for feature extraction and selection.

6.3.2.3 Feature Selection and Extraction

The Kruskal–Wallis (KW) test is applied to identify significant features, using probabilistic values to assess their discriminative ability across four groups. The p-value indicates similarity, with $p < 0.05$ recommended for effective discrimination; lower p-values denote higher statistical significance. The KW test merges all decomposed signal sub-bands, retaining features with $p < 0.05$ as statistically significant for emotion discrimination, while excluding those above this threshold. Since classification spans multiple dimensions, a few weak features have negligible impact.

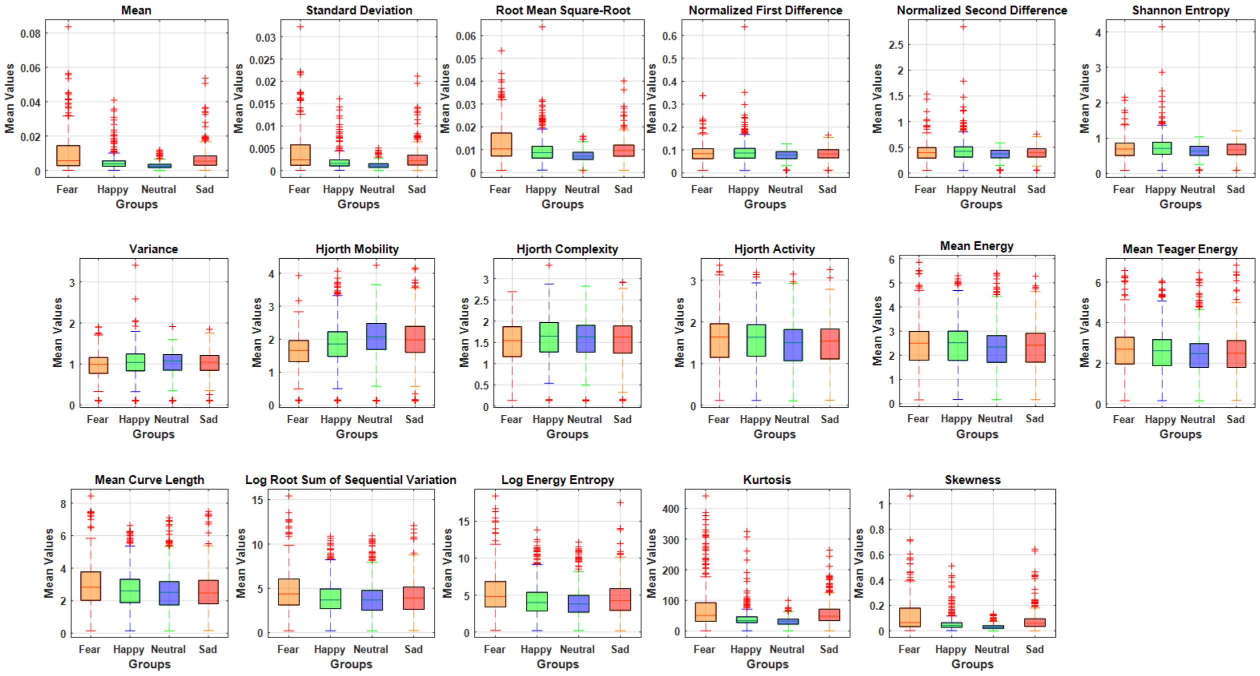


Figure 6.13: The box plots of the extracted feature values for Dataset 1 illustrate the ability to distinguish between class-wise distributions.

Box plots (Figures 6.13 and 6.14) display mean feature distributions across groups, using color coding for clarity, and confirm that all 17 extracted time-domain features significantly distinguish between four and three emotions. From TQWT sub-bands, 17 features are computed, yielding column matrices of size 16 × 17 (272) for Dataset-I and 8 × 17 (136) for Dataset-II, which are then applied to classification algorithms.

In the classification stage, 75% of the features are used for training and 25% for testing (subject-dependent). Figures 6.15 show classification accuracies for datasets 1 and 2 across multiple classifiers. Among tree models, the optimized tree performs best, followed by fine, medium, and coarse variants. For SVM, cubic and optimized versions achieve the highest accuracy, followed by quadratic, fine-gaussian, medium, linear, and gaussian. In k-NN, optimized, fine, and weighted variants outperform medium, cosine, cubic, and coarse versions. For ensemble models, optimized, bagged, and subspace variants perform better than subspace k-NN, boosted, and rusboosted. Neural networks show strong performance across all variants, with optimized, wide, and medium architectures outperforming narrow, bi-layered, and tri-layered versions. Overall, SVM, ensemble, and neural networks are the most effective classifiers, with the best results obtained by cubic SVM and optimizable neural networks. Ensemble models also provide competitive performance, while tree-based and simpler k-NN models generally underperform. Optimized variants across all classifiers consistently yield superior accuracy, highlighting the importance of hyperparameter tuning. This work employs Bayesian optimization for hyperparameter tuning, where an acquisition function balances exploration (searching unknown

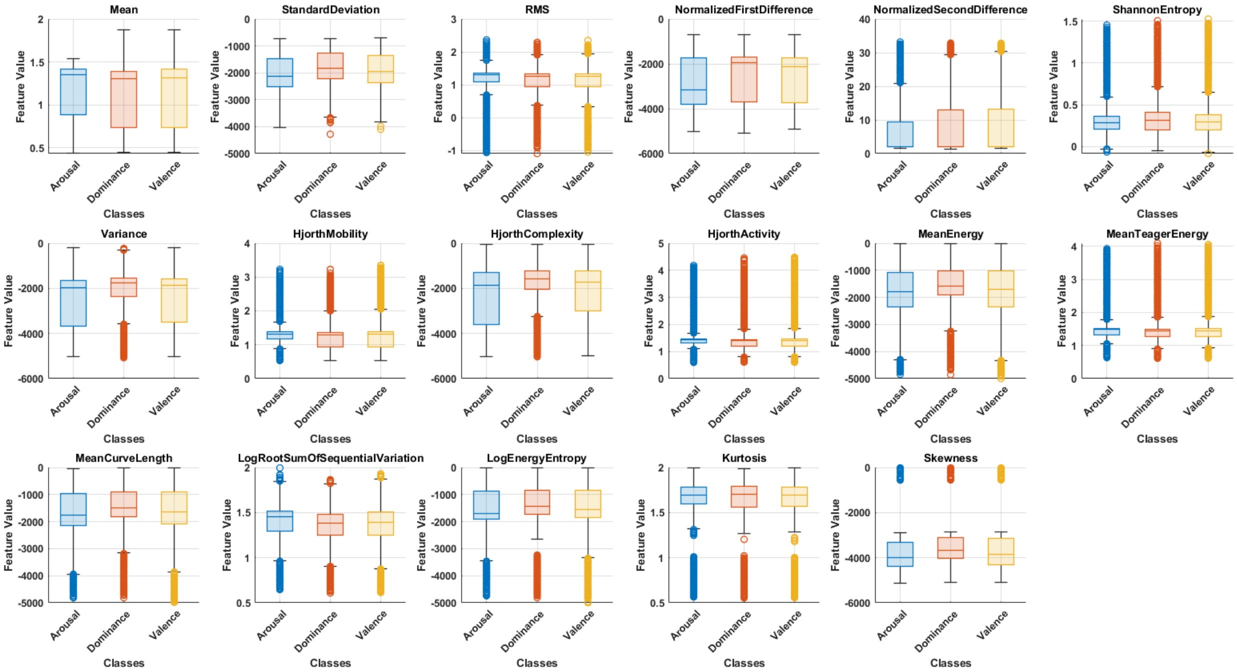


Figure 6.14: The box plots of the extracted feature values for Dataset 2 illustrate the ability to distinguish between class-wise distributions.

regions) and exploitation (refining promising regions) when selecting the next evaluation point. Compared with grid and random search, Bayesian optimization is more time-efficient. Cross-validation validates the final hyperparameters, which proved robust by achieving the highest validation accuracy. The optimized hyperparameters yielded strong performance across subjects in Datasets I and II.

6.3.2.4 Performance Comparison

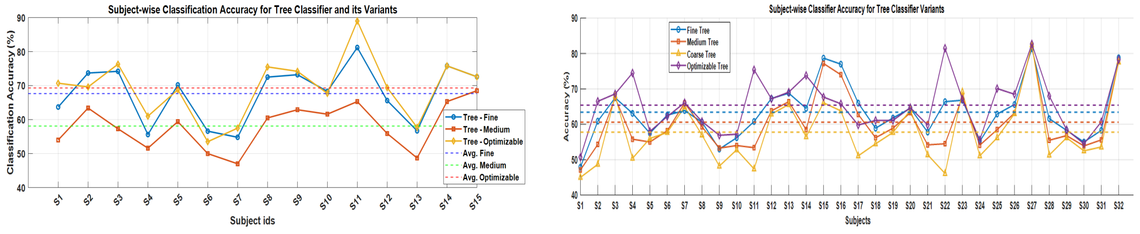
Table 6.12 presents the mean values of performance measures, including Precision (PRE), Accuracy (ACC), Error (ERR), Recall (RC), F1-Score (F1), Kappa (KP), and Matthew's Correlation Coefficient (MCC). Accuracy reflects correct predictions, ERR indicates misclassification rate, and F1 balances precision and recall.

For Dataset-I, ERR, PRE, RC, F1, MCC, and KP ranged from 04.70–31.24%, 68.90–95.3%, 68.76–95.3%, 68.71–95.3%, 59.44–88.35%, and 0.6861–0.9372, respectively. For Dataset-II, these values ranged from 11.50–34.61%, 64.89–88.90%, 65.32–87.80%, 64.28–88.%, 39.48–72.63%, and 0.52–0.8159. The values are close to ideal, confirming the method's reliability and stability.

Cubic SVM achieved the best performance with mean accuracies of 91.46% (Dataset-I, Subject-Dependent) and 77.21% (Dataset-II, Subject-Dependent), showing consistent results across datasets. Furthermore, neural networks, SVM, and ensemble methods yielded lower classification errors compared to k-NN and decision tree.

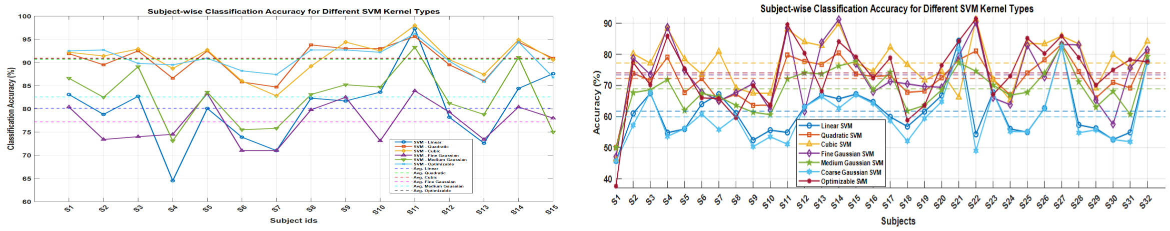
Figure 6.15 shows that subject 11 (Dataset-I) and subject 22 (Dataset-II) achieved the highest mean classification accuracies, with the optimized neural network reaching 98.1% and 91.7%, respectively. The confusion matrix (Figure 6.16) indicates high true positives for Dataset-I: Fear (98.9%), Happy (97.8%), Neutral (98.9%), and Sad (96.8%). For Dataset-II, true positives were Arousal (89.2%), Dominance (90.1%), and Valence (92.9%), confirming strong agreement between actual and predicted values.

ROC analysis (Figs. 6.16 reported AUCs of 0.9899, 0.9953, 0.9966, 0.9987 (Dataset-I), and 0.9722, 0.9755, 0.9694 (Dataset-II), demonstrating precise class separability. Precision-recall AUCs were



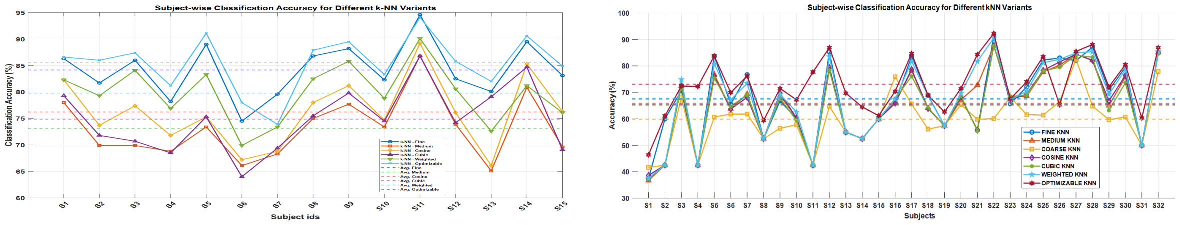
(a) Tree - Dataset I

(b) Tree - Dataset II



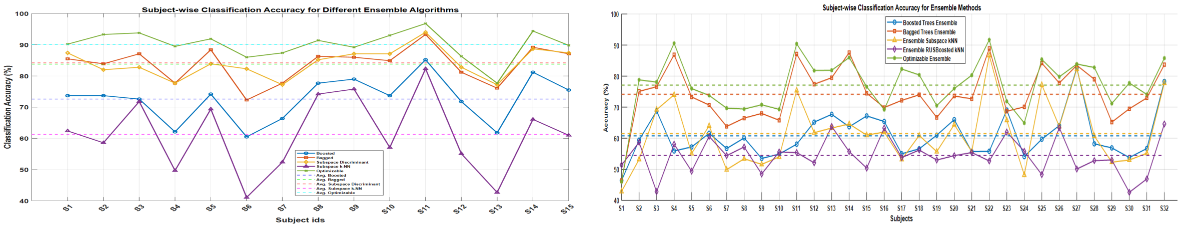
(c) SVM - Dataset I

(d) SVM - Dataset II



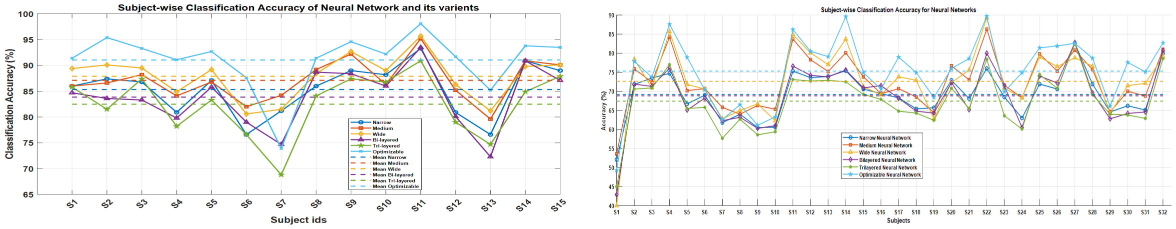
(e) k-NN - Dataset I

(f) k-NN - Dataset II



(g) Ensemble - Dataset I

(h) Ensemble - Dataset II



(i) Neural Network - Dataset I

(j) Neural Network - Dataset II

Figure 6.15: Classification performance comparison across five classifiers on Datasets I and II.

Table 6.12: Classifier performance comparison: Highlight shows best overall performer (Cubic SVM).

	Dataset I						**Dataset II**					
Classifier	**ACC**	**ERR**	**PRE**	**REC**	**F1**	**MCC**	**ACC**	**ERR**	**PRE**	**REC**	**F1**	**MCC**
O. Tree	68	31	68	68	68	59	65	34	64	65	64	63
O. SVM	90	9	91	90	90	87	74	25	74	74	74	67
Cubic SVM	**90**	**9**	**91**	**90**	**90**	**87**	**77**	**22**	**77**	**77**	**76**	**71**
O. k-NN	85	14	86	85	85	81	73	26	73	72	72	72
O. Ensemble	89	10	90	90	90	87	77	22	76	77	76	72
O. Neural Net.	91	8	91	91	91	88	75	24	75	75	75	72

0.9916, 0.993, 0.9732, 0.9956 (Dataset-I) and 0.9454, 0.9308, 0.9665 (Dataset-II), showing high precision maintained across increasing recall. Overall, results highlight the model's strong discriminative ability and robustness across both datasets.

6.3.2.5 Computational Time Comparison

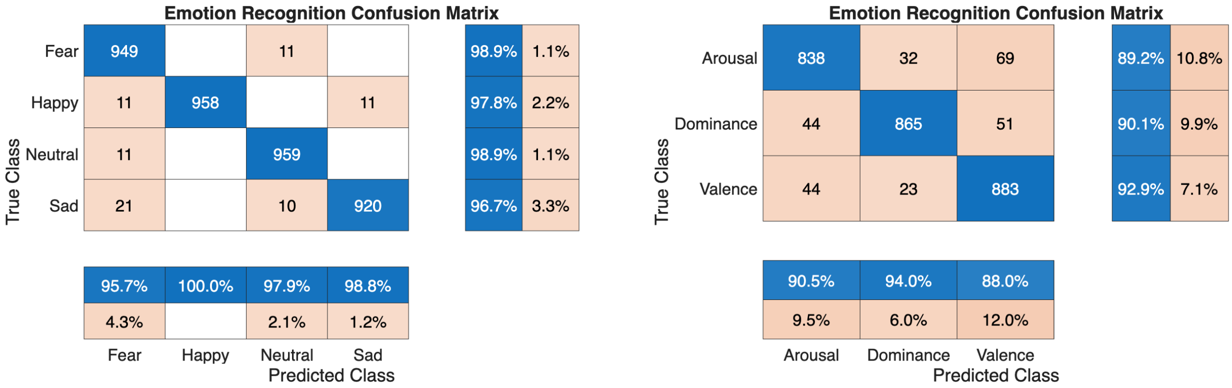
The computational time of the proposed work comprises four stages: optimization, signal decomposition, feature extraction, and model training. TQWT parameters (Q and r) are optimized once offline using crayfish optimization algorithm and then reused for EEG decomposition. Owing to its adaptability, TQWT is more efficient than fixed filter banks. Feature extraction (mean, variance, entropy, etc.) is lightweight, making it suitable for real-time use. Subject-wise computational times for datasets 1 and 2. For Dataset-I, the average times are 3.96 s for decomposition, 12.88 s for feature extraction, 277.34 s for training, and 294.18 s total. For Dataset-II, the corresponding times are 11.90 s, 36.26 s, 1166.88 s, and 1215.04 s. The higher computational cost of Dataset-II arises from its larger size (32 subjects, more epochs), which generates more features and increases processing time.

6.3.3 Comparison With Previous Studies

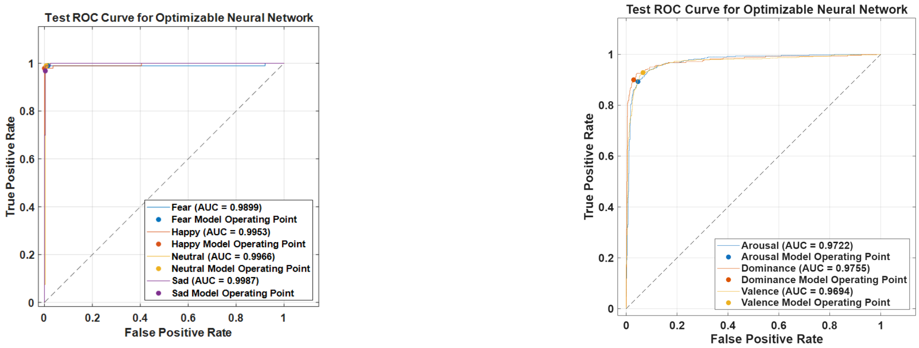
Table 6.13 summarizes major findings on SEED IV and DEAP datasets. On SEED IV, Zhong et al. [74] achieved 79.% accuracy using a regularized GNN, leveraging interchannel asymmetry learning to reduce overfitting. Jiyao et al. [75] introduced a spatiotemporal transformer, improving accuracy by 2.96% over the best spatio-temporal model. Sijin et al. [76] proposed SOFNN, yielding 84% accuracy for happy emotion and improving sadness, fear, and neutral by 11%, 12%, and 9%. Temporal fusion further raised accuracy to 82.31%, outperforming spatial (74.62%) and entropy (71.82%) features. Guarneros et al. [152] presented FSA-TSP with MMD losses, improving accuracy by 1.05–2.66%. Chang et al. [153] developed a spatial-gated graph transformer, reaching 88.62% mean accuracy, with sadness recognition at 93.48% and other emotions between 85–90%.

On DEAP, Zhang et al. [101] proposed RDFKM, improving accuracy by 0.6–11% and F1-score by 0.1–6%. Li et al. [154] achieved 90.3% accuracy using Hjorth parameters in the beta band, showing linear features are both efficient and accurate. Kumar et al. [118] applied the superlet transform with AdaBoost, achieving 62.57% accuracy.

The proposed method identifies significant features and optimized classifiers, consistent with the "no free lunch theorem" [155], highlighting the importance of both feature selection and model choice in emotion recognition.



(a) Dataset 1 – Confusion Matrix (b) Dataset 2 Confusion Matrix



(c) Dataset 1 – ROC Curve (d) Dataset 2 – ROC Curve

Figure 6.16: Performance metrics of optimal TQWT parameter selection using Crayfish Optimization: confusion matrices and ROC curves.

6.4 Adaptive Quality Factor Tuning in Wavelet Scattering Transform Using Particle Swarm Optimization

6.4.1 Methodology

The block diagram of the proposed methodology is displayed in Figure 6.17. The block diagram consists of preprocessing, band-pass and notch filtering, VMD-based mode filtering, optimization of the wavelet scattering quality factor, deep scattering feature extraction, and classification using k-NN, SVM, and ensemble classifiers. Two datasets were used, with details provided below.

6.4.1.1 Data Pre-processing

EEG signals are processed using a 0.5–60 Hz band-pass filter to retain relevant brain activity, followed by a 50 Hz notch filter to remove power-line interference. For emotion detection, 10-second segments from Dataset I and 15-second segments from Dataset II of the filtered EEG data are used.

6.4.1.2 Variational Mode Decomposition (VMD)

Variational mode decomposition [124] decomposes non-stationary signals like EEG into several oscillatory modes, adapting to time-varying frequencies and non-stationary features. EEG signals (0.5–60 Hz) may include noise or frequency fluctuations, which can be mitigated using VMD-based

Table 6.13: Comparison of previous studies on EEG-based emotion recognition using SEED-IV and DEAP datasets.

Author (Year)	**Classes**	**Dataset**	**Method / Features / Classifier**	**Accuracy (%)**
SEED-IV Dataset Studies				
Zhong et al.[74] (21)	4	SEED-IV	Regularized Graph Neural Network (RGNN)	79.37 ± 10.54
Jiyao et al.[75] (22)	4	SEED-IV	Transformer model	83.27 ± 8.37
Sijin et al. [76] (22)	4	SEED-IV	SOFNN	86.27 ± 10.16
Guarneros et al.[152](23)	4	SEED-IV	FSA-TSP	85.45 ± 9.81
Yadong et al. [153] (24)	4	SEED-IV	SGGT	88.62 ± 8.01
Proposed Method (2025)	4	SEED-IV	**Time-Domain Features + Cubic SVM**	**91.47 ± 4.9**
DEAP Dataset Studies				
Zange et al.[101] (21)	2	DEAP	PSD features + SVM	58.00
Li et al.[154] (18)	2	DEAP	STFT + SVM	59.06
Kumar et al.[118] (25)	2	DEAP	Superlet Transform + Adaboost	62.57
Proposed (2025)	3	DEAP	**Time-Domain Features + Cubic SVM**	**77.81 ± 9.15**

IMF filtering. The original EEG signal $y(t)$ is represented as a sum of AM-FM modes:

$$y(t) = \sum_{k=1}^{K} u_k(t), \quad u_k(t) = A_k(t)\cos(\phi_k(t)), \tag{6.16}$$

where $A_k(t)$ and $\phi_k(t)$ are slowly varying envelopes and phases centered around frequencies f_k.

VMD computes all mode waveforms and their central frequencies simultaneously by solving a constrained variational problem via the augmented Lagrangian. Iterative updates determine the modes $\hat{u}_k^{(n+1)}(f)$ and center frequencies $f_k^{(n+1)}$, controlled by parameters such as the balancing factor α and Lagrangian multiplier λ. The extracted modes oscillate around their center frequencies, and the instantaneous frequency (IF) evaluates their frequency ranges. Residual signals represent the portion not decomposed, enabling noise removal through mode filtering.

Key VMD parameters include Absolute Tolerance (AT), Relative Tolerance (RT), Max Iterations (MI), number of IMFs (nIMF), Penalty Factor (PF), Lagrange Multiplier Update Rate (LMUR), initialization method (IM), and Pearson correlation coefficient (PCC) for assessing mode similarity.

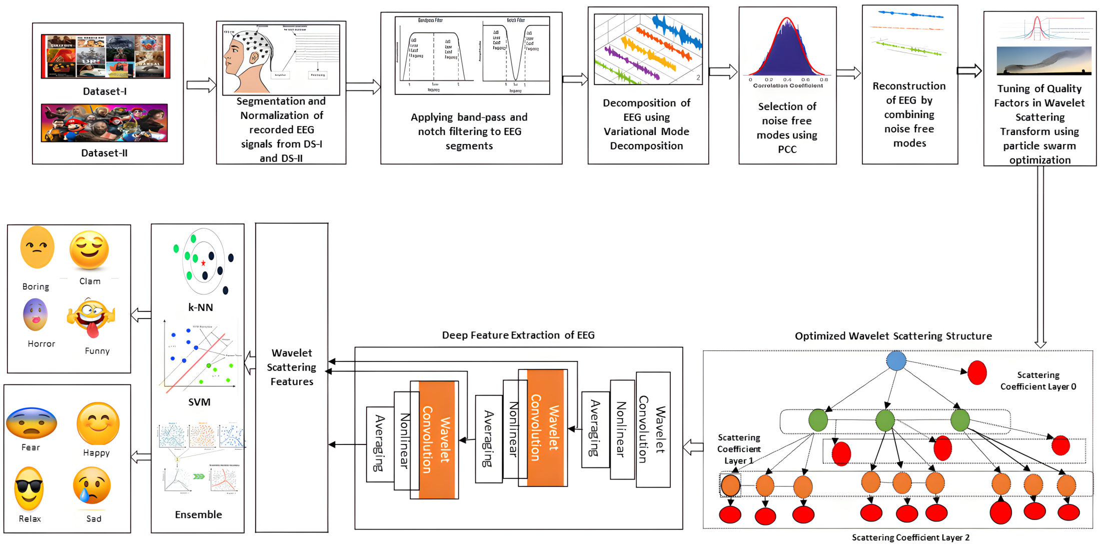


Figure 6.17: Block diagram of the proposed methodology for adaptive quality factor tuning in wavelet scattering transform using PSO.

Pearson's correlation ρ is calculated as:

$$\rho = \frac{\mathrm{COV}(P,Q)}{\sigma_P \sigma_Q} = \frac{\sum(p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum(p_i - \bar{p})^2}\sqrt{\sum(q_i - \bar{q})^2}}, \tag{6.17}$$

Algorithm 6.3 : Noise Removal from EEG Signals using VMD

```
Input: EEG signals, VMD parameters
Output: Reconstructed EEG signals without noisy modes
Initialize parameters:
    $AT \leftarrow 5 \times 10^{-6}$, $RT \leftarrow 0.005$, $MI \leftarrow 500$
    $nIMF \leftarrow N$, $PF \leftarrow 1000$, $LMUR \leftarrow 0.01$
    $IM \leftarrow$ peaks, $PCC \leftarrow 0.4$
Initialize reconstructed array
for $i \leftarrow 1$ to $M$ do
    Perform VMD on signal $i$
    Initialize selected IMF array
    for $j \leftarrow 1$ to $N$ do
        Calculate PCC between $\mathrm{IMF}_j$ and original signal
        if $PCC \geq 0.4$ then
            Add $\mathrm{IMF}_j$ to selected IMF array
        end if
    end for
    Reconstruct signal using selected IMFs and residual
    Store reconstructed signal in array
end for
```

facilitating selection of noise-free modes [132]. Algorithm 6.3 outlines the step-by-step VMD-based noise removal.

It is possible to determine the instantaneous frequency (IF) of the m-th mode $u_m(t)$ by its analytical representation as:

$$V_m(t) = u_m(t) + j\mathcal{H}\{u_m(t)\} = A_m(t)e^{j\phi_m(t)}, \tag{6.18}$$

where $\mathcal{H}\{u_m(t)\}$ is the Hilbert transform of $u_m(t)$, $A_m(t)$ is the amplitude, and $\phi_m(t)$ is the phase. The intermediate frequency (IF) of the m-th mode can be calculated as:

$$\text{IF} = \frac{1}{2\pi}\frac{d\phi_m(t)}{dt} \tag{7}$$

The m-th mode's IF range is described as:

$$L_m = \mu_m + \sigma_m \tag{8}$$

Where μ_m, L_m, and σ_m represent the mean, range, and standard deviation, respectively. The part of the original signal $y(t)$ that VMD cannot decompose is represented by the residual. By using these details VMD can be used to remove the noise. The step-by-step details for removing noise using mode filtering are given in Algorithm 6.3. The value of N is chosen according to the number of IMFs selected. The parameters used for VMD decomposition are Absolute Tolerance (AT), Relative Tolerance (RT), Max Iterations (MI), Num IMF (nIMF), Penalty Factor (PF), LM Update Rate (LMUR), Initialize Method (IM), and Pearson's correlation coefficient (PCC). The Pearson's correlation can be defined as

$$\begin{aligned}\rho &= \frac{\text{COV}(P,Q)}{\sigma_M\sigma_N}\\ &= \frac{\mathbb{E}(PQ)-\mathbb{E}(P)\mathbb{E}(Q)}{\sqrt{\mathbb{E}(P^2)-\mathbb{E}(P)^2}\sqrt{\mathbb{E}(Q^2)-\mathbb{E}(Q)^2}}\\ &= \frac{\sum(p_i-\bar{p})(q_i-\bar{q})}{\sqrt{\sum(p_i-\bar{p})^2}\sqrt{\sum(q_i-\bar{q})^2}}\end{aligned} \tag{9}$$

The range of values of PCC and its correlation intensity is defined in Table 3.1 [156].

The threshold PCC defined for the algorithm is 0.4 [156]. Noisy signals are weakly correlated with the original signal. The VMD algorithm decomposes the non-stationary signal into IMFs and residues. IMFs with weak correlation are filtered out, and IMFs with moderate to high correlation are kept for signal reconstruction.

6.4.1.3 Wavelet Scattering Transform (WST)

The Wavelet Scattering Transform [122] performs time-frequency analysis, offering strong discrimination, stability against time-warping, non-linearity, and translation invariance. Figure 6.18 shows that WST resembles a CNN by using iterative complex wavelet transforms, replacing data-driven filters with wavelets, and employing complex-modulus operations instead of pooling and low-pass averaging.

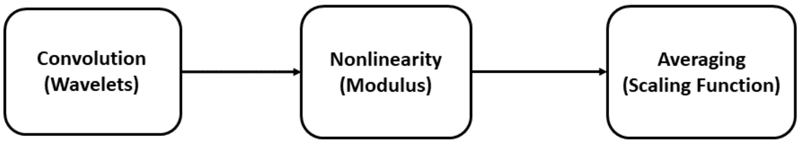


Figure 6.18: Stages of the wavelet scattering transform.

Figure 6.19 illustrates the wavelet scattering layout, which involves the input data I, wavelets $\phi_{i,j}$, and the scaling function τ_J. The process begins with the convolution of I with τ_J to obtain the zero-order scattering coefficients $S[0]$. Multiple wavelet rotations are then applied, and the modulus of

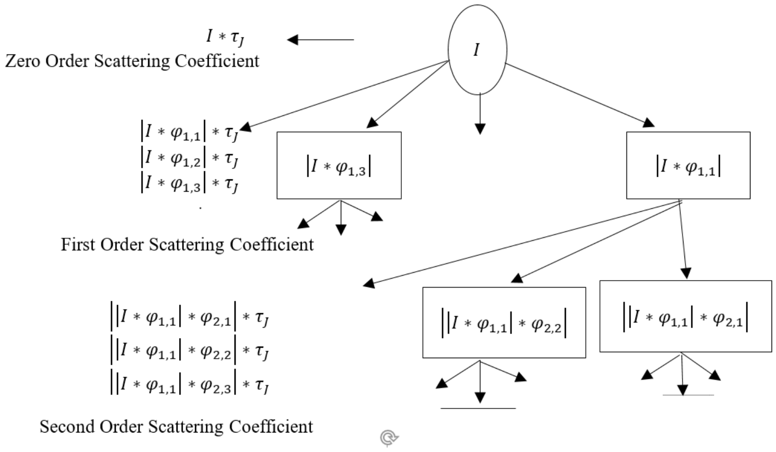


Figure 6.19: Two-stage framework of the wavelet scattering transform.

each filtered output produces the scalogram $U[1]$. Convolution with the scaling function generates the first-order scattering coefficients $S[1]$, while iterative application of wavelet transforms and modulus operations yields the second-order coefficients $S[2]$. Although low-pass filtering initially leads to information loss, this is recovered in subsequent coefficient calculations. Default filter banks use eight wavelets per octave for the first-order and one wavelet per octave for the second-order scattering, forming the wavelet temporal scattering network.

6.4.1.4 Optimization of Quality factor in WST

The wavelet scattering transform important parameters include the invariance scale, quality factor (Q), number of filter banks, number of wavelets per octave, number of rotations (for 2D scattering), sampling frequency, number of scattering orders, and wavelet type. Among these, the quality factor is the most critical, as it directly influences the scattering characteristics by determining the number of wavelet filters per octave.

The quality factor controls the division of the frequency axis within each octave. A higher value results in closely spaced wavelet filters, enhancing frequency resolution and enabling the WST to detect subtle frequency variations. Conversely, a lower quality factor reduces computational complexity but may lead to the loss of important frequency information. Therefore, optimizing the quality factor is essential to balance feature extraction quality and computational cost.

In each iteration, a machine learning algorithm (e.g., k-NN or SVM) classifies the features extracted from EEG data using the current quality factor for each particle. Particle swarm optimization [104] then updates the quality factor based on particle performance, using classification error as the objective function. Through multiple iterations, this approach identifies optimal quality factors that maximize classification performance, achieving a balance between exploration (searching for new quality factors) and exploitation (refining the best-known values). Figure 6.20 illustrates the quality Factor optimization procedure in WST, while algorithm 6.4 details the step-by-step process of combining WST with PSO to maximize classification accuracy.

6.4.1.5 Feature Matrix using WST

The wavelet scattering algorithm divides a signal into multiple windows, extracts features from each using wavelet transforms, and categorizes each window separately. The final categorization is based on uniform weighting, giving equal influence to each window. Scattering coefficients are robust to small input variations and efficiently represent key features.

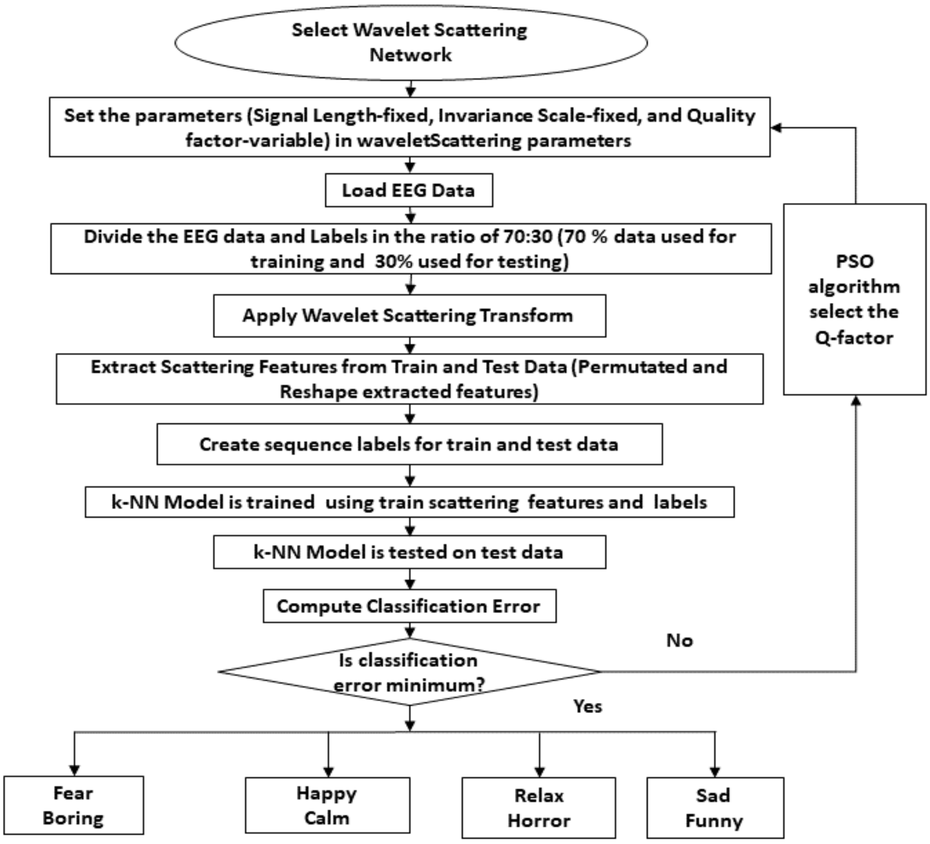


Figure 6.20: Flowchart of the algorithm for optimizing the wavelet scattering transform.

6.4.2 Results and Discussion

This section details the proposed procedure, covering experimental parameters and outcomes, including data preprocessing, mode filtering, WST quality factor optimization, scattering feature extraction, and classification.

6.4.2.1 Segmentation and Filtering

Datasets I and II are segmented into 10- and 15-second intervals and normalized, using frontal electrode recordings due to their relevance in emotion detection. EEG signals are filtered with a 0.5–60 Hz Butterworth bandpass and a 50 Hz notch IIR filter. Dataset I has a 256 Hz sampling frequency (Nyquist 128 Hz), and Dataset II has 128 Hz (Nyquist 64 Hz). Filter order is computed based on the passband and Nyquist frequency, ensuring an even order. The filtered EEG segments are then processed using VMD.

6.4.2.2 Noisy Mode Filtering

VMD decomposes signals into multiple intrinsic mode functions (IMFs) representing high, moderate, and low frequencies. The Pearson correlation coefficient (PCC) measures similarity between waveforms, with Table 3.1 defining correlation intensity and a threshold set at 0.4 [156].

Algorithm 6.4 : Wavelet Scattering Transform Optimization using Particle Swarm for Quality Factor Analysis

```
Initialize PSO Parameters: Set number of particles $P$, iterations $T$, and inertia weight $\omega$. Initialize particle positions (quality factors $Q_i$) and velocities.
Load EEG Signal: Import EEG data for training and testing.
Objective Function: Minimize classification error.
for $t \leftarrow 1$ to $T$ do
    for $i \leftarrow 1$ to $P$ do
        Apply WST for Particle $i$:
            Compute scattering coefficients using current $Q_i$
            Extract scattering features
        Classification: Use k-NN or SVM to classify extracted features.
        Error Calculation: Compute classification error for particle $i$.
        Update $P_{\text{best}}$ (local best) and $G_{\text{best}}$ (global best).
    end for
    Update Velocities and Positions:
    for $i \leftarrow 1$ to $P$ do
        $v_i \leftarrow \omega v_i + c_1 r_1 (P_{\text{best}} - Q_i) + c_2 r_2 (G_{\text{best}} - Q_i)$
        $Q_i \leftarrow Q_i + v_i$
    end for
end for
Return: $G_{\text{best}}$ as optimal quality factors.
```

IMFs with PCC below 0.4 indicate strong dissimilarity and are considered noise, while those with PCC ≥ 0.4 strongly resemble the original EEG and are used for reconstruction. For example, IMF3, IMF4, and IMF5 (PCCs 0.4001, 0.6637, 0.8057) are retained, producing a reconstructed signal free of redundant noise. Algorithm 6.3 applies this filtering to all signals in Datasets I and II.

6.4.2.3 Optimization of Wave Scattering Transform (O-WST)

A wavelet scattering transform has been used to extract the features from the VMD-filtered EEG signals. The performance of WST is improved by selecting the optimum value of the quality factor in WST using particle swarm optimization. Table 6.14 displays the values of parameters in PSO. In PSO, the number of variables represents the parameters to optimize, the population size is the number of particles, and the maximum iterations define the stopping condition. Inertia weights W_{Max} and W_{Min} balance exploration and exploitation, while c_1 and c_2 control the influence of personal and global best particles. The quality factor is constrained to integers, and PSO aims to minimize classification error using WST-generated scattering features. These features are divided into training and testing sets for kNN classification, and multiple quality factor values are evaluated to find the optimum minimizing error. The optimized quality factors are presented in Table 6.15, with Dataset-I and Dataset-II optimization illustrated in Figures 6.21a and 6.21b, respectively, using Algorithm 6.4.

6.4.2.4 Feature Extraction From Optimized Wave Scattering Transform

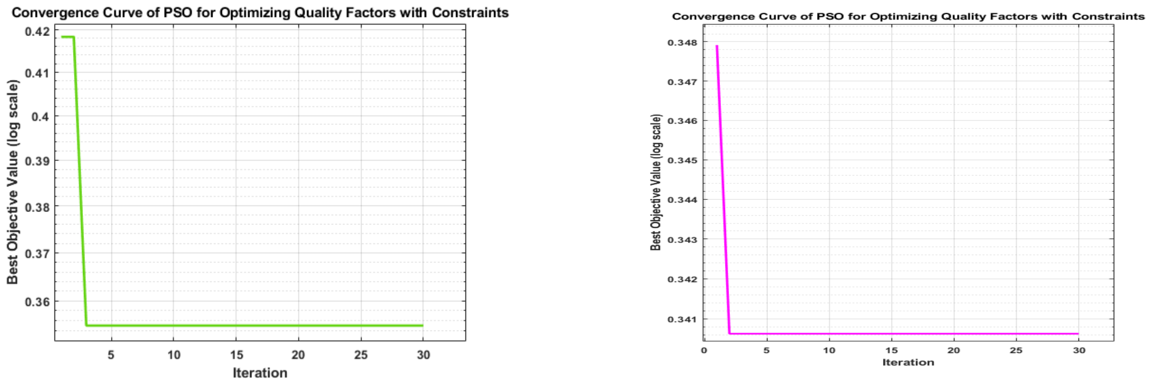
The filtered EEG signals are processed using an optimized wave scattering transform (O-WST) for feature extraction. O-WST yields stable, translation-invariant representations in the form of zeroth-, first-, and second-order scattering coefficients across multiple filters. These coefficients capture the signal's structure and frequency information, represented through scalograms that highlight nonlinear dynamics. The resulting feature matrix is a three-dimensional tensor (a×b×c), where a denotes the scattering route, b the wavelet scale, and c the number of signals. The wavelet scattering function constructs a time scattering network using default settings. For Dataset I, the first filter bank employs two wavelets per octave and the second uses one, while for Dataset II, the first uses 21 wavelets per octave and the second uses eight. The invariance scale equals the signal length, and the optimized path parameter is set to true. Dataset I contains 7,904 filtered recordings (1,976 per emotion, each representing 25% of the total), with 5,533 used for training and 2,371 for testing. Dataset II comprises 28,224 filtered recordings (7,056 per emotion, each 25%), with 19,757 for training and 8,467 for testing.

Table 6.14: PSO parameter values for quality factor optimization.

Parameter	**Value**
Number of variables	2
Maximum Iteration	30
Population Size	50
W_{Max}	0.8
W_{Min}	0.6
Acceleration Coefficient	$c_1 = 2, c_2 = 1$
Lower Bound (lb)	[1, 1]
Upper Bound (ub)	[30, 30]
v_{Max}	(ub - lb)×0.2
v_{Min}	$-v_{\text{Max}}$

Table 6.15: Optimized parameters of the wavelet scattering network.

O-WST Parameters	**Dataset-I**	**Dataset-II**
Signal Length	2560	1920
Invariance Scale	2560	1920
Quality Factor	[8, 1]	[21, 8]



(a) Quality factor optimization for Dataset I.

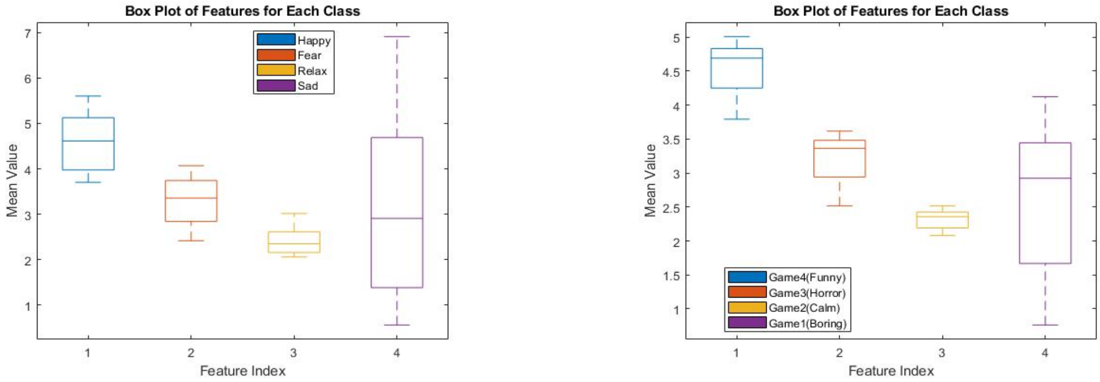
(b) Quality factor optimization for Dataset II.

Figure 6.21: Minimization of classification error as the objective function in PSO.

6.4.2.5 Features from O-WST

For Dataset I, the scattering transform is computed for 5,533 training and 2,371 test (subject-independent) recordings, treating each column of the multivariate time series as a separate signal. The natural logarithm of the scattering coefficients is applied, producing a 3-D feature matrix of size 166 × 5 × 5533 for training, where 166 is the number of scattering paths and 5 is the scattering window. This tensor is reshaped into 27,665×166 (5×5533) for classifier input, with each row representing one window across 166 paths. The test set yields features of size 11,855×166. For Dataset II, the training data dimension is 19,757×1920, and the scattering feature matrix is 162×4×19757, where 162 denotes scattering paths and 4 the scattering window. After reshaping, this becomes 79,028×162 (4×19757).

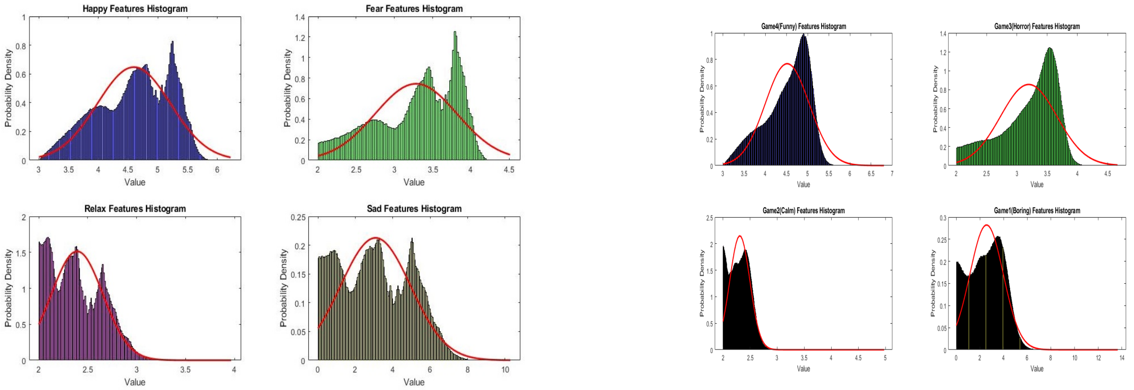
The test feature matrix has dimensions 33,868×162 (subject-independent).



(a) Box-plot of features from four classes of emotions Dataset-I.

(b) Box-plot of features from four classes of emotions Dataset-II.

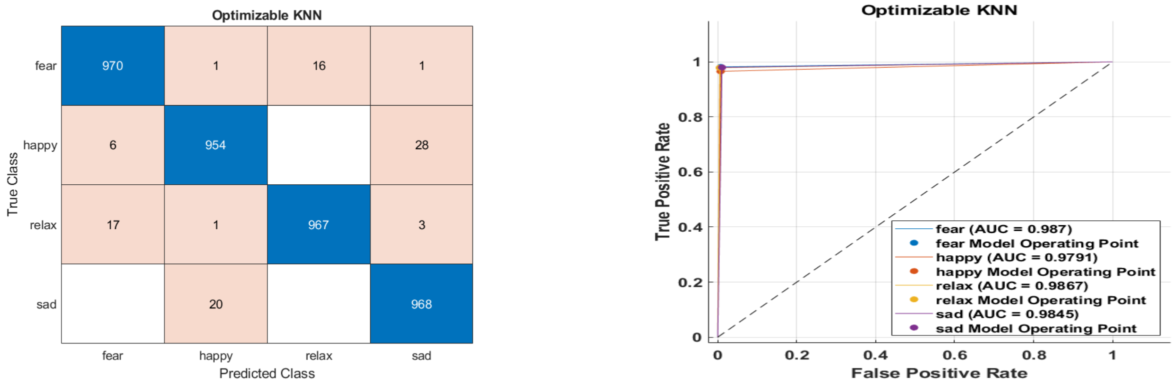
Figure 6.22: Box-plots of features from two datasets of emotions.



(a) Histogram plot of features from four classes of emotions Dataset-I.

(b) Histogram plot of features from four classes of emotions Dataset-II.

Figure 6.23: Histogram plot of feature from two datasets of emotions.



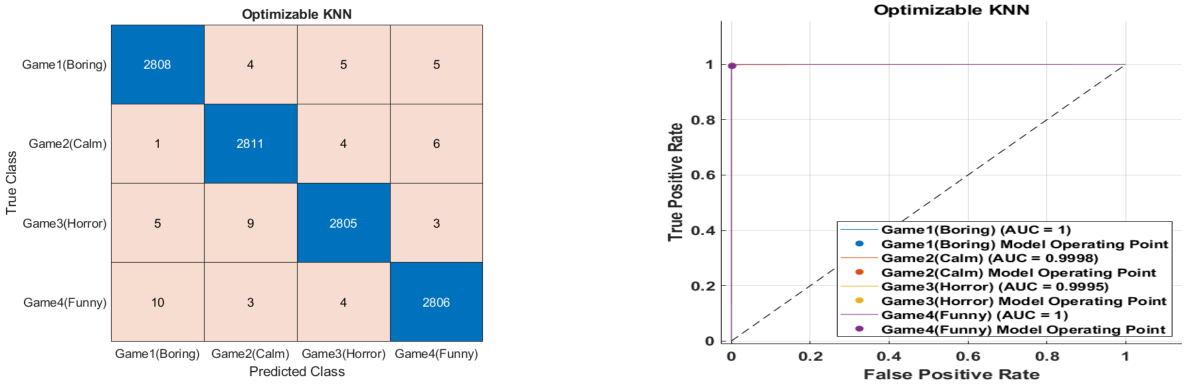
	fear	happy	relax	sad
fear	970	1	16	1
happy	6	954		28
relax	17	1	967	3
sad		20		968

(a) Confusion Matrix of Dataset-I.

(b) ROC of four classes of Dataset-I.

Figure 6.24: Confusion matrix and ROC curve for Dataset I.

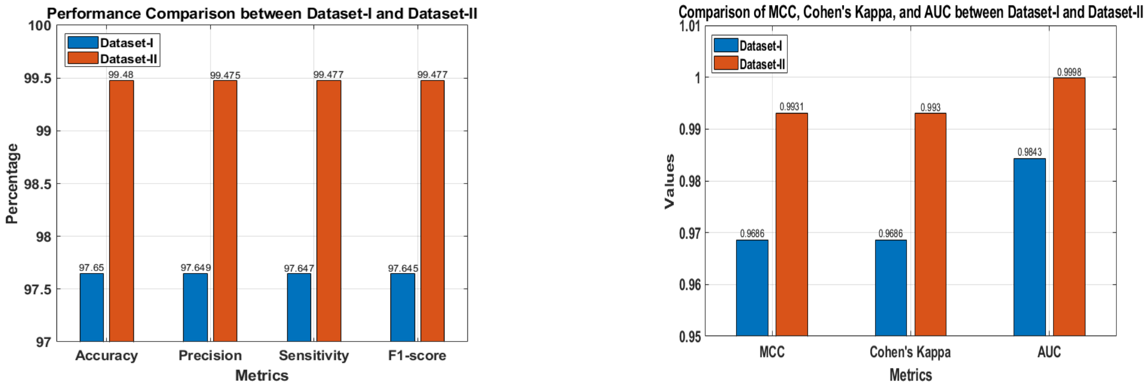
The scattering features from train and test data are combined, and their means are analyzed using box plots. In a box plot, the central mark indicates the median, while the box edges represent the 25th and 75th percentiles. Figures 6.22a and 6.22b show box plots for the four emotion classes, where distinct medians and quartile ranges confirm that the features across classes are completely separable.



(a) Confusion Matrix of Dataset-II.

(b) ROC of four classes of Dataset-II.

Figure 6.25: Confusion matrix and ROC curve for Dataset II.



(a) Accuracy, Precision, Recall, and F1-Score

(b) MCC, AUC, and Cohen's Kappa.

Figure 6.26: Summary of performance parameters for Dataset I and Dataset II.

Figures 6.23a and 6.23b present histograms of retrieved feature sets for Dataset I (happiness, fear, relaxation, sadness) and Dataset II (Game 1–boring, Game 2–calm, Game 3–horror, Game 4–funny), respectively. With normalization applied to show probability density, the histograms reveal distinct feature distributions, confirming the strong discriminatory ability of scattering feature sets for accurate emotion differentiation.

6.4.2.6 Classification of Scattering Features using Multiple Classifiers

The combined scattering features from Datasets I and II are split into training (70%) and testing (30%) sets (subject-independent), with random sequencing to prevent overfitting. The training set is input to multiple classifiers for emotion recognition, including SVM, kNN, quadratic discriminant (QDA), ensemble subspace kNN, and wide neural networks (WNN). Accuracy is evaluated using 10-fold cross-validation. The study explores kNN variants (cubic, fine, weighted, medium), SVM variants (cubic, fine Gaussian), and ensemble methods (subspace kNN, bagged trees) to optimize classification performance. After hyperparameter tuning of kNN using Bayesian optimization, the optimized kNN (O-kNN) achieved the lowest classification error.

6.4.2.7 Comparative Analysis of Classification Results

A 10-fold cross-validation evaluates each classification model by dividing the dataset into ten subsets, iteratively using one for testing and nine for training. This ensures all subsets serve as both

Table 6.16: Performance comparison of emotion recognition methods on two EEG datasets.

Authors	**Method**	**Classes**	**Acc. (%)**
Dataset I: Game-induced Emotions (GAMEEMO)			
Bajaj et al.[32] (2018)	FAWT and Weighted kNN	4	86.10
Krishna et al.[33] (2019)	TQWT and ELM	4	87.10
Taran and Bajaj et al.[34] (2019)	EMD-VMD and MC LS-SVM	4	90.63
Khare and Bajaj et al.[135] (2021a)	SPWVD and CNN	4	93.01
Khare et al. [57] (2020)	ATQWT and MC-LS-SVM	4	95.70
Khare and Bajaj et al.[136] (2021b)	OVMD and ELM	4	97.24
Proposed (Dataset I)	**VMD, O-WST, and O-kNN**	**4**	**97.60**
Dataset II: Video-induced Emotions (LUMED-2)			
Alakus et al.[35] (2020)	Hjorth, Entropy, SD, VAR, ZC + MLPNN	4	73.21
Alakus et al.[35] (2020)	Hjorth, Entropy, SD, VAR, ZC + MLPNN	2	81.35
Alakus and Turkoglu et al.[125](2020)	Spectral Entropy and Bi-LSTM	2	76.93
Aslan et al.[126] (2022)	Scalogram, GoogleNet, and ELM	2	98.41
Tuncer et al.[127] (2022)	Fused features (LEDPatNet19 + Cubic SVM)	2	99.29
Su et al.[128] (2023)	GA-PDPL	4	49.01
Su et al. [128] (2023)	GA-PDPL	2	64.34
M. Aslan et al.[129] (2024)	Energy, R.E., S.D., VAR, SK + CNN	4	95.30
Proposed (Dataset II)	**VMD, O-WST, and O-kNN**	**4**	**99.40**

training and test sets. Model performance is assessed using accuracy, precision, recall, f1-score, MCC, AUC, and Cohen's kappa, with precision, recall, f1-score, MCC, and kappa verifying robustness, stability, and reliability. All classifiers except QDA and WNN exhibit very good κ, while QDA and WNN show moderate κ. Both Dataset-I and Dataset-II, optimized kNN, fine kNN, ensemble subspace kNN, and fine Gaussian SVM achieve very good precision; weighted kNN and cubic SVM achieve good precision; while medium kNN, cubic kNN, ensemble bagged tree, QDA, and WNN demonstrate moderate precision. Optimized kNN, fine kNN, ensemble subspace kNN, fine Gaussian SVM, weighted kNN, and cubic SVM achieve high F1-scores and MCC values above 0.5, indicating strong precision-recall balance and significant correlation between predicted and observed classes. Medium kNN, cubic kNN, ensemble bagged tree, QDA, and WNN yield moderate F1-scores and MCC values (0.25–0.5). All models achieve AUC > 0.5, confirming discrimination ability, with optimized kNN reaching > 0.97, indicating near-perfect class separation. Overall, optimized kNN outperforms all models, with weighted and fine kNN superior to medium and cubic kNN; ensemble subspace kNN better than ensemble bagged tree; fine Gaussian SVM outperforming cubic SVM; WNN surpassing other neural networks; and QDA performing better than linear discriminant.

6.4.2.8 Confusion Matrix and Receiver Operating Characteristic Curves

Figures 6.24a and 6.25a show that the optimized kNN confusion matrices align predicted and true classes closely, achieving accuracies of 97.6% (Dataset I) and 99.4% (Dataset II). The ROC curves in Figures 6.24b and 6.25b demonstrate strong class discrimination, with all classes achieving AUC > 0.98, confirming excellent overall classification performance. Figures 6.26a and 6.26b show that the best classifier achieves accuracy, precision, recall, and F1-score above 97.6% for Dataset-I and 99.4% for Dataset-II, while MCC, AUC, and Cohen's kappa exceed 0.95 and 0.96, respectively. These high values confirm the robustness and stability of the proposed approach.

6.4.3 Comparison of Previous Study

(a) Comparison of Previous Study with Dataset-I

Table 6.16 summarizes studies on Dataset I. Bajaj et al.[32] achieved 86.10% accuracy using weighted kNN with FAWT. Krishna et al.[33] reported 87.1% accuracy with TQWT and ELM. Taran and Bajaj et al.[34] obtained 90.63% accuracy using MC-LS SVM with VMD and EMD. Khare and Bajaj[135] achieved 93.01% with CNN and SPWVD, while Khare et al.[57] reported 95.70% using MCLS-SVM with adaptive TQWT. Khare and Bajaj et al.[136] further improved results with EVCM and O-VMD using ELM kernels, achieving 96.33%, 96.43%, and 97.24% accuracy.

(b) Comparison of Previous Study with Dataset-II.

Table 6.16 summarizes studies on Dataset II (GAMEEMO dataset) [35]. Alakus et al.[35] (2020) used MLPNN with Hjorth, entropy, SD, VAR, and ZC features, achieving 73.21% accuracy, while kNN and SVM yielded 42.85% and 55%. For two-class classification, MLPNN, SVM, and kNN achieved 81.3%, 72.5%, and 66%, respectively. Alakus and Turkoglu[125] proposed a BiLSTM using spectral entropy, reaching 76.91%. Aslan[126] (2022) employed CWT with GoogLeNet, where kNN, SVM, and ELM achieved 98.53%, 98.41%, and 98.78% for two-class classification. Tuncer et al.[127] (2022) integrated TQWT, statistical and nonlinear textural features with LEDPatNet 19, achieving 99.29% for channel-wise two-class classification. Su et al.[128] (2023) introduced PDPL with GA-based optimization, reporting 64.34% for two classes and 49.01% for four classes. Aslan et al.[157] used DWT with statistical features and CNN, achieving 95% accuracy.

The proposed method effectively denoises EEG signals, from which wavelet scattering transform (WST) coefficients are extracted. WST ensures stable, translation-invariant time–frequency resolution while preserving high-frequency details essential for classification. Using optimized kNN, the method achieved classification accuracies of 97.7% and 98.80% on Datasets I and II, respectively. WST produces translation-invariant and deformation-stable representations, which reduce sensitivity to signal variability and noise, thereby enhancing generalization. Furthermore, the consistently high accuracies achieved across different datasets (97.6% and 99.4%), combined with cross-validation and randomized data sequencing during training, suggest that the performance is due to effective generalization rather than overfitting to specific datasets.

6.5 Evolution of Adaptive Wavelets

This chapter explores adaptive wavelet optimization as a method for improving EEG emotion identification in non-stationary and noise-prone settings. Rather than presenting isolated wavelet variants, the proposed research establishes a progressive methodological framework in which successive decomposition strategies overcome the limitations of previous approaches. Initially, an adaptive flexible analytic wavelet transform optimized with PSO was used to reduce reliance on empirically obtained decomposition parameters by adaptively modifying sub-band generation based on reconstruction quality. Despite increased decomposition flexibility, oscillatory rhythm characterization remained limited. As a result, a tunable Q wavelet transform combined with COA was introduced to better match the resonance features of EEG rhythms through adaptive filter tuning. Following on these advancements, the optimized wavelet scattering transform (O-WST) was subsequently developed to provide deformation-stable and translation-invariant signal representations. The methodological progression shows that each framework was established to address a specific decomposition difficulty, rather than to represent marginal algorithmic variation.

6.6 Summary

This chapter centres on improving emotion recognition from EEG signals by employing adaptive signal decomposition techniques. Recent advances demonstrate the effectiveness of adaptive wavelet frameworks combined with nature-inspired optimization algorithms. The adaptive flexible analytic wavelet transform employs particle swarm optimization to automatically tune decomposition parameters (s_1, s_2, s_3, s_4, β), minimizing reconstruction error, generating meaningful sub-bands, and improving time–frequency localization. Statistical features such as Hjorth mobility and complexity, Renyi and Shannon entropy, wavelet energy, and autoregressive coefficients are extracted and classified using optimized machine learning models, achieving accuracies up to 90% while also identifying channels most relevant to emotion detection.

Similarly, the tunable Q wavelet transform optimized with the crayfish optimization algorithm (TQWT–COA) adaptively selects filter parameters, balancing exploration and exploitation, while features from the resulting sub-bands are classified using fine-tuned machine learning and neural models. This approach achieved notable accuracies on datasets like SEED IV and DEAP, outperforming conventional deep learning baselines. Furthermore, wavelet scattering transforms optimized by PSO (O-WST) integrate VMD-based signal reconstruction with adaptive quality factor selection, allowing the optimized k-NN classifier to achieve exceptionally high accuracies of 97.6% and 99.4% across multiple datasets.

These studies collectively demonstrate that optimization-driven preprocessing and adaptive decomposition reduce noise and artifacts, improve feature discriminability, and enhance classifier performance. Building upon these optimized features, the subsequent research objective focuses on deep learning hyperparameter optimization, wherein systematic tuning of parameters such as learning rate, batch size, number of layers, and network architecture further improves model accuracy, generalization, and robustness.

Chapter 7

Emotion Recognition Using Optimally Tuned Deep Learning Models

7.1 Introduction

Emotions are physiological phenomena influencing motivation, personality, temperament, behavior, and feelings. Research increasingly explores their relationship with physiological functions. Emotions strongly affect daily life, including work, play, and relationships, and are broadly categorized as positive or negative. Positive emotions support good mental health [158], while negative emotions can harm both mental and physical health and may lead to psychiatric issues.

EEG signals are widely used to identify emotional states and enhance human–machine interaction through brain–computer interfaces (BCIs) [159]–[29]. Extracting useful EEG features is crucial for emotion recognition. Although EEG often involves over 30 electrodes, spatial characteristics reflecting functional interactions among channels are rarely analyzed. Recent studies using multichannel EEG highlight these spatial aspects [160]. Emotion recognition, influenced by both external and psychological factors, is a rapidly growing interdisciplinary field bridging computer engineering and neuroscience [161].

A study integrating machine learning with EEG signals revealed a strong link between emotional stimuli and EEG responses across different models [162]. In recent years, deep learning (DL) has advanced rapidly, gaining wide research interest. Architectures such as recurrent neural networks (RNNs) [163], convolutional neural networks (CNNs) [58], and deep belief networks (DBNs) [164] have been applied to emotion recognition. These models also demonstrate high performance and efficiency in domains like speech recognition [165], image classification [166], and time-series prediction [167], [168].

A 2D-CNN with four convolutional and seven fully connected layers achieved accuracies of 76.56% for arousal and 80.46% for valence [59]. In [169], a CNN with a residual block (three convolutional layers, the third serving as a shortcut), max pooling, fully connected, and softmax layers reached a mean accuracy of 82.84% on the DEAP dataset. Another CNN-based method [170] generated a 3D feature matrix, applied three 1×1 univariate convolutional layers, followed by a four-way multivariate convolutional layer and a fully connected layer, achieving 85.53% (valence) and 85.88% (arousal) on DEAP.

In [60], a regional-asymmetric CNN (RACNN) with three feature extractors achieved 95% accuracy on the DEAP and DREAMER datasets. While deep learning approaches perform well, their practical use is limited by time and expertise demands, highlighting the need for automated methods. This work proposes EEG-based emotion recognition using CNNs and time–frequency images, where time-

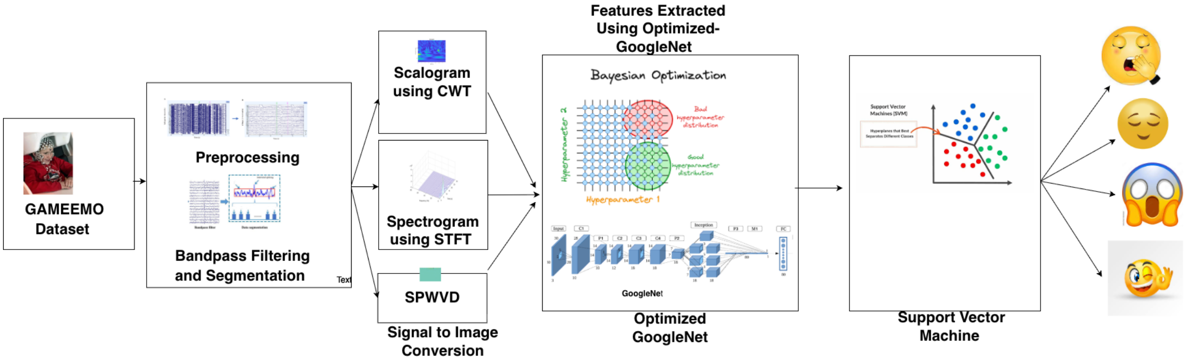


Figure 7.1: Block diagram of the proposed work: optimally tuned deep learning model for EEG emotion recognition.

domain signals are transformed into scalograms, spectrograms, and smoothed pseudo-Wigner–Ville distributions. These representations are input to CNN and optimized CNN models, with the optimized version achieving superior accuracy.

7.2 Methodology

The block diagram of the proposed methodology is displayed in Figure 7.1. EEG signals produced from playing various video games.

7.2.1 Continuous Wavelet Transform

Time-Frequency Representation (TFR) is a useful technique for displaying the spectral variations of a signal over time. This article's methodology in Figure 7.1 makes use of the Short-Time Fourier Transform (STFT). The Continuous Wavelet Transform (CWT), and the SPWVD. It is simpler to convert EEG signals into their time-frequency representations when these transformation methods are coupled. The unique insights that each technique provides into the behavior of the signal allow for a more thorough and sophisticated study of the dynamic frequency components of the signal over time. A band-pass filter with finite energy and a zero mean function that oscillates within a particular frequency range is called a wavelet. There are two ways to accomplish the wavelet transform: the CWT and the discrete wavelet transform (DWT). The signal x(t)'s CWT is represented by [171]

$$W_{c,d}[x(t)] = \frac{1}{\sqrt{c}} \int_{-\infty}^{\infty} x(t)\, \Psi^{*}\left(\frac{t-d}{c}\right) dt \tag{7.1}$$

Where the scaling parameter c regulates the function's spread and d represents the time shift. The scaled and shifted form of the mother wavelet $\psi(t)$ is $\psi(t)$ is $\psi\left(\frac{t-d}{c}\right)$. CWT has the benefit of having adjustable window sizes. Thus, low-frequency components can be analyzed with bigger windows, and high-frequency components can be examined with smaller windows. The scalogram of EEG data during the boring game from a single participant.

7.2.2 Smoothed Pseudo Wigner Ville Distribution

SPWVD [172] is a robust transformation method that gives complete details about the time-frequency localization of a signal's energy. SPWVD gives information on the time-domain localization of the signal and makes it easier to examine higher harmonics.

Cross-term interference is efficiently minimized in SPWVD by utilizing both time and frequency windows. The formula for SPWVD is defined as [172]

$$SPWVD(t,f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m\left(\frac{\tau}{2}\right) m^{*}\left(-\frac{\tau}{2}\right) x(t' + \frac{\tau}{2}) \tag{7.2}$$

$$\times x^*\left(t' - \frac{\tau}{2}\right)\exp(-j2\pi f t')\,dt'\,d\tau \tag{7.3}$$

SPWVD utilizes windows in the time $(m(t))$ and frequency $(n(t))$ domains to minimize interference. These windows are one by one adjusted for precise control over time and frequency details. The optimum between interference and time-frequency precision in conventional techniques is overcome by SPWVD, in contrast to STFT and CWT.

7.2.3 Short-Time Fourier Transform

The Fourier transform in its enhanced form is known as the Short-Time Fourier transform (STFT). It evaluates the frequency variations of a signal over time. The time-domain signal is split into discrete blocks, and each block is then subjected to the Fourier transform. STFT has found several uses with EEG signals because of its ability to divide a signal into blocks [173], [174], and [175]. Several features are extracted from each time-domain signal segment by applying the Fourier transform using a window function. In symmetric bandpass filtering, the frequency bands are evenly spaced in STFT. The signal x(t)'s STFT is calculated as [176]

$$X_\omega(t, f) = \int_0^T x(\tau)\,\omega(t-\tau)\,\exp(-j2\pi f\tau)\,d\tau \tag{7.4}$$

The squared magnitude of the STFT defines the spectrogram. The spectrogram can be expressed mathematically as [176]

$$S_\omega(t, f) = |X_\omega(t, f)|^2 \tag{7.5}$$

where the window function is denoted by . There must be uniformity in the length of the windows in every block. During the window's duration, a signal x(t) is taken to be almost stationary.

Table 7.1: Validation accuracy of baseline and optimized deep learning model across EEG channels AF3, AF4, F3, F4, F7, F8, and FC5.

Deep-Learning Model	**Time-Frequency Technique**	**AF3**	**AF4**	**F3**	**F4**	**F7**	**F8**	**FC5**
Google-Net	Scalogram	60	55.4	63	64	48	60	65.5
	Spectrogram	42.7	38.4	56	40	37	37	38
	SPWVD	70	64	74	78	60	70	76
Optimized Google-Net	Scalogram	71	64	74	73	59	71	76
	Spectrogram	53	47	67	52	48	48	50
	SPWVD	81	85	87	92	87	81	95

Table 7.2: Validation accuracy of baseline and optimized deep learning model across EEG channels FC6, T7, T8, P7, P8, O1, and O2.

Deep-Learning Model	**Time-Frequency Technique**	**FC6**	**T7**	**T8**	**P7**	**P8**	**O1**	**O2**
Google-Net	Scalogram	63	59	63	61	57	66	62
	Spectrogram	40	37	47	39	38	38	39
	SPWVD	74	71	74	73	68	75	71
Optimized Google-Net	Scalogram	76	71	75	72	68	74	73
	Spectrogram	51	48	56	48	49	49	51
	SPWVD	83	81	88	82	78	87	83

Table 7.3: Training parameters used for GoogLeNet.

Parameters	**Value**
Minimization of Loss Function algorithm	‘sgdm’
Max-Epochs	20
Minimum Batch Size	15
Initial-Learn-Rate	1×10^{-4}
Validation Data	imgsValidation
Validation Frequency	10
Momentum	0.9
L2 Regularization	0.0001

7.2.4 Bayesian Optimization

Bayesian optimization integrates deep learning with probabilistic, model-based optimization to identify optimal hyperparameters, which must be set before training (e.g., batch size, learning rate, dropout rate). Unlike grid search, random search, evolutionary, or gradient-based methods, Bayesian optimization is more efficient and effective. It iteratively explores and exploits the hyperparameter space by building a probabilistic surrogate model of the objective function, guiding the search toward promising regions. This approach requires fewer evaluations, adapts sampling based on prior results, and handles noisy or stochastic objective functions through uncertainty modeling, enabling better decision-making.

7.3 Results and Discussions

This work uses the GAMEEMO dataset, comprising 14-channel preprocessed EEG signals with 38,252 samples from 28 subjects playing four games. Each EEG signal is divided into nine segments of 3,840 samples. For channel-wise analysis, data from all 28 subjects in one epoch of the same channel are combined, giving dimensions of (4×28)×3840. Combining all nine epochs yields a total signal dimension of (4×28×9)×3840 = 1008×3840 per channel. A sequential transformation is then applied to convert EEG signals into EEG images, with all 14 channels analyzed individually.

Time–frequency analysis is performed using three techniques: continuous wavelet transform (CWT), smoothed pseudo-Wigner–Ville distribution (SPWVD), and short-time Fourier transform (STFT). In SPWVD, a Kaiser window is applied to reduce cross-terms in time and frequency. Window size selection is critical: large windows increase image size, while very small ones lower resolution. Hence, a medium-sized window of 31 is empirically chosen, constrained to 2^{n-1} for computational efficiency.

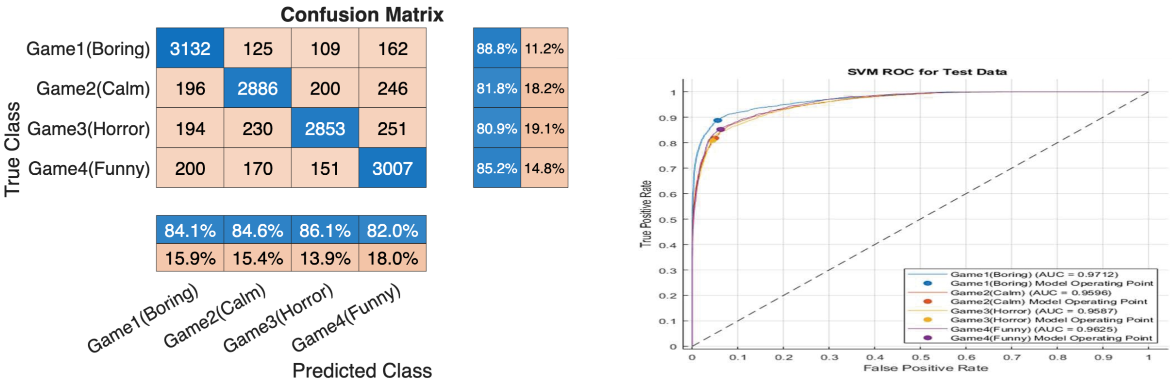
The short-time Fourier transform (STFT) analyzes the temporal evolution of frequency components in nonstationary signals. Its squared magnitude, the spectrogram, provides time–frequency characteristics. A Hann window is applied to 128-sample segments with 64 overlaps, and spectrogram images are generated via water plots. These are scaled into greyscale images of 1008×3840, then mapped with a jet128 colormap to create 1024×1024×3 EEG colour images. The images are resized to 224×224×3 for compatibility with GoogLeNet. Similarly, scalograms, spectrograms, and SPWVD images are imported into an image datastore, labeled by folder names, and split into 70% training, 10% validation, and 10% testing sets (subject-independent).

In deep learning, layers progressively extract features from simple (edges, blobs, colors) to complex representations. Pre-trained networks such as GoogLeNet are effective for EEG image classification. Since GoogLeNet was originally trained for 1000 object categories, its final layers (loss3-classifier and

Table 7.4: Range of hyperparameter values employed in Bayesian optimization.

Hyperparameters	**Range**	**Data Type**	**Transform**
Section Depth	[1 5]	Integer	None
Initial Learning Rate	[1e-4 1]	Real	log
Momentum	[0.5 0.95]	Real	None
L2 Regularization	[1e-12]	Real	log
Loss Function algorithm	[‘Adam’, ‘sgdm’]	Categorical	None
Optimized values of hyperparameters			
Section Depth	3	Integer	None
Initial Learning Rate	0.01327	Real	log
Momentum	0.65334	Real	None

output) are replaced with customized layers for EEG-based emotion recognition, enabling classification using extracted features.



(a) Confusion matrix for test data.

(b) ROC curve for test data.

Figure 7.2: Performance metrics of the model on test data.

Tables 7.1 and 7.2 present validation accuracy for four emotions, showing that optimized GoogLeNet outperforms the standard version, with SPWVD images yielding higher accuracy than scalogram and STFT. Channels AF3, F3, F4, F8, O1, T8, P7, FC5, and O2 perform better than others. Table 7.3 lists the empirical hyperparameter values used to train GoogLeNet with time–frequency images.

GoogLeNet performance is further improved using Bayesian optimization to tune hyperparameters such as section depth, initial learning rate, momentum, regularization, and loss function. Section depth balances model complexity and efficiency. The initial learning rate influences learning speed and stability, while momentum affects convergence smoothness and the impact of past gradients. L_2 regularization prevents overfitting by penalizing large weights, balancing training fit and generalization. The loss function measures prediction errors and influences model behavior during training. Table 7.4 shows the optimized ranges. The optimized GoogLeNet, trained on combined scalogram, spectrogram, and SPWVD images from all channels, extracts features which are then classified using SVMs.

The trained SVM classifier was tested on unseen image data, with performance shown in Figures 7.2a and 7.2b. Figure 7.2a indicates true and predicted class accuracies above 82%, confirming model reliability. Figure 7.2b shows AUC values of 0.9712, 0.9596, 0.9597, and 0.9625 for the four classes, demonstrating clear separability.

Table 7.5: Comparative study with existing EEG-based emotion recognition methods.

Author (Year)	**Feature Representation**	**Classifier**	**Classes**	**Performance (%)**
Alakus et al. (2020) [35]	Hjorth parameters, entropy, standard deviation, variance, zero crossing	kNN, SVM, MLP-NN	4	kNN: 42.85, SVM: 55.00, MLP: 73.21
Alakus et al. (2020) [35]	Hjorth parameters, entropy, standard deviation, variance, zero crossing	kNN, SVM, MLP-NN	2	kNN: 66.00, SVM: 72.50, MLP: 81.30
Alakus and Turkoglu (2020) [125]	Spectral entropy	Bi-LSTM	2	76.93
Aslam et al. (2021) [126]	EEG scalogram features extracted using pretrained GoogleNet	kNN, SVM, ELM	2	kNN: 98.53, SVM: 98.87, ELM: 98.41
Tuncer et al. (2021) [127]	LEDPatNet19 fused features	Cubic SVM	2	99.00
Su et al. (2023) [128]	GA-PDPL features	GA-PDPL	4	49.01, 64.34
Proposed Method	**Optimized GoogleNet features**	**SVM**	**4**	**84.20**

7.4 Comparative Study with Previous Work

Table 7.5 compares results with prior studies. Alakus et al. [35] achieved four-class accuracies of 73.21%, 42.85%, and 55% using MLPNN, while kNN and SVM performed worse; in 2-class classification, MLPNN, SVM, and kNN achieved 81.3%, 72.5%, and 66%, respectively. Alakus and Turkoglu [125] used spectral entropy with BiLSTM, obtaining 76.91%. Aslam et al.[126] reported 2-class accuracies of 98.78%, 98.53%, and 98.41% using SVM, kNN, and ELM. Tuncer et al. [127] reached 99.29% channel-wise accuracy with a cubic SVM, while Su et al. [128] achieved 64.34% (2-class) and 49.01% (4-class) using projection dictionary learning.

This study evaluated CNN-based emotion recognition during gaming conditions using three time-frequency representations: STFT, CWT, and SPWVD. SPWVD was superior to the other two due to better joint time-frequency localization and less cross-term interference. Rather than proposing a new CNN architecture, the contribution is to optimize and adapt CNNs for EEG-specific representations, systematically aligning representation quality with network optimisation to improve classification accuracy, convergence, and computational efficiency in real-time applications. According to the interpretability analysis, the higher performance is due to better capturing of emotion-related spectral-temporal patterns in EEG rhythms.

7.5 Summary

This study investigated the effectiveness of CNN-based architectures for EEG-based emotion recognition under gaming conditions, using multiple time–frequency representation techniques. These techniques, including Short-Time Fourier Transform, Continuous Wavelet Transform, and Smoothed Pseudo Wigner–Ville Distribution (SPWVD), are systematically compared. Among these, SPWVD demonstrated superior performance, achieving higher classification accuracy due to its ability to capture richer joint time–frequency features and suppress cross-term interference more effectively than STFT and CWT.

The proposed optimized CNN framework, customized to process these time–frequency representations, consistently outperformed baseline variants. In addition to achieving the highest classification accuracy, it also demonstrated faster convergence and enhanced computational efficiency, underscoring its suitability for real-time emotion recognition systems.

This demonstrates the importance of aligning deep learning models with robust time–frequency analysis techniques for improved classification reliability. Building on this foundation, future work will extend the framework by integrating EEG signals with multimodal data streams from additional wearable sensors (e.g., physiological and behavioral signals), with the goal of enhancing cross-domain learning and achieving higher accuracy in emotion recognition.

Chapter 8

Conclusions, Future Directions and Social Impact

8.1 Conclusions

Emotion recognition has substantial importance in affective computing, healthcare, and human computer interaction. Emotional recognition using EEG signals remains challenging due to signal non-stationarity, inter-subject variability, and sensitivity to noise and errors. This thesis investigated multiple methods for enhancing EEG-based emotion recognition, including channel selection optimization, adaptive preprocessing, wavelet-based decomposition, and deep learning-based classification.

An optimal subset of EEG electrodes was investigated utilizing correlation-based channel selection to reduce redundancy while retaining discriminative emotional information. Statistical and entropy-based characteristics extracted from rhythm-decomposed EEG signals displayed the potential of lowering channel complexity while maintaining recognition performance.

To address signal quality limitations, adaptive filtering, decomposition-based denoising, and combined time-frequency representations were examined for increased reliability against physiological and environmental artifacts. These methods contributed to improve feature stability and more reliable signal representations in noisy environments.

Furthermore, adaptive wavelet parameter optimization with nature-inspired algorithms was investigated to reduce dependence on empirically selected decomposition parameters. Optimization-driven tuning in frameworks such as AFAWT, TQWT, and optimized WST resulted in better time-frequency localization and feature discriminability on the tested datasets.

Finally, CNN-based architectures combining time-frequency representations were investigated for EEG emotion recognition in gaming situations. A comparison of STFT, CWT, and SPWVD representations revealed that SPWVD-based representations provided higher temporal-spectral information in the tested experimental setup. Deep-learning hyperparameters tuning enhances both convergence behavior and classification performance.

Overall, this thesis conducts a systematic investigation of multiple EEG emotion recognition strategies and shows that combining channel optimization, adaptive preprocessing, optimized decomposition, and deep learning-based classification improves recognition performance in the experimental settings studied.

8.2 Future Directions

Although this thesis demonstrates promising advancements in EEG-based emotion recognition, several research challenges remain open and provide opportunities for further improvement.

(a) Explainable and Interpretable Emotion Recognition: Deep learning models can be difficult to interpret in healthcare and affective computing applications due to their black-box nature. Future studies can use XAI approaches like Grad-CAM, SHAP, and attention-based visualization to find discriminatory EEG channels, rhythms, and temporal patterns that contribute to emotional classification. This can improve reliability, interpretability, and clinical acceptance.

(b) Adaptive and Self-Configuring Signal Decomposition: Although improved wavelet and decomposition frameworks were investigated in this thesis, future research should concentrate on self-adaptive decomposition strategies capable of dynamically picking the best transform and parameters based on EEG signal properties. Reinforcement learning and metaheuristic-driven adaptive frameworks may enhance resilience under a variety of emotional and recording situations.

(c) Lightweight Models for Real-Time Deployment: The computational complexity of optimal decomposition and deep learning frameworks continues to pose a hurdle for wearable and real-time applications. Future study may look into lightweight neural architectures, model compression, pruning, and knowledge distillation approaches to enable efficient deployment on portable EEG devices and edge computing platforms.

(d) Multimodal Emotion Recognition Frameworks: Future research could broaden the suggested EEG-based framework by incorporating complementing physiological and behavioral data such as electrocardiography (ECG), galvanic skin response (GSR), respiration, facial expressions, and eye-tracking. Transformer-based and attention-driven fusion frameworks may increase robustness, especially in noisy or emotionally complex environments.

(e) Integration of Adaptive Brain-Computer Interface Systems: The suggested framework could be developed to include adaptive brain-computer interface (BCI) systems for mental health monitoring, neurofeedback, stress assessment, and assistive technologies. Future research may look into closed-loop emotional state adaptation to enable intelligent and individualized human-computer interaction.

8.3 Social Impact

This research on EEG-based emotion identification transcends traditional technological advancements to address fundamental human rights: the right to emotional expression, the right to mental healthcare, and the right to participate fully in an increasingly digital society. By grounding emotion recognition in the universal language of brain activity, we create technologies that are fundamentally equal, scalable, and transformative. This technology represents not only scientific progress but also a commitment to fostering cultures that recognize, support, and enhance emotional well-being through inclusive technological frameworks.

References

[1] James A Russell. "A circumplex model of affect." In: *Journal of personality and social psychology* 39.6 (1980), p. 1161.

[2] Md Rabiul Islam et al. "Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques". In: *IEEe Access* 9 (2021), pp. 94601–94624.

[3] Albert Mehrabian. "Correlations of the PAD Emotional Scales with self-reported satisfaction in marriage and work". In: *Genetic, Social, and General Psychology Monographs* 124.3 (1998), p. 311.

[4] Paul Ekman, Tim Dalgleish, and M Power. "Basic emotions". In: *San Francisco, USA* (1999).

[5] William James. "The James-Lange theory of emotion". In: *Visceral sensory neuroscience: interoception* 9 (2002).

[6] Magda B Arnold. *The emotions: facts, theories and a new model*. 1964.

[7] Otniel E Dror. "Deconstructing the "two factors": The historical origins of the Schachter–Singer theory of emotions". In: *Emotion Review* 9.1 (2017), pp. 7–16.

[8] Hany Ferdinando et al. "Emotion recognition by heart rate variability". In: *Australian Journal of Basic and Applied Science* 8.14 (2014), pp. 50–55.

[9] Trisha Paul et al. "Human emotion recognition using GSR and EEG". In: *International Journal of Scientific and Research Publication* 10.5 (2020), pp. 394–400.

[10] Margaret M Bradley and Peter J Lang. "Measuring emotion: the self-assessment manikin and the semantic differential". In: *Journal of behavior therapy and experimental psychiatry* 25.1 (1994), pp. 49–59.

[11] Burkhard Matzke et al. "Facial reactions during emotion recognition in borderline personality disorder: a facial electromyography study". In: *Psychopathology* 47.2 (2014), pp. 101–110.

[12] Nima TaheriNejad, Axel Jantsch, and David Pollreisz. "Comprehensive Observation and its Role in Self-Awareness; An Emotion Recognition System Example." In: *Fed-CSIS (Position Papers)*. 2016, pp. 117–124.

[13] Spiros V Ioannou et al. "Emotion recognition through facial expression analysis based on a neurofuzzy network". In: *Neural Networks* 18.4 (2005), pp. 423–435.

[14] CUI Mingming, FANG Jiandong, and ZHAO Yudong. “Emotion recognition of human body’s posture in open environment”. In: *2020 Chinese Control And Decision Conference (CCDC)*. Ieee. 2020, pp. 3294–3299.

[15] Jia Zheng Lim, James Mountstephens, and Jason Teo. “Emotion recognition using eye-tracking: taxonomy, review and current challenges”. In: *Sensors* 20.8 (2020), p. 2384.

[16] Ralf Kredel et al. “Eye-tracking technology and the dynamics of natural gaze behavior in sports: A systematic review of 40 years of research”. In: *Frontiers in psychology* 8 (2017), p. 287392.

[17] K Sreenivasa Rao and Shashidhar G Koolagudi. *Emotion recognition using speech features*. Springer Science & Business Media, 2012.

[18] Amol Patwardhan. “Three-dimensional, kinematic, human Behavioral pattern-based features for multimodal emotion recognition”. In: *Multimodal Technologies and Interaction* 1.3 (2017), p. 19.

[19] Judy Willis. “Review of research: Brain-based teaching strategies for improving students’ memory, learning, and test-taking success”. In: *Childhood education* 83.5 (2007), pp. 310–315.

[20] Juan J López-Ibor et al. “The perception of emotion-free faces in schizophrenia: a magneto-encephalography study”. In: *Schizophrenia research* 98.1-3 (2008), pp. 278–286.

[21] Coenraad J Hattingh et al. “Functional magnetic resonance imaging during emotion recognition in social anxiety disorder: an activation likelihood meta-analysis”. In: *Frontiers in human neuroscience* 6 (2013), p. 347.

[22] Piotr Olejniczak. “Neurophysiologic basis of EEG”. In: *Journal of clinical neurophysiology* 23.3 (2006), pp. 186–189.

[23] Mouhannad Ali et al. “EEG-based emotion recognition approach for e-healthcare applications”. In: *2016 eighth international conference on ubiquitous and future networks (ICUFN)*. IEEE. 2016, pp. 946–950.

[24] Essam H Houssein, Asmaa Hammad, and Abdelmgeid A Ali. “Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review”. In: *Neural Computing and Applications* 34.15 (2022), pp. 12527–12557.

[25] Meng-Hsien Lin et al. “Applying EEG in consumer neuroscience”. In: *European Journal of Marketing* 52.1/2 (2018), pp. 66–91.

[26] Yi Wei. “Assessment study on brain wave predictive ability to policemen’s safety law enforcement”. In: *Journal of Discrete Mathematical Sciences and Cryptography* 20.1 (2017), pp. 193–204.

[27] Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. “EEG-based emotion recognition: a state-of-the-art review of current trends and opportunities”. In: *Computational intelligence and neuroscience* 2020.1 (2020), p. 8875426.

[28] Madiha Tariq, Pavel M Trivailo, and Milan Simic. "EEG-based BCI control schemes for lower-limb assistive-robots". In: *Frontiers in human neuroscience* 12 (2018), p. 312.

[29] Sander Koelstra et al. "Deap: A database for emotion analysis; using physiological signals". In: *IEEE transactions on affective computing* 3.1 (2011), pp. 18–31.

[30] Stamos Katsigiannis and Naeem Ramzan. "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices". In: *IEEE journal of biomedical and health informatics* 22.1 (2017), pp. 98–107.

[31] Wei-Long Zheng et al. "Emotionmeter: A multimodal framework for recognizing human emotions". In: *IEEE transactions on cybernetics* 49.3 (2018), pp. 1110–1122.

[32] Varun Bajaj, Sachin Taran, and Abdulkadir Sengur. "Emotion classification using flexible analytic wavelet transform for electroencephalogram signals". In: *Health information science and systems* 6.1 (2018), p. 12.

[33] Anala Hari Krishna et al. "Emotion classification using EEG signals based on tunable-Q wavelet transform". In: *IET Science, Measurement & Technology* 13.3 (2019), pp. 375–380.

[34] Sachin Taran and Varun Bajaj. "Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method". In: *Computer methods and programs in biomedicine* 173 (2019), pp. 157–165.

[35] Talha Burak Alakus, Murat Gonen, and Ibrahim Turkoglu. "Database for an emotion recognition system based on EEG signals and various computer games–GAMEEMO". In: *Biomedical Signal Processing and Control* 60 (2020), p. 101951.

[36] Yucel Cimtay, Erhan Ekmekcioglu, and Seyma Caglar-Ozhan. "Cross-subject multimodal emotion recognition based on hybrid fusion". In: *IEEE Access* 8 (2020), pp. 168865–168878.

[37] Rosalind W Picard. *Affective computing*. MIT press, 2000.

[38] Maria Egger, Matthias Ley, and Sten Hanke. "Emotion recognition from physiological signal analysis: A review". In: *Electronic Notes in Theoretical Computer Science* 343 (2019), pp. 35–55.

[39] Manuel G Calvo and Lauri Nummenmaa. "Detection of emotional faces: salient physical features guide effective visual search." In: *Journal of Experimental Psychology: General* 137.3 (2008), p. 471.

[40] Pawe l Tarnowski et al. "Emotion recognition using facial expressions". In: *Procedia Computer Science* 108 (2017), pp. 1175–1184.

[41] Morteza Zangeneh Soroush et al. "A review on EEG signals based emotion recognition". In: *International Clinical Neuroscience Journal* 4.4 (2017), pp. 118–129.

[42] Rab Nawaz et al. "Comparison of different feature extraction methods for EEG-based emotion recognition". In: *Biocybernetics and Biomedical Engineering* 40.3 (2020), pp. 910–926.

[43] Starlet Ben Alex, Leena Mary, and Ben P Babu. “Attention and feature selection for automatic speech emotion recognition using utterance and syllable-level prosodic features”. In: *Circuits, Systems, and Signal Processing* 39.11 (2020), pp. 5681–5709.

[44] David Feess, Mario M Krell, and Jan H Metzen. “Comparison of sensor selection mechanisms for an ERP-based brain-computer interface”. In: *PloS one* 8.7 (2013), e67543.

[45] Yuan-Pin Lin et al. “EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine”. In: *2009 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2009, pp. 489–492.

[46] Zhong-Min Wang, Shu-Yuan Hu, and Hui Song. “Channel selection method for EEG emotion recognition using normalized mutual information”. In: *IEEE access* 7 (2019), pp. 143303–143311.

[47] Wei-Long Zheng and Bao-Liang Lu. “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks”. In: *IEEE Transactions on autonomous mental development* 7.3 (2015), pp. 162–175.

[48] Esen Yildirim, Yasn Kaya, and Fatih Kiliç. “A channel selection method for emotion recognition from EEG based on swarm-intelligence algorithms”. In: *IEEE Access* 9 (2021), pp. 109889–109902.

[49] Mahnaz Arvaneh et al. “Optimizing the channel selection and classification accuracy in EEG-based BCI”. In: *IEEE transactions on biomedical engineering* 58.6 (2011), pp. 1865–1873.

[50] Michael Schröder et al. “Robust EEG channel selection across subjects for brain-computer interfaces”. In: *EURASIP Journal on Advances in Signal Processing* 2005.19 (2005), p. 174746.

[51] Murugappan Murugappan, Nagarajan Ramachandran, and Yaacob Sazali. “Classification of human emotion from EEG using discrete wavelet transform”. In: *Journal of biomedical science and engineering* 3.4 (2010), pp. 390–396.

[52] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. “Deep learning for electroencephalogram (EEG) classification tasks: a review”. In: *Journal of neural engineering* 16.3 (2019), p. 031001.

[53] Vipin Gupta, Mayur Dahyabhai Chopda, and Ram Bilas Pachori. “Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals”. In: *IEEE Sensors Journal* 19.6 (2018), pp. 2266–2274.

[54] Ashwin Kamble, Pradnya H Ghare, and Vinay Kumar. “Optimized rational dilation wavelet transform for automatic imagined speech recognition”. In: *IEEE Transactions on Instrumentation and Measurement* 72 (2023), pp. 1–10.

[55] Ahmed Elrefaiy et al. “EEG emotion recognition framework based on invariant wavelet scattering convolution network”. In: *Journal of Ambient Intelligence and Humanized Computing* 15.4 (2024), pp. 2181–2199.

[56] Joakim Andén, Vincent Lostanlen, and Stéphane Mallat. “Joint time–frequency scattering”. In: *IEEE Transactions on Signal Processing* 67.14 (2019), pp. 3704–3718.

[57] Smith K Khare, Varun Bajaj, and Ganesh Ram Sinha. “Adaptive tunable Q wavelet transform-based emotion identification”. In: *IEEE transactions on instrumentation and measurement* 69.12 (2020), pp. 9609–9617.

[58] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[59] Yea-Hoon Kwon, Sae-Byuk Shin, and Shin-Dug Kim. “Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system”. In: *Sensors* 18.5 (2018), p. 1383.

[60] Heng Cui et al. “EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network”. In: *Knowledge-Based Systems* 205 (2020), p. 106243.

[61] Dongrui Gao et al. “EEG driving fatigue detection based on log-Mel spectrogram and convolutional recurrent neural networks”. In: *Frontiers in Neuroscience* 17 (2023), p. 1136609.

[62] Zhe Jia et al. “A novel dual-task model for EEG-based emotion and cognition recognition”. In: *IEEE Transactions on Instrumentation and Measurement* (2024).

[63] Jingzhao Hu et al. “ScalingNet: Extracting features from raw EEG data for emotion recognition”. In: *Neurocomputing* 463 (2021), pp. 177–184.

[64] Zhuoling Cheng et al. “EEG-based emotion recognition using multi-scale dynamic CNN and gated transformer”. In: *Scientific Reports* 14.1 (2024), p. 31319.

[65] Nirmal Varghese Babu and E Grace Mary Kanaga. “A novel framework for multiclass emotion detection using EEG signals with advanced preprocessing and explainable graph attention networks”. In: *International Journal of Computers and Applications* (2025), pp. 1–17.

[66] Jibin Yin et al. “EEG-based emotion recognition with autoencoder feature fusion and MSC-TimesNet model”. In: *Computer Methods in Biomechanics and Biomedical Engineering* (2025), pp. 1–18.

[67] Monira Islam and Tan Lee. “An automated extraction of spectral-temporal and spatial-temporal features of EEG for emotion detection”. In: *Brain Informatics* 12.1 (2025), pp. 1–19.

[68] Mengmeng Yan et al. “An improved common spatial pattern combined with channel-selection strategy for electroencephalography-based emotion recognition”. In: *Medical Engineering & Physics* 83 (2020), pp. 130–141.

[69] Håvard Vold and Jan Leuridan. *High resolution order tracking at extreme slew rates, using Kalman tracking filters*. Tech. rep. SAE Technical Paper, 1993.

[70] Zhu Yan, Liang Wang, Aijun Hu, et al. “Feature extraction by enhanced time-frequency representation based on Vold-Kalman filter”. In: *Liang and Hu, Aijun, Feature extraction by enhanced time-frequency representation based on Vold-Kalman filter* (2022).

[71] WY Choong et al. “Correlation analysis of emotional EEG in alpha, beta and gamma frequency bands”. In: *Journal of Physics: Conference Series*. Vol. 1997. 1. IOP Publishing. 2021, p. 012029.

[72] Bo Hjorth. “EEG analysis based on time domain properties”. In: *Electroencephalography and clinical neurophysiology* 29.3 (1970), pp. 306–310.

[73] Claude E Shannon. “A mathematical theory of communication”. In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[74] Peixiang Zhong, Di Wang, and Chunyan Miao. “EEG-based emotion recognition using regularized graph neural networks”. In: *IEEE Transactions on Affective Computing* 13.3 (2020), pp. 1290–1301.

[75] Jiyao Liu et al. “Spatial-temporal transformers for EEG emotion recognition”. In: *Proceedings of the 6th International Conference on Advances in Artificial Intelligence*. 2022, pp. 116–120.

[76] Sijin Zhou et al. “Objectivity meets subjectivity: A subjective and objective feature fused neural network for emotion recognition”. In: *Applied Soft Computing* 122 (2022), p. 108889.

[77] Thalía Harmony. “The functional significance of delta oscillations in cognitive processing”. In: *Frontiers in integrative neuroscience* 7 (2013), p. 83.

[78] Ljubomir I Aftanas et al. “Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands”. In: *International Journal of Psychophysiology* 44.1 (2002), pp. 67–82.

[79] GG Knyazev, J Yu Slobodskoj-Plusnin, and AV Bocharov. “Event-related delta and theta synchronization during explicit and implicit emotion processing”. In: *Neuroscience* 164.4 (2009), pp. 1588–1600.

[80] Giacomo Novembre, Daniela Sammler, and Peter E Keller. “Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action”. In: *Neuropsychologia* 89 (2016), pp. 414–425.

[81] James A Coan and John JB Allen. “Frontal EEG asymmetry as a moderator and mediator of emotion”. In: *Biological psychology* 67.1-2 (2004), pp. 7–50.

[82] Richard J Davidson. *Emotion and affective style: Hemispheric substrates*. 1992.

[83] Chia-Yen Yang and Ching-Po Lin. “Magnetoencephalography study of different relationships among low-and high-frequency-band neural activities during the induction of peaceful and fearful audiovisual modalities among males and females”. In: *Journal of Neuroscience Research* 95.1-2 (2017), pp. 176–188.

[84] Bahar Güntekin and Erol Başar. “A review of brain oscillations in perception of faces and emotional pictures”. In: *Neuropsychologia* 58 (2014), pp. 33–51.

[85] Matthias M Müller, Thomas Gruber, and Andreas Keil. “Modulation of induced gamma band activity in the human EEG by attention and visual information processing”. In: *International Journal of Psychophysiology* 38.3 (2000), pp. 283–299.

[86] Andreas Keil et al. “Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials”. In: *Clinical neurophysiology* 112.11 (2001), pp. 2057–2068.

[87] Baloju Revanth et al. “Multi-channel EEG-based multi-class emotion recognition from multiple frequency bands”. In: *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*. IEEE. 2023, pp. 1–5.

[88] Parthana Sarma and Shovan Barma. “Emotion recognition by distinguishing appropriate EEG segments based on random matrix theory”. In: *Biomedical Signal Processing and Control* 70 (2021), p. 102991.

[89] Yisi Liu and Olga Sourina. “EEG-based dominance level recognition for emotion-enabled interaction”. In: *2012 IEEE International Conference on Multimedia and Expo*. IEEE. 2012, pp. 1039–1044.

[90] Guiyoung Son and Yaeri Kim. “EEG-Based Emotion Classification for Verifying the Korean Emotional Movie Clips with Support Vector Machine (SVM)”. In: *Complexity* 2021.1 (2021), p. 5497081.

[91] Sourabh Banik et al. “Assessment of Emotion Elicitation using Multimodal Physiological Sensors and Phase Synchronization”. In: *IEEE Sensors Letters* (2024).

[92] Raveendrababu Vempati and Lakhan Dev Sharma. “EEG rhythm based emotion recognition using multivariate decomposition and ensemble machine learning classifier”. In: *Journal of Neuroscience Methods* 393 (2023), p. 109879.

[93] Mu Li and Bao-Liang Lu. “Emotion classification based on gamma-band EEG”. In: *2009 Annual International Conference of the IEEE Engineering in medicine and biology society*. ieee. 2009, pp. 1223–1226.

[94] Jianzhuo Yan, Shangbin Chen, and Sinuo Deng. “A EEG-based emotion recognition model with rhythm and time characteristics”. In: *Brain informatics* 6.1 (2019), p. 7.

[95] Metin Bilgin and Ahmet Mert. “Gated transformer network based EEG emotion recognition”. In: *Signal, Image and Video Processing* 18.10 (2024), pp. 6903–6910.

[96] Md Mahinur Alam et al. “TMNet: Transformer-fused multimodal framework for emotion recognition via EEG and speech”. In: *ICT Express* (2025).

[97] Xuheng Jiang et al. “WNOTNet: A Hybrid Wavelet Neural Operator and Transformer Framework for Enhanced EEG Denoising”. In: *IEEE Transactions on Instrumentation and Measurement* (2025).

[98] Richard C Burgess. “Filtering of neurophysiologic signals”. In: *Handbook of clinical neurology* 160 (2019), pp. 51–65.

[99] Heming Jia et al. “Crayfish optimization algorithm”. In: *Artificial Intelligence Review* 56.Suppl 2 (2023), pp. 1919–1979.

[100] Wai-Cheong Lincoln Lew et al. “EEG-based emotion recognition using spatial-temporal representation via Bi-GRU”. In: *2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*. IEEE. 2020, pp. 116–119.

[101] Xiaowei Zhang et al. “Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine”. In: *IEEE transactions on cybernetics* 51.9 (2020), pp. 4386–4399.

[102] Yuzhe Zhang et al. “EEG-based emotion recognition with emotion localization via hierarchical self-attention”. In: *IEEE transactions on affective computing* 14.3 (2022), pp. 2458–2469.

[103] Amit Kumar Dwivedi, Om Prakash Verma, and Sachin Taran. “EEG-Based Emotion Recognition Using Optimum Number of Channels”. In: *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE)*. IEEE. 2024, pp. 1–6.

[104] James Kennedy and Russell Eberhart. “Particle swarm optimization”. In: *Proceedings of ICNN’95-international conference on neural networks*. Vol. 4. ieee. 1995, pp. 1942–1948.

[105] Vasile V Moca et al. “Time-frequency super-resolution with superlets”. In: *Nature communications* 12.1 (2021), p. 337.

[106] Sara Bagherzadeh et al. “A hybrid EEG-based emotion recognition approach using wavelet convolutional neural networks and support vector machine”. In: *Basic and Clinical Neuroscience* 14.1 (2023), p. 87.

[107] Sara Bagherzadeh et al. “Automated depression detection via cloud based EEG analysis with transfer learning and synchrosqueezed wavelet transform”. In: *Scientific Reports* 15.1 (2025), p. 18008.

[108] Sara Bagherzadeh et al. “A subject-independent portable emotion recognition system using synchrosqueezing wavelet transform maps of EEG signals and ResNet-18”. In: *Biomedical Signal Processing and Control* 90 (2024), p. 105875.

[109] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet classification with deep convolutional neural networks”. In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[110] Forrest N Iandola et al. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size”. In: *arXiv preprint arXiv:1602.07360* (2016).

[111] Keiron O’shea and Ryan Nash. “An introduction to convolutional neural networks”. In: *arXiv preprint arXiv:1511.08458* (2015).

[112] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[113] Wei Li et al. “TMLP+ SRDANN: A domain adaptation method for EEG-based emotion recognition”. In: *Measurement* 207 (2023), p. 112379.

[114] Gopal Chandra Jana, Anshuman Sabath, and Anupam Agrawal. “Capsule neural networks on spatio-temporal EEG frames for cross-subject emotion recognition”. In: *Biomedical Signal Processing and Control* 72 (2022), p. 103361.

[115] Haipeng Liu et al. “EEG Emotion Recognition via a Lightweight 1DCNN-BiLSTM Model in Resource-Limited Environments”. In: *IEEE Sensors Journal* (2024).

[116] Chang Li et al. “EEG-based emotion recognition via efficient convolutional neural network and contrastive learning”. In: *IEEE Sensors Journal* 22.20 (2022), pp. 19608–19619.

[117] Sheeraz Ahmad Khan, Eamin Chaudary, and Wajid Mumtaz. “EEG-ConvNet: Convolutional networks for EEG-based subject-dependent emotion recognition”. In: *Computers and Electrical Engineering* 116 (2024), p. 109178.

[118] Himanshu Kumar, Nagarajan Ganapathy, and Ramakrishnan Swaminathan. “Analysis of dynamics of EEG signals in emotional valence using super-resolution superlet transform”. In: *IEEE Sensors Letters* (2025).

[119] Fuwang Wang et al. “Real driving environment EEG-based detection of driving fatigue using the wavelet scattering network”. In: *Journal of neuroscience methods* 400 (2023), p. 109983.

[120] Samaneh Madanian et al. “Speech emotion recognition using machine learning—A systematic review”. In: *Intelligent systems with applications* 20 (2023), p. 200266.

[121] Haoyang Liu. “Emotion Detection through Body Gesture and Face”. In: *arXiv preprint arXiv:2407.09913* (2024).

[122] Joakim Andén and Stéphane Mallat. “Deep scattering spectrum”. In: *IEEE Transactions on Signal Processing* 62.16 (2014), pp. 4114–4128.

[123] Peter J Rousseeuw. “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.

[124] Konstantin Dragomiretskiy and Dominique Zosso. “Variational mode decomposition”. In: *IEEE transactions on signal processing* 62.3 (2013), pp. 531–544.

[125] TB Alakus and IJEL Turkoglu. “Emotion recognition with deep learning using GAMEEMO data set”. In: *Electronics Letters* 56.25 (2020), pp. 1364–1367.

[126] Muzaffer Aslan. “CNN based efficient approach for emotion recognition”. In: *Journal of King Saud University-Computer and Information Sciences* 34.9 (2022), pp. 7335–7346.

[127] Turker Tuncer, Sengul Dogan, and Abdulhamit Subasi. “LEDPatNet19: Automated emotion recognition model based on nonlinear LED pattern feature extraction function using EEG signals”. In: *Cognitive Neurodynamics* 16.4 (2022), pp. 779–790.

[128] Jipu Su et al. “Subject-independent eeg emotion recognition based on genetically optimized projection dictionary pair learning”. In: *Brain Sciences* 13.7 (2023), p. 977.

[129] Musa Aslan, Muhammet Baykara, and Talha Burak Alakuş. “Analysis of brain areas in emotion recognition from EEG signals with deep learning methods”. In: *Multimedia Tools and Applications* 83.11 (2024), pp. 32423–32452.

[130] Aftab Alam, Shabana Urooj, and Abdul Quaiyum Ansari. “Human emotion recognition models using machine learning techniques”. In: *2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON)*. IEEE. 2023, pp. 329–334.

[131] Nasser Mourad. “Group-sparse mode decomposition: A signal decomposition algorithm based on group-sparsity in the frequency domain”. In: *Digital Signal Processing* 122 (2022), p. 103375.

[132] Jingyi Lu et al. “Variational mode decomposition denoising combined with improved Bhattacharyya distance”. In: *Measurement* 151 (2020), p. 107283.

[133] Ivan W Selesnick. “Wavelet transform with tunable Q-factor”. In: *IEEE transactions on signal processing* 59.8 (2011), pp. 3560–3575.

[134] Donald B Percival and Harold O Mofjeld. “Analysis of subtidal coastal sea level fluctuations using wavelets”. In: *Journal of the American Statistical Association* 92.439 (1997), pp. 868–880.

[135] Smith K Khare and Varun Bajaj. “Time–frequency representation and convolutional neural network-based emotion recognition”. In: *IEEE transactions on neural networks and learning systems* 32.7 (2020), pp. 2901–2909.

[136] Smith K Khare and Varun Bajaj. “An evolutionary optimized variational mode decomposition for emotion recognition”. In: *IEEE Sensors Journal* 21.2 (2020), pp. 2035–2042.

[137] Smith K Khare, Varun Bajaj, and U Rajendra Acharya. “SPWVD-CNN for automated detection of schizophrenia patients using EEG signals”. In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–9.

[138] Kah Phooi Seng, Li-Minn Ang, and Chien Shing Ooi. “A combined rule-based & machine learning audio-visual emotion recognition approach”. In: *IEEE Transactions on Affective Computing* 9.1 (2016), pp. 3–13.

[139] Oyenuga Wasiu Olakunle. “Discrete Wavelet Packet Transform for Electroencephalogram Based Valence-arousal Emotion Recognition”. PhD thesis. Universiti Utara Malaysia, 2015.

[140] Alejandro Gomez et al. “An approach to emotion recognition in single-channel EEG signals using stationary wavelet transform”. In: *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*. Springer. 2017, pp. 654–657.

[141] Smith K Khare et al. “Classification of mental states from rational dilation wavelet transform and bagged tree classifier using EEG signals”. In: *Artificial Intelligence-Based Brain-Computer Interface*. Elsevier, 2022, pp. 217–235.

[142] Ning Zhuang et al. “Emotion recognition from EEG signals using multidimensional information in EMD domain”. In: *BioMed research international* 2017.1 (2017), p. 8317357.

[143] Ilker Bayram. “An analytic wavelet transform with a flexible time-frequency covering”. In: *IEEE Transactions on Signal Processing* 61.5 (2012), pp. 1131–1142.

[144] Petr Jizba and Toshihico Arimitsu. “The world according to Rényi: thermodynamics of multifractal systems”. In: *Annals of Physics* 312.1 (2004), pp. 17–59.

[145] IS Iwueze et al. “Comparison of Two Time Series Decomposition Methods: Least Squares and Buys-Ballot Methods”. In: *Open Journal of Statistics* 6.6 (2016), pp. 1123–1137.

[146] Stephane G Mallat. “A theory for multiresolution signal decomposition: the wavelet representation”. In: *IEEE transactions on pattern analysis and machine intelligence* 11.7 (2002), pp. 674–693.

[147] John G Proakis. *Digital signal processing: principles, algorithms, and applications, 4/E*. Pearson Education India, 2007.

[148] Gareth James et al. *An introduction to statistical learning: with applications in R*. Vol. 103. Springer, 2013.

[149] Rui Yan et al. “Multi-modality of polysomnography signals’ fusion for automatic sleep scoring”. In: *Biomedical Signal Processing and Control* (2019).

[150] Digambar V Puri, Sanjay L Nalbalwar, and Pallavi P Ingle. “EEG-based systematic explainable Alzheimer’s disease and mild cognitive impairment identification using novel rational dyadic biorthogonal wavelet filter banks”. In: *Circuits, Systems, and Signal Processing* 43.3 (2024), pp. 1792–1822.

[151] Serap Aydın, Hamdi Melih Saraoğlu, and Sadık Kara. “Log energy entropy-based EEG classification with multilayer neural networks in seizure”. In: *Annals of biomedical engineering* 37.12 (2009), pp. 2626–2630.

[152] Magdiel Jiménez-Guarneros and Gibran Fuentes-Pineda. “Cross-subject EEG-based emotion recognition via semisupervised multisource joint distribution adaptation”. In: *IEEE Transactions on Instrumentation and Measurement* 72 (2023), pp. 1–11.

[153] Yadong Chang et al. “Spatiotemporal gated graph transformer for EEG-based emotion recognition”. In: *IEEE Signal Processing Letters* 31 (2024), pp. 1630–1634.

[154] Xiang Li et al. “Exploring EEG features in cross-subject emotion recognition”. In: *Frontiers in neuroscience* 12 (2018), p. 162.

[155] Stavros P Adam et al. “No free lunch theorem: A review”. In: *Approximation and optimization: Algorithms, complexity and applications* (2019), pp. 57–82.

[156] Yaping Huang, Hanyong Bao, and Xuemei Qi. “Seismic random noise attenuation method based on variational mode decomposition and correlation coefficients”. In: *Electronics* 7.11 (2018), p. 280.

[157] Hanie Asemi and Nacer Farajzadeh. “Improving EEG signal-based emotion recognition using a hybrid GWO-XGBoost feature selection method”. In: *Biomedical Signal Processing and Control* 99 (2025), p. 106795.

[158] Anne-Sophie Baudry et al. “Sub-dimensions of trait emotional intelligence and health: A critical and systematic review of the literature”. In: *Scandinavian journal of psychology* 59.2 (2018), pp. 206–222.

[159] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. “Toward machine emotional intelligence: Analysis of affective physiological state”. In: *IEEE transactions on pattern analysis and machine intelligence* 23.10 (2001), pp. 1175–1191.

[160] Xiang Li et al. “Emotion recognition from multi-channel EEG data through convolutional recurrent neural network”. In: *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE. 2016, pp. 352–359.

[161] Zhihong Zeng et al. “A survey of affect recognition methods: audio, visual and spontaneous expressions”. In: *Proceedings of the 9th international conference on Multimodal interfaces*. 2007, pp. 126–133.

[162] Rania Alhalaseh and Suzan Alasasfeh. “Machine-learning-based emotion recognition system using EEG signals”. In: *Computers* 9.4 (2020), p. 95.

[163] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[164] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. “A fast learning algorithm for deep belief nets”. In: *Neural computation* 18.7 (2006), pp. 1527–1554.

[165] Ossama Abdel-Hamid et al. “Convolutional neural networks for speech recognition”. In: *IEEE/ACM Transactions on audio, speech, and language processing* 22.10 (2014), pp. 1533–1545.

[166] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012).

[167] Wenbin Yu et al. “Integrating spatio-temporal and generative adversarial networks for enhanced nowcasting performance”. In: *Remote Sensing* 15.15 (2023), p. 3720.

[168] Tengfei Song et al. “EEG emotion recognition using dynamical graph convolutional neural networks”. In: *IEEE Transactions on Affective Computing* 11.3 (2018), pp. 532–541.

[169] Fei Wang et al. “Emotion recognition with convolutional neural network and EEG-based EFDMs”. In: *Neuropsychologia* 146 (2020), p. 107506.

[170] Hao Chao and Liang Dong. “Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals”. In: *IEEE sensors journal* 21.2 (2020), pp. 2024–2034.

[171] Jean-Michel Combes, Alexander Grossmann, and Philippe Tchamitchian. *Wavelets: time-frequency methods and phase space proceedings of the international conference, marseille, france, december 14–18, 1987*. Springer Science & Business Media, 2012.

[172] Edmundo Pereira de Souza Neto et al. “Smoothed pseudo Wigner–Ville distribution as an alternative to Fourier transform in rats”. In: *Autonomic Neuroscience* 87.2-3 (2001), pp. 258–267.

[173] Alejandro Bachiller et al. “Decreased entropy modulation of EEG response to novelty and relevance in schizophrenia during a P300 task”. In: *European archives of psychiatry and clinical neuroscience* 265.6 (2015), pp. 525–535.

[174] Umit Budak et al. “An effective hybrid model for EEG-based drowsiness detection”. In: *IEEE sensors journal* 19.17 (2019), pp. 7624–7631.

[175] Shalu Chaudhary et al. “Convolutional neural network based approach towards motor imagery tasks EEG signals classification”. In: *IEEE Sensors Journal* 19.12 (2019), pp. 4494–4500.

[176] S Nawab, T Quatieri, and Jae Lim. “Signal reconstruction from short-time Fourier transform magnitude”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 31.4 (1983), pp. 986–998.