Published online on the page : https://ejournal.uinbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Human Emotion Classification Based on EEG Using FFT Band Power and LSTM Classifier

Dwi Wahyu Prabowo

Universitas Darwan Ali, Sampit, Indonesia

Article Information

Article History:

Submitted: October 10, 2025
Revision: December 3, 2025
Accepted: December 20, 2025
Published: December 23, 2025

Keywords

EEG
Emotion
FFT
SEED
LSTM

Correspondence

E-mail: dwi.wahyu9@unda.ac.id*

A B S T R A C T

This study investigates human emotion recognition using electroencephalogram (EEG) signals, focusing on the Shanghai Jiao Tong University Emotion EEG Dataset (SEED), which consists of recordings from 62 EEG channels categorized into three emotion classes: positive, neutral, and negative. The main challenges in EEG-based emotion classification include the limited amount of available data and the nonlinear, non-stationary nature of EEG signals. To address these challenges, this study evaluates the effectiveness of the Fast Fourier Transform (FFT) band power as input features and employs a stacked Long Short-Term Memory (LSTM) network as the classifier. Model validation was conducted using stratified 10-fold cross-validation, and performance was assessed using accuracy, F1-score, and Cohen's kappa metrics. Experimental results show that the proposed method achieved an average accuracy of 89.87%, an F1-score of 90.10%, and a Cohen's kappa value of 0.848, with minimal variation across folds, demonstrating high model stability. Unlike many recent studies that rely on image-based representations or Generative Adversarial Networks (GAN)-driven data augmentation, this study demonstrates that FFT band power combined with a sequential LSTM classifier can achieve strong performance without synthetic data generation or complex feature transformations. These findings indicate that the combination of FFT band power features and the LSTM classifier can serve as a solid baseline for further research.

1. Introduction

The development of brain–computer interface (BCI) technology has advanced rapidly, enabling a wide range of applications in healthcare, education, human–computer interaction, and entertainment. One of the most prominent research areas in this domain is emotion recognition based on electroencephalogram (EEG) signals [1]–[3]. The ability to detect emotions in real time is important in applications such as psychological therapy, mental health monitoring, and enhancing user experience in virtual reality environments. EEG-based approaches are considered superior to those relying on facial expressions or speech, as EEG directly reflects neural activity and is relatively resistant to manipulation [4]–[6].

However, several major challenges remain. First, EEG signals are nonlinear, nonstationary, and highly complex, making feature extraction difficult. These characteristics often render conventional methods, such as time- or frequency-domain analysis, insufficient to capture the full dynamics of EEG signals [7]–[9]. Second, the limited amount of data poses a serious constraint. EEG data collection requires specialized equipment, high costs, extended recording sessions, and limited subject participation. Consequently, classification models often suffer from generalization issues, particularly when trained on small datasets [10]–[13].

Various approaches have been proposed to address these challenges. For instance, EEG signals have been represented as images such as recurrence plots (RP) and spectrograms (SP), which are then processed using convolutional neural networks (CNNs). This approach leverages the power of deep learning to capture complex patterns within brain signals [14], [15]. Meanwhile, to mitigate data scarcity, recent studies have employed data augmentation techniques based on generative adversarial networks (GANs), which can generate synthetic samples with characteristics similar to real data. As a result, research in this field is evolving toward the integration of advanced representation methods and data augmentation strategies [16]–[19].

Nevertheless, image-based representation and data augmentation approaches often require substantial computational resources and complex processing pipelines. As an alternative, Fast Fourier transform (FFT) band power remains a widely used feature extraction method due to its simplicity and computational efficiency [20], [21]. The FFT decomposes EEG signals into several frequency bands—δ, θ, α, β, and γ—each associated with specific cognitive and emotional states, thus providing a solid foundation for emotion classification. Furthermore, FFT is relatively easy to implement and computationally efficient, making it a relevant baseline before incorporating image-based representations [22], [23].

Based on this background, the present study focuses on EEG-based emotion classification using features derived from FFT band power. The SEED dataset [24] was employed, as it is widely acknowledged as a reliable benchmark for emotion recognition research due to its high quality and experimental richness. Model validation was carried out through a stratified 10-fold cross-validation scheme to ensure robustness and generalizability. Model performance was assessed using three evaluation metrics—accuracy, F1-score, and Cohen's kappa—with the latter providing a more rigorous measure of agreement that accounts for random chance. The outcomes of this research are intended to establish a strong baseline for future comparisons with more advanced methods, including multi-representation convolutional neural networks (CNNs) and generative adversarial network (GAN)-based data augmentation techniques.

To situate this work within recent developments, it is important to note that substantial progress has been made through GAN-based augmentation, nonlinear image representations, and attention-based deep learning models. However, relatively little research has examined how effectively a lightweight and interpretable spectral approach can perform without synthetic data or high-capacity architectures. Existing studies often report accuracy gains after augmentation or complex transformations, yet the performance of a purely frequency-domain baseline has not been revisited systematically. Therefore, this study fills that gap by evaluating FFT band power combined with an LSTM classifier as an efficient baseline that can perform competitively even without augmentation. This contribution provides a clear comparison point for future research exploring GANs, multi-representation models, or hybrid pipelines.

2. Related Works

Research on emotion recognition based on electroencephalogram (EEG) signals has grown rapidly in recent years, with diverse approaches focusing on improving signal representation, model generalization, and interpretability of classification results. Several recent studies have demonstrated that the success of EEG-based emotion recognition systems is strongly influenced by how signals are represented prior to classification, as well as by the model's ability to capture the temporal and spatial dynamics of brain activity.

Prabowo et al. [25] introduced the Asymmetric Windowing Recurrence Plot (AWRP) as a novel strategy for transforming EEG signals into more informative nonlinear image representations. This method addresses

the limitations of conventional symmetric recurrence plots (RP), which often contain redundant information. By applying asymmetric windowing, AWRP generates richer and more discriminative texture patterns, significantly improving classification accuracy on the DEAP and SEED datasets, achieving results above 99%. This study highlights the importance of signal representation in emotion classification systems and demonstrates that nonlinear, image-based transformations can provide substantial accuracy.

In contrast, Roshanaei et al. [26] adopted a different perspective by analyzing functional and effective connectivity using graph theory to better understand inter-regional brain relationships during emotional experiences. Their findings emphasize that emotional information is not only reflected in local signal power (such as band power) but also in cross-regional brain interactions. Using coherence analysis and Granger causality, the study revealed that beta and gamma frequency bands play key roles in high-intensity negative emotions, while alpha activity correlates with relaxed states. These results reinforce the view that connectivity-based representations can complement conventional spectral analysis in understanding the neural dynamics of emotion.

Efforts to improve EEG model generalization have also become a major research focus. Song et al. [12] proposed Domain Generalization through Latent Distribution Exploration for motor imagery EEG classification. Leveraging a multi-branch deep learning architecture that combines EEGNet and ShallowConvNet, the model learns latent distributions from multiple source domains, allowing it to adapt to new subjects without recalibration. This approach significantly improved cross-subject accuracy on the BCI Competition IV and PhysioNet datasets, demonstrating the potential applicability of domain generalization concepts to EEG-based emotion recognition, which faces similar inter-individual variability challenges.

A broader generalization approach was proposed by Zhong et al. [11] through the EEG-DG (Multi-Source Domain Generalization) framework. This model is designed to construct domain-invariant feature representations by optimizing both marginal and conditional distribution alignment across source domains using Maximum Mean Discrepancy (MMD). By integrating an Inception-ResNet architecture with a dual-alignment mechanism, EEG-DG achieved accuracies exceeding 80% on the BCI Competition IV dataset without requiring target data during training. This framework represents a key advancement in developing adaptive and efficient EEG systems and opens opportunities for similar applications in emotion recognition involving heterogeneous data across subjects and sessions.

Unlike the aforementioned studies emphasizing generalization, Tao et al. [3] introduced an attention-based strategy through the Attention-based Convolutional Recurrent Neural Network (ACRNN). The model combines channel-wise attention to evaluate the contribution of individual EEG channels and self-attention to capture temporal dependencies between signal segments. Results on the DEAP and DREAMER datasets showed accuracy improvements exceeding 93%, while also enhancing interpretability by identifying brain regions most relevant to specific emotions. The main contribution of this study lies in the simultaneous integration of spatial and temporal attention mechanisms, which strengthens spatio-temporal EEG representation without compromising model transparency.

Furthermore, de Paula et al. [14] proposed an approach that transforms EEG signals into image representations using image encoding techniques such as Recurrence Plot (RP), Gramian Angular Field (GAF), and Markov Transition Field (MTF). By employing two-dimensional CNN architectures, the study achieved classification accuracies of up to 97% in SSVEP-based EEG, confirming that converting signals into the visual domain can improve the model's ability to capture nonlinear patterns. This approach reinforces the growing trend of image-based representations increasingly adopted in EEG emotion recognition using CNN and transformer architectures.

From the above review, it can be concluded that recent advances in EEG-based emotion recognition predominantly focus on nonlinear signal representations (e.g., RP and GAF), attention and connectivity

mechanisms, and models emphasizing generalization and computational efficiency. However, most of these approaches require large computational resources, extensive training data, and complex model configurations. In this context, the use of FFT band power as a fundamental feature remains essential as a strong and interpretable baseline because it effectively captures key spectral information with high efficiency. When combined with LSTM networks, this approach can model the temporal dynamics of EEG signals without requiring complex transformations or image-based augmentation.

Therefore, this study remains relevant as it reaffirms the potential of the FFT band power + LSTM method as a solid starting point before integrating more complex frameworks such as multi-representation CNNs or GAN-based augmentation. By emphasizing a balance between accuracy, computational efficiency, and interpretability, this work provides an empirical foundation for developing adaptive, lightweight, and practical EEG-based emotion recognition systems suitable for real-world applications.

3. Method

3.1. Dataset

This study employs the SEED dataset [24] as the primary data source due to its rich characteristics and strong relevance to EEG-based emotion classification. The dataset comprises 62-channel EEG recordings collected from 15 subjects, with each subject participating in three separate experimental sessions conducted approximately one week apart to enhance data reliability. In each session, subjects were instructed to watch 15 film clips of approximately four minutes each, specifically designed to elicit positive, neutral, and negative emotional states. This protocol yielded a total of 675 trials (15 subjects × 15 films × 3 sessions), providing substantial data variability for analysis. Table 1 show the characteristics of the SEED dataset.

Table 1. Characteristics of the SEED Dataset

Attribute	Description
Number of Subjects	15 (recorded across three separate sessions)
Number of EEG Channels	62
Sampling Frequency	200 Hz
Stimulus Duration	Approximately 4 minutes per film clip
Number of Film Clips	15 (each designed to elicit a specific emotion)
Emotion Types	Positive, Neutral, Negative
Data Format	Raw EEG (.mat) with ground truth labels from self-assessment
Experimental Scheme	Each subject watched 15 film clips, repeated across three sessions at one-week intervals
Total Data	15 subjects × 15 films × 3 sessions = 675 trials

The dataset's high temporal resolution (200 Hz) and larger number of channels compared to other datasets, such as DEAP dataset [27], enable a more detailed and comprehensive exploration of EEG signals. Furthermore, SEED provides discrete and explicit emotion labels, making it particularly suitable for this study, which focuses on direct classification of three emotional categories without requiring transformation from a valence–arousal scale. Given these advantages, SEED is an appropriate choice to support the emotion classification experiments based on FFT band power features.

3.2. Preprocessing

The original SEED recordings were sampled at 200 Hz. In this study, the signals were intentionally resampled to 70 Hz before segmentation. This downsampling retains all relevant EEG frequency components up to the low gamma band, reduces high-frequency noise, and ensures that the 256-sample window corresponds to approximately 3.66 seconds, which is a commonly used duration for stable FFT band-power estimation. According to the Nyquist theorem, the maximum frequency that can be represented is half of the sampling rate. With a 70 Hz sampling frequency, the preserved spectrum extends up to 35 Hz, which fully covers the delta, theta, alpha, beta, and lower gamma bands that are widely used for EEG-based emotion recognition [28], [29]. Reducing the sampling rate also suppresses high-frequency noise and muscle artifacts that fall outside the range of cognitive and affective EEG activity [28], [30]. Through this resampling step, the preprocessing pipeline produces cleaner, noise-reduced signals while preserving all spectral components relevant to subsequent FFT band-power extraction.

The preprocessing phase is an essential part of EEG-based analysis since raw recordings typically contain artifacts, noise, and subject-specific variations that may disrupt the classification process. To address this, the present study applies the FFT to convert EEG data from the temporal representation into the frequency domain. Through this transformation, the signal is decomposed into its constituent frequency components, which are more interpretable because brain activity is strongly linked to distinct oscillatory patterns. For instance, delta waves are commonly linked to deep sleep, whereas beta waves correspond to levels of concentration. Consequently, FFT provides a feature representation that is not only compact but also physiologically meaningful [31], [32]. The results of the FFT transformation are then used to compute the band power across five major frequency ranges: δ, θ, α, β, and γ. Each frequency band holds particular relevance to cognitive and emotional states. For example, increased activity in the theta band is often associated with emotional engagement, while the gamma band is related to complex information processing. By extracting the signal power from each band, the resulting features capture psychological aspects pertinent to emotion classification [26], [33]. This approach was chosen for its simplicity, computational efficiency, and ability to provide clearer physiological interpretations compared to more complex image-based representation methods.

3.3. Feature Extraction Based on Fast Fourier Transform (FFT)

To capture the spectral characteristics of EEG signals associated with emotional states, this study employs the FFT as the basis for feature extraction [34]. Each raw EEG recording from 62 channels was resampled to 70 Hz prior to segmentation and feature extraction, ensuring that all subsequent FFT computations operated on the same frequency basis.. The raw EEG signalis divided into overlapping segments (windows), each consisting of 256 samples (≈ 3.66 seconds) with a step size of 16 samples (≈ 0.23 seconds). For each time segment, the time-domain signal $x[n]$ is transformed into its frequency-domain representation using the Discrete Fourier Transform (DFT), defined as follows:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, k = 0,1,...,N-1 \quad (1)$$

where N denotes the FFT window length. The transformation yields an amplitude spectrum $|X[k]|$, which represents the contribution of each frequency component in the EEG signal. Next, the average spectral power (band power) is computed across five primary frequency bands based on the SEED dataset configuration with cutoff frequencies at [3, 4, 8, 14, 32] Hz. For each frequency band $[f_1, f_2]$, the average spectral power (Power Spectral Density, PSD) is calculated using the following equation:

$$P_{[f_1,f_2]} = \frac{1}{M} \sum_{k=f_1}^{f_2} |X[k]|^2 \quad (2)$$

where M represents the number of frequency bins within the defined range. This process produces five spectral power values for each channel, resulting in a total of 62 × 5 = 310 features per EEG segment. This approach preserves the spectral information of each EEG channel in both spatial and temporal dimensions while significantly reducing data dimensionality compared to raw signals. The resulting features in the frequency domain are then provided as inputs to the Long Short-Term Memory (LSTM) classifier, as detailed in Subsection 2.4.

The FFT band power was selected as the primary feature because EEG rhythms are well localized within canonical frequency ranges with well-established physiological relevance. Computing the power within δ, θ, α, β, and γ bands provides a compact summary of neural oscillations associated with arousal, attention, and affective processing, while eliminating the need for high-capacity or image-based models. A band-pass filter between 1–50 Hz was applied to remove slow drifts and muscle artifact. Limiting the analysis to frequencies below 50 Hz minimizes contamination from high-frequency artifacts yet retains lower γ activity, which has

been reported as informative for emotion recognition. This configuration achieves a balance between physiological interpretability, robustness to noise, and computational efficiency [26], [31]–[33].

3.4. Long Short-Term Memory (LSTM) Classifier

The classification model employed in this study is a five-layer stacked Long Short-Term Memory (LSTM) network implemented using Keras/TensorFlow. LSTM-based models, including stacked, have demonstrated strong performance in emotion recognition tasks using FFT or similar frequency-domain features [35], [36]. The input consists of FFT band power features derived from 62 EEG channels across five frequency bands, formatted into a three-dimensional structure [batch, timestep, feature] to ensure compatibility with the LSTM architecture. Each LSTM layer is followed by batch normalization to stabilize activation distributions and dropout with ratios ranging from 0.2 to 0.3 to reduce the risk of overfitting. As illustrated in Figure 1, the architecture begins with an LSTM layer containing 512 units, followed sequentially by layers with 256, 128, 64, and 32 units, respectively. The final output is projected onto a Dense layer with three units—corresponding to the number of emotion classes—using a softmax activation function. This configuration enables the model to capture nonlinear temporal dynamics in EEG feature representations while preserving relevant sequential information through the gating mechanisms inherent to LSTM networks. The detailed configuration and training parameters of the classifier are summarized in Table 2, which lists the key hyperparameter specifications, including layer structure, normalization strategy, regularization methods, and training settings such as loss function, optimizer, learning rate, batch size, and validation scheme.

Table 2. Hyperparameter Specifications of the LSTM classifier

Component	Specification
Input	FFT band power features (62 channels × 5 frequency bands), formatted as [batch, timestep, feature]
LSTM Architecture	512 → 256 → 128 → 64 → 32 units (each layer followed by BN + Dropout)
Normalization	Batch normalization after each LSTM layer
Regularization	Dropout (0.2–0.3)
Output Layer	Dense (3) + Softmax activation
Loss Function	Categorical cross-entropy
Optimizer	Adam (learning rate = 1e-4)
Epochs	50
Batch Size	150
Validation	Stratified 10-fold cross-validation
Evaluation Metrics	Model performance assessed using accuracy, F1-measure, Cohen's kappa coefficient, and confusion matrix analysis

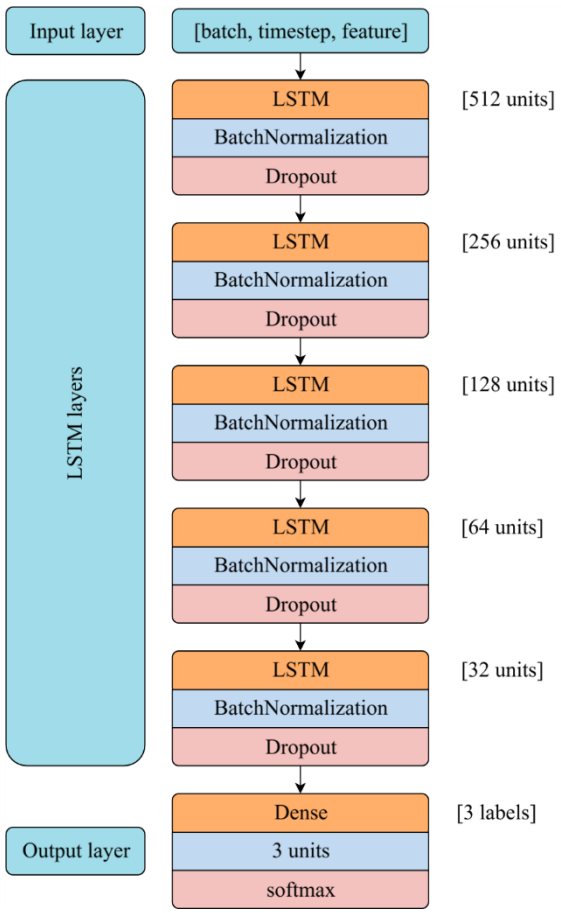

Figure 1. LSTM Architecture

3.5. *Experimental setting*

The experimental design in this study was structured systematically to ensure that all stages of data processing, feature extraction, model training, and performance evaluation were conducted in a consistent and reproducible manner. The experimental workflow consisted of five main stages: (1) EEG signal acquisition and segmentation, (2) transformation and feature extraction using the FFT, (3) data normalization using *StandardScaler*, (4) classification model training based on LSTM, and (5) evaluation using stratified 10-fold cross-validation with quantitative performance metrics. Each stage was executed sequentially with consistent parameters and configurations across all iterations.

Stage 1: EEG Signal Acquisition and Segmentation

The EEG data used in this study were obtained from the SEED dataset, which consists of multichannel EEG recordings comprising 62 channels with a sampling frequency of 200 Hz. Each EEG recording was collected from 15 participants across three separate sessions, with each session containing 15 film clips of approximately four minutes in duration. All EEG signals were stored in *raw EEG* (.mat) format, represented as amplitude matrices over time.

Before applying the transformation, each EEG channel was segmented into overlapping time windows. The window length was set to 256, which corresponds to approximately 3.66 seconds after the signals were resampled to 70 Hz. All segmentation and feature extraction steps were performed on the resampled data. The step size is 16 samples, resulting in an overlap of 93.75% between consecutive windows. The overlapping segmentation was applied to retain temporal continuity within the EEG signal and to ensure that the model received sufficient temporal dynamics for each observation. Each window was treated as an independent observation unit during feature extraction. To ensure consistency across the entire pipeline, all SEED recordings were uniformly resampled from 200 Hz to 70 Hz before segmentation.

The segmentation structure is illustrated in Figure 2, showing that each EEG channel was divided into multiple overlapping short-duration segments. Each segment was subsequently transformed into the frequency domain, and the resulting FFT representations were aggregated to form a comprehensive feature set for the classification stage.

Stage 2: FFT Transformation and Band Power Feature Extraction

After segmentation, each EEG segment $x[n]$ was transformed from the time domain to the frequency domain using the FFT. FFT decomposes the EEG signal into its constituent frequency components. The transformation was applied to each channel and time window. The resulting output was a complex-valued frequency spectrum vector $X[k]$, representing the amplitude and phase contributions at each frequency index (k). From the FFT output, only the magnitude spectrum $(|X[k]|)$ was utilized, as it reflects the relative strength of each frequency component without considering phase information.

To obtain features representing the spectral characteristics of brain activity, the magnitude values $(|X[k]|)$ were grouped into five canonical EEG frequency bands: δ (1–3 Hz), θ (4–7 Hz), α (8–13 Hz), β (14–31 Hz), and γ (32–35 Hz). For each frequency band $[f_1, f_2]$, the mean magnitude value was computed using the following equation:

$$P_{[f_1,f_2]} = \frac{1}{M}\sum_{k=f_1}^{f_2}|X[k]| \tag{3}$$

where (M) is the number of frequency bins within the specified band. This computation produced five *band power* values per EEG channel. Given 62 channels, each EEG segment yielded a total of 62 × 5 = 310 features. All channel-wise features were then concatenated into a single feature vector of 310 dimensions, representing the spectral energy distribution of the EEG signal within that segment.

The complete process is depicted in Figure 2, which consists of four main blocks: (1) raw EEG signal input, (2) overlapping segmentation with 256-sample windows and 16-sample step size, (3) FFT transformation and mean magnitude calculation within each frequency band, and (4) feature assembly by concatenating all channels. The diagram clearly illustrates how the EEG signal was converted into a numerical representation suitable for subsequent classification.

Stage 3: Feature Normalization

After feature extraction, all feature vectors were normalized. Normalization was applied separately within each training fold to prevent *data leakage*. Then the resulting mean and standard deviation parameters were subsequently used to transform the corresponding test data within the same fold. This ensured that all extracted features followed a standardized distribution with zero mean and unit variance. This normalization step is important because FFT-derived features across frequency bands show different numerical ranges. Without normalization, frequency bands with higher magnitudes (such as β and γ) might dominate the learning process and bias the model toward those bands. Normalization also facilitates faster and more stable convergence during neural network training, as gradients remain consistent across feature dimensions.

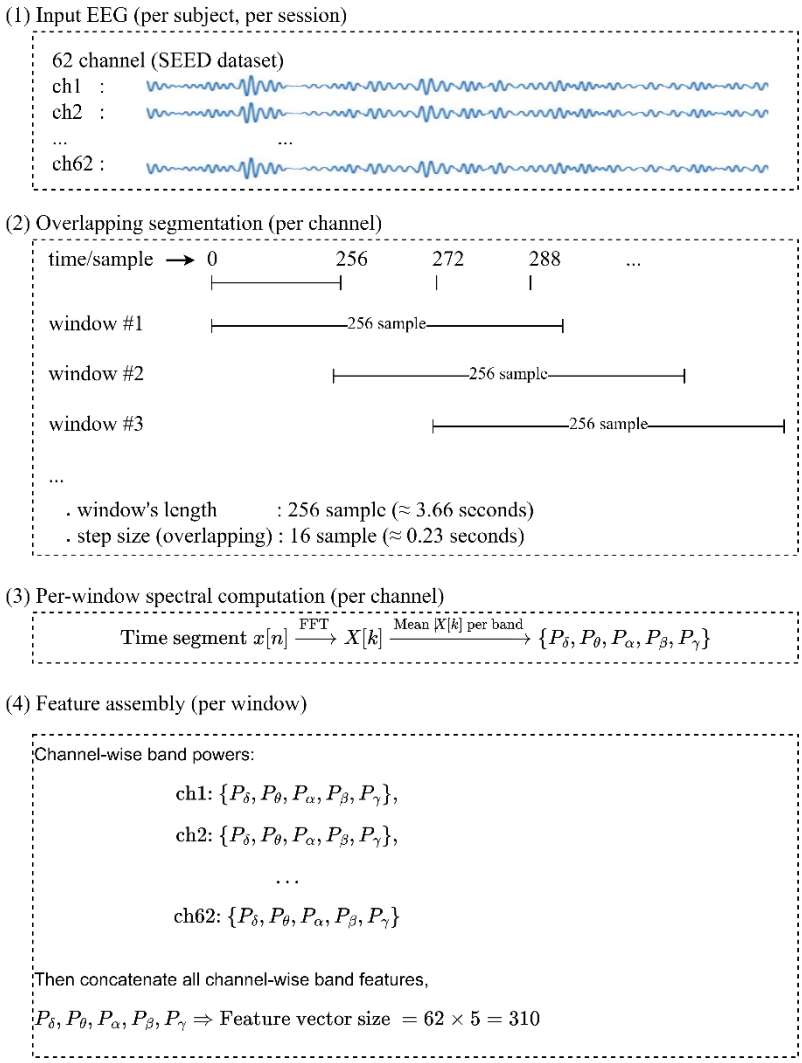

Figure 2. Illustration of EEG signal segmentation and FFT-based feature extraction. Each channel is divided into overlapping windows, transformed into the frequency domain using FFT, averaged within five canonical frequency bands (δ, θ, α, β, γ), and concatenated into a 310-dimensional feature vector per segment. All EEG recordings were resampled to 70 Hz prior to segmentation, ensuring consistent temporal resolution across all overlapping windows.

Stage 4: LSTM Model Training

The next stage involved training the classification model using a five-layer stacked LSTM network. The detailed architecture is described in Subsection 3.4, but in summary, it consisted of sequential LSTM layers with 512, 256, 128, 64, and 32 units, respectively. Each LSTM layer was followed by *batch normalization* and *dropout* (ranging from 0.2 to 0.3) to stabilize training and reduce overfitting. The final output layer was a *Dense* layer with three neurons and a *softmax* activation function, corresponding to the three emotion categories: positive, neutral, and negative. The model was compiled using the *categorical cross-entropy* loss function and optimized with the *Adam* optimizer at a learning rate of 1×10^{-4}. Training was performed for 50 epochs with a batch size of 150, without *early stopping*, to maintain consistency across all folds. All experiments were executed in a GPU-accelerated (DGX-1 V100) TensorFlow environment to ensure uniform training speed and reproducibility. Identical initialization and parameter configurations were applied across iterations to eliminate random variability in training outcomes.

Stage 5: Model Evaluation and Validation

Model performance was evaluated using stratified 10-fold cross-validation. The dataset was divided into ten subsets with balanced distributions across the three emotion classes (positive, neutral, and negative). In each iteration, nine subsets were used for training and one for testing. This process was repeated ten times so that every subset served once as the test data. This validation approach provided more stable and representative performance estimates than hold-out validation and better assessed the model's ability to generalize across subject-level variations. Three key metrics were used to evaluate performance: *accuracy*, *F1-score*, and *Cohen's kappa*.

- *Accuracy* measured the overall proportion of correctly predicted samples.
- *F1-score* evaluated the harmonic balance between *precision* and *recall*, which is particularly relevant in cases of minor class imbalance.
- Cohen's kappa served as an additional, stricter validation metric by accounting for the probability of random agreement between model predictions and actual labels. A kappa value greater than 0.80 indicates almost perfect agreement, signifying a high level of consistency between the model's predictions and the ground truth.

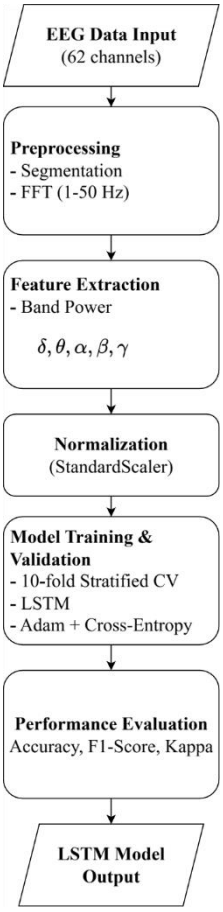

Figure 3. The complete experimental pipeline of the proposed EEG-based emotion recognition framework. The process includes data acquisition from 62 EEG channels, segmentation and FFT-based preprocessing, band power feature extraction across δ, θ, α, β, and γ frequency bands, feature normalization using StandardScaler, model training and validation with a 10-fold stratified cross-validation scheme, and final performance evaluation using accuracy, F1-score, and Cohen's kappa metrics.

The complete experimental pipeline is summarized in Figure 3, which illustrates the systematic workflow from raw EEG acquisition, segmentation, and FFT-based feature extraction to normalization, LSTM training, and cross-validation-based evaluation. The diagram provides a comprehensive overview of how each processing block contributes to the construction of an efficient and generalizable EEG-based emotion

classification system. Through this structured experimental design and validation procedure, the study ensures that the classification results reflect stable, measurable, and unbiased model performance, free from data imbalance or information leakage across folds. In this context, the SEED dataset does not require any balancing procedure because its emotion labels are uniformly distributed. Each subject provides 15 samples for each class (positive, neutral, and negative), which results in an overall balanced dataset as reported by Zheng & Lu [24] and Li et al. [37]. The approach also guarantees that the results can be replicated by future studies using the same dataset and configuration.

4. Results and Discussion

4.1. Experimental Results per Fold

The experimental findings summarized in Table 3 report the model's performance in classifying emotions using FFT band power features evaluated on the SEED dataset through a stratified 10-fold validation scheme. The obtained accuracy scores across folds vary only slightly, ranging from 0.8978 to 0.9001 with an average of 0.8987. This indicates that the model performs consistently across all data partitions without any significant performance degradation in specific folds. Such consistency confirms that the implemented pipeline is stable and not strongly affected by variations in data subsets.

In addition to accuracy, the F1-score values ranging from 0.9000 to 0.9023 demonstrate a good balance between precision and recall. With an average of 0.9010, the results indicate that the model is capable of recognizing all emotion classes (positive, neutral, and negative) in a balanced manner rather than focusing only on the dominant class. The stability of the F1-score further supports the claim that FFT band power features provide sufficient discriminative information to distinguish emotional states effectively, despite being a relatively simple representation compared to image-based EEG features.

Cohen's kappa metric also provides an important perspective, yielding an average value of 0.8481. According to Landis and Koch's interpretation, this falls into the "almost perfect agreement" category, meaning the model's predictions show a very high level of agreement with the true labels, even after accounting for random chance. This demonstrates that the model's performance is not only characterized by high accuracy but also statistically valid in reflecting its classification capability. In other words, the model truly differentiates between emotion classes rather than relying on data distribution biases.

Meanwhile, the loss values remain stable, with an average of 0.1763, indicating that the model consistently minimizes prediction errors during both training and testing. The very small variation in loss across folds (0.1745–0.1779) further confirms that the model exhibits strong generalization ability toward unseen data. The combination of accuracy, F1-score, kappa, and loss results demonstrates that FFT band power serves as a robust baseline for EEG-based emotion classification and provides a solid foundation for future research involving more complex approaches, such as multi-representation CNNs or GAN-based data augmentation.

Table 3. Emotion classification results on the SEED dataset using FFT band power

Fold	Accuracy	F1 score	Kappa	Loss
1	0.8997	0.9019	0.8495	0.1753
2	0.8990	0.9012	0.8485	0.1767
3	0.8985	0.9007	0.8477	0.1779
4	0.8980	0.9001	0.8470	0.1755
5	0.8978	0.9000	0.8467	0.1768
6	0.9001	0.9023	0.8501	0.1745
7	0.8974	0.8996	0.8461	0.1765
8	0.8994	0.9016	0.8491	0.1765
9	0.8991	0.9013	0.8486	0.1765
10	0.8987	0.9009	0.8480	0.1765

4.2. Overall Performance and Discussion

The experimental results indicate that the average accuracy nearly reached 90%, specifically 89.87%. This is a relatively high value for a baseline approach that solely utilizes FFT band power as the main feature. Given that FFT is a simple method focusing on energy representation across specific frequency bands, this finding demonstrates that frequency information in EEG signals has a strong correlation with emotional states. Although image-based approaches such as recurrence plot (RP) and spectrogram (SP) are often reported to yield superior performance, the present results emphasize that the classical FFT-based approach remains relevant and competitive for emotion classification [25], [38], [39].

Beyond accuracy, the stable F1-score around 0.90 provides a more comprehensive perspective on the model's performance. The F1-score, which harmonizes precision and recall, shows that the model is capable of identifying the majority class and maintains balanced recognition across all emotion categories (positive, neutral, and negative). This is essential because EEG experiments often involve class imbalance. When relying solely on accuracy, models may exhibit bias toward dominant classes. However, the consistently high F1-score confirms that the FFT band power pipeline successfully maintains balanced performance across emotion classes.

Cohen's kappa further reinforces the validity of these findings. With an average of 0.848, the score falls within the "almost perfect agreement" category according to the Landis and Koch interpretation [40]. This indicates that the model's predictions exhibit a very high level of agreement with the ground truth labels, even after adjusting for random chance. The kappa metric is particularly important in EEG-based emotion classification, which is susceptible to inter-subject variability and signal noise. With Kappa > 0.80, it can be concluded that the model's performance is not random but genuinely reflects its capability to recognize emotional patterns in EEG signals.

Result stability is another key point in this analysis. The variation across folds is extremely small, with a standard deviation in accuracy of less than 0.001. This demonstrates that the experimental pipeline is robust against variations in training and testing data. In other words, the model is not overly sensitive to changes in data subsets, indicating good generalization ability. Such stability is essential for real-world applications of EEG-based emotion recognition systems, where incoming data may differ from training data.

Overall, the consistently strong results across all evaluation metrics (classification accuracy, F1-measure, and Cohen's kappa coefficient) demonstrate that the FFT band power method serves as a strong baseline for EEG-based emotion recognition research. Although the results do not yet surpass those of deep learning approaches involving multi-representations or data augmentation, the FFT-based pipeline provides competitive performance with higher interpretability and lower computational complexity. This foundation can support future research aimed at developing more advanced models, such as combining FFT with RP/SP or employing GANs, to further improve accuracy without compromising the demonstrated stability and interpretability of the current approach.

4.3. Analysis of Confusion Matrix

The results presented and visualized in Figure 4 indicate that the emotion classification model based on FFT band power can accurately recognize all three emotion classes with relatively high performance. Across the 10-fold cross-validation, the average number of correct predictions reached 64,269 for the positive class, 70,368 for the neutral class, and 65,378 for the negative class. These results show that the distribution of correct predictions is well balanced across all classes, suggesting that no emotion category was neglected. In other words, the model does not show bias toward any specific emotion class. This finding demonstrates that band power distributions contribute substantially to distinguishing emotional states.

However, Figure 4 reveals a consistent confusion pattern, primarily involving the neutral class. Specifically, 7,614 positive samples and 7,574 negative samples were misclassified as neutral. The darker shades in the [Positive → Neutral] and [Negative → Neutral] cells highlight the model's tendency to safely

assign uncertain predictions to the neutral category. This phenomenon is reasonable, as the neutral emotional state is physiologically positioned near the midpoint of the positive–negative spectrum, causing its EEG patterns to often resemble both other classes.

Conversely, misclassifications from the neutral class to other classes were relatively small, with 1,351 instances to the positive class and 2,464 to the negative class. This indicates that the model tends to be more conservative when classifying neutral signals. Such behavior reduces the risk of extreme prediction errors, such as incorrectly labeling negative emotions as positive. However, the trade-off is a slight decrease in sensitivity when detecting positive and negative emotions. This pattern is consistent with multiple studies who reported that the neutral emotion class frequently serves as a buffer or intermediary zone between positive and negative emotions in EEG-based emotion classification systems [41], [42].

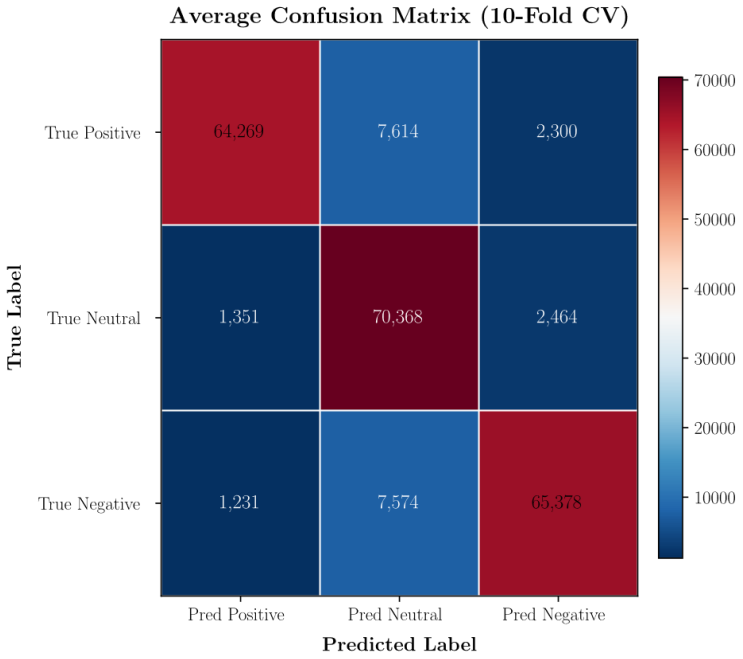

Figure 4. Heatmap Visualization of the Average Confusion Matrix (10-Fold CV)

Overall, Figure 4 demonstrates that the FFT band power pipeline produces a stable model with a high level of agreement with the ground truth labels. This is supported by a Cohen's kappa value of 0.848, which falls within the "almost perfect agreement" category according to Landis and Koch (1977) [40]. Therefore, although some confusion remains within the neutral class, the model's performance remains consistent and statistically valid. This confusion matrix analysis further reinforces the claim that FFT band power serves as a strong baseline for EEG-based emotion recognition, providing a foundation for future improvements through multi-representation approaches [43] or GAN-based data augmentation [16], [18], [19], [44] to enhance inter-class separability.

4.4. Comparison with Related Works

To contextualize the performance of the proposed FFT-based LSTM model, a comparison with several representative studies on SEED is presented in Table 4. These studies employ various data augmentation strategies, primarily through noise injection or the use of generative adversarial networks (GANs), which are designed to increase data diversity and improve classification performance in limited-data scenarios such as EEG emotion recognition.

Wang et al. [45]enhanced the SEED dataset by adding Gaussian noise and trained a ResNet18 classifier, achieving an accuracy of 75.00 percent. Despite the additional artificially generated samples, the performance remains relatively modest, suggesting that noise-based augmentation may not effectively capture the complex temporal–spectral characteristics of EEG emotional dynamics. Luo and Lu [46] introduced a conditional

Wasserstein GAN (cWGAN) to generate synthetic samples and used an SVM classifier, yielding an accuracy of 86.96 percent. Luo et al. [47] further extended this approach by employing a conditional Boundary Equilibrium GAN (cBEGAN), producing a slightly improved accuracy of 87.56 percent. Zhang et al. [48] developed a multi-generator cWGAN (MG-cWGAN) framework to better model EEG variability, yet the reported accuracy was 84.00 percent.

Studies	Methods	Accuracy (%)
Wang et al. [45]	Gaussian noise + ResNet18	75.00
Luo & Lu [46]	cWGAN + SVM	86.96
Luo et al. [47]	cBEGAN + SVM	87.56
Zhang et al. [48]	MG-cWGAN + SVM	84.00
This study	No augmentation + (FFT & LSTM)	**89.87**

Compared to these augmentation-based methods, the proposed approach demonstrates stronger performance even without any synthetic sample generation. As shown in Table 4, this study achieved an accuracy of 89.87 percent using only the original SEED data and a straightforward feature extraction pipeline based on FFT band power. This performance advantage highlights the discriminative power of spectral features when combined with a deep sequential classifier such as LSTM. Furthermore, the stability demonstrated across the folds in this study indicates that high-quality spectral representations can offer a reliable foundation without the need for computationally intensive augmentation strategies.

It is also important to note that Section 4.3 identified remaining confusion between positive and negative classes converging toward the neutral class, suggesting potential benefits from controlled data expansion. Therefore, although the present study demonstrates competitive performance without augmentation, the findings are well aligned with future research directions involving GAN-based augmentation, particularly using cWGAN, cBEGAN, or MG-cWGAN frameworks. Integrating these generative models with FFT-based representations may further enhance class separability and improve robustness in emotion classification systems.

5. Conclusion

This study successfully demonstrates that FFT band power can serve as a strong baseline for EEG-based emotion classification. With an average accuracy of 89.87%, an F1-score of 90.10%, and a Cohen's kappa of 0.848, the proposed method exhibits stable performance and high statistical validity. The consistent stability across folds indicates that the experimental pipeline is not only reliable for specific data subsets but also possesses good generalization capability. These findings confirm that although FFT band power is a relatively simple method, it remains relevant and effective for extracting essential information from EEG signals, particularly in the context of human emotion analysis. For future research, this baseline can be extended by integrating more advanced approaches to enhance model accuracy and generalization. One potential direction is the use of multi-representation frameworks that combine recurrence plot (RP) and spectrogram (SP) features to capture both nonlinear characteristics and time–frequency distributions of EEG signals. Another alternative is the application of generative adversarial network (GAN)-based data augmentation to overcome the limitations of sample size and increase training data diversity. By combining the strengths of classical methods such as FFT with modern deep learning approaches, future studies are expected to develop EEG-based emotion recognition systems that are more accurate, robust, and suitable for real-world applications. Furthermore, comparative results reported in the literature show that a straightforward FFT and LSTM pipeline can surpass several augmentation-based methods, underscoring its value as a strong and computationally efficient baseline for future EEG-based emotion recognition research.

References

[1] X. Li *et al.*, "EEG Based Emotion Recognition: A Tutorial and Review," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–57, Apr. 2023, doi: 10.1145/3524499.

[2] C. Li, Z. Zhang, X. Zhang, G. Huang, Y. Liu, and X. Chen, "EEG-based Emotion Recognition via Transformer," vol. 19, no. 4, pp. 1–10, 2022.

[3] W. Tao *et al.*, "EEG-Based Emotion Recognition via Channel-Wise Attention and Self Attention," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 382–393, Jan. 2023, doi: 10.1109/TAFFC.2020.3025777.

[4] D. W. Prabowo, H. A. Nugroho, N. A. Setiawan, and J. Debayle, "A systematic literature review of emotion recognition using EEG signals," *Cogn. Syst. Res.*, vol. 82, no. July, p. 101152, 2023, doi: 10.1016/j.cogsys.2023.101152.

[5] N. S. Suhaimi, J. Mountstephens, and J. Teo, "EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–19, Sep. 2020, doi: 10.1155/2020/8875426.

[6] H. Liu *et al.*, "EEG-Based Multimodal Emotion Recognition: A Machine Learning Perspective," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–29, 2024, doi: 10.1109/TIM.2024.3369130.

[7] A. K. Singh and S. Krishnan, "Trends in EEG signal feature extraction applications," *Front. Artif. Intell.*, vol. 5, Jan. 2023, doi: 10.3389/frai.2022.1072801.

[8] X. Geng, D. Li, H. Chen, P. Yu, H. Yan, and M. Yue, "An improved feature extraction algorithms of EEG signals based on motor imagery brain-computer interface," *Alexandria Eng. J.*, vol. 61, no. 6, pp. 4807–4820, Jun. 2022, doi: 10.1016/j.aej.2021.10.034.

[9] I. Rakhmatulin, M.-S. Dao, A. Nassibi, and D. Mandic, "Exploring Convolutional Neural Network Architectures for EEG Feature Extraction," *Sensors*, vol. 24, no. 3, p. 877, Jan. 2024, doi: 10.3390/s24030877.

[10] S. Liang *et al.*, "Domain-Generalized EEG Classification With Category-Oriented Feature Decorrelation and Cross-View Consistency Learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3285–3296, 2023, doi: 10.1109/TNSRE.2023.3300961.

[11] X.-C. Zhong *et al.*, "EEG-DG: A Multi-Source Domain Generalization Framework for Motor Imagery EEG Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 29, no. 4, pp. 2484–2495, Apr. 2025, doi: 10.1109/JBHI.2024.3431230.

[12] H. Song, Q. She, F. Fang, S. Liu, Y. Chen, and Y. Zhang, "Domain generalization through latent distribution exploration for motor imagery EEG classification," *Neurocomputing*, vol. 614, p. 128889, Jan. 2025, doi: 10.1016/j.neucom.2024.128889.

[13] H. Zhao, Q. Zheng, K. Ma, H. Li, and Y. Zheng, "Deep Representation-Based Domain Adaptation for Nonstationary EEG Classification," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 2, pp. 535–545, Feb. 2021, doi: 10.1109/TNNLS.2020.3010780.

[14] P. O. de Paula, T. B. da Silva Costa, R. R. de Faissol Attux, and D. G. Fantinato, "Classification of image encoded SSVEP-based EEG signals using Convolutional Neural Networks," *Expert Syst. Appl.*, vol. 214, p. 119096, Mar. 2023, doi: 10.1016/j.eswa.2022.119096.

[15] Z. Gao *et al.*, "Complex networks and deep learning for EEG signal analysis," *Cogn. Neurodyn.*, vol. 15, no. 3, pp. 369–388, Jun. 2021, doi: 10.1007/s11571-020-09626-1.

[16] X. Du *et al.*, "Electroencephalographic Signal Data Augmentation Based on Improved Generative Adversarial Network," *Brain Sci.*, vol. 14, no. 4, p. 367, Apr. 2024, doi: 10.3390/brainsci14040367.

[17] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative Adversarial Networks-Based Data Augmentation for Brain–Computer Interface," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 9, pp. 4039–4051, Sep. 2021, doi: 10.1109/TNNLS.2020.3016666.

[18] G. Bao *et al.*, "Data Augmentation for EEG-Based Emotion Recognition Using Generative Adversarial Networks," *Front. Comput. Neurosci.*, vol. 15, Dec. 2021, doi: 10.3389/fncom.2021.723843.

[19] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in EEG analysis: an overview," *J. Neuroeng. Rehabil.*, vol. 20, no. 1, p. 40, Apr. 2023, doi: 10.1186/s12984-023-01169-w.

[20] M.-A. Li, J.-F. Han, and L.-J. Duan, "A Novel MI-EEG Imaging With the Location Information of Electrodes," *IEEE Access*, vol. 8, pp. 3197–3211, 2020, doi: 10.1109/ACCESS.2019.2962740.

[21] R. Alam, H. Zhao, A. Goodwin, O. Kavehei, and A. McEwan, "Differences in Power Spectral Densities and Phase Quantities Due to Processing of EEG Signals," *Sensors*, vol. 20, no. 21, p. 6285, Nov. 2020, doi: 10.3390/s20216285.

[22] Y. Zhang, G. Yan, W. Chang, W. Huang, and Y. Yuan, "EEG-based multi-frequency band functional connectivity analysis and the application of spatio-temporal features in emotion recognition," *Biomed. Signal Process. Control*, vol. 79, p. 104157, Jan. 2023, doi: 10.1016/j.bspc.2022.104157.

[23] K. P. Wagh and K. Vasanth, "Performance evaluation of multi-channel electroencephalogram signal (EEG) based time frequency analysis for human emotion recognition," *Biomed. Signal Process. Control*, vol. 78, p. 103966, Sep. 2022, doi: 10.1016/j.bspc.2022.103966.

[24] Wei-Long Zheng and Bao-Liang Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 3, pp. 162–175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[25] D. Wahyu Prabowo *et al.*, "Enhancing EEG-Based Emotion Recognition Using Asymmetric Windowing Recurrence Plots," *IEEE Access*, vol. 12, no. June, pp. 85969–85982, 2024, doi: 10.1109/ACCESS.2024.3409384.

[26] M. Roshanaei, H. Norouzi, J. Onton, S. Makeig, and A. Mohammadi, "EEG-based functional and effective connectivity patterns during emotional episodes using graph theoretical analysis," *Sci. Rep.*, vol. 15, no. 1, p. 2174, Jan. 2025, doi: 10.1038/s41598-025-86040-9.

[27] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.

[28] M. Klug and N. A. Kloosterman, "Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG," *Hum. Brain Mapp.*, vol. 43, no. 9, pp. 2743–2758, Jun. 2022, doi: 10.1002/hbm.25832.

[29] M. Tröndle, T. Popov, A. Pedroni, C. Pfeiffer, Z. Barańczuk-Turska, and N. Langer, "Decomposing age effects in EEG alpha power," *Cortex*, vol. 161, pp. 116–144, Apr. 2023, doi: 10.1016/j.cortex.2023.02.002.

[30] N. Avital, N. Shulkin, and D. Malka, "Automatic Calculation of Average Power in Electroencephalography Signals for Enhanced Detection of Brain Activity and Behavioral Patterns," *Biosensors*, vol. 15, no. 5, p. 314, May 2025, doi: 10.3390/bios15050314.

[31] E. Tan *et al.*, "Theta activity and cognitive functioning: Integrating evidence from resting-state and task-related developmental electroencephalography (EEG) research," *Dev. Cogn. Neurosci.*, vol. 67, p. 101404, Jun. 2024, doi: 10.1016/j.dcn.2024.101404.

[32] G.-H. Shin, Y.-S. Kweon, M. Lee, K.-Y. Jung, and S.-W. Lee, "Quantifying Sleep Quality Through Delta-Beta Coupling Across Sleep and Wakefulness," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 33, pp. 1905–1915, 2025, doi: 10.1109/TNSRE.2025.3569283.

[33] S. Chikhi, N. Matton, and S. Blanchet, "EEG power spectral measures of cognitive workload: A meta-analysis," *Psychophysiology*, vol. 59, no. 6, Jun. 2022, doi: 10.1111/psyp.14009.

[34] S. Sanei and J. A. Chambers, *EEG Signal Processing*. West Sussex, England: John Wiley & Sons Ltd, 2007. doi: 10.1002/9780470511923.

[35] X. Du *et al.*, "An Efficient LSTM Network for Emotion Recognition From Multichannel EEG Signals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1528–1540, Jul. 2022, doi: 10.1109/TAFFC.2020.3013711.

[36] A. Samavat, E. Khalili, B. Ayati, and M. Ayati, "Deep Learning Model With Adaptive Regularization for EEG-Based Emotion Recognition Using Temporal and Frequency Features," *IEEE Access*, vol. 10, pp. 24520–24527, 2022, doi: 10.1109/ACCESS.2022.3155647.

[37] J. Li, S. Li, J. Pan, and F. Wang, "Cross-Subject EEG Emotion Recognition With Self-Organized Graph Neural Network," *Front. Neurosci.*, vol. 15, no. June, pp. 1–10, Jun. 2021, doi: 10.3389/fnins.2021.611653.

[38] L. Farokhah, R. Sarno, and C. Fatichah, "Cross-Subject Channel Selection Using Modified Relief and Simplified CNN-Based Deep Learning for EEG-Based Emotion Recognition," *IEEE Access*, vol. 11, no. September, pp. 110136–110150, 2023, doi: 10.1109/ACCESS.2023.3322294.

[39] M. Aslan, "CNN based efficient approach for emotion recognition," *J. King Saud Univ. - Comput. Inf. Sci.*, Aug. 2021, doi: 10.1016/j.jksuci.2021.08.021.

[40] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, p. 159, Mar. 1977, doi: 10.2307/2529310.

[41] C. Wei, L. Chen, Z. Song, X. Lou, and D. Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomed. Signal Process. Control*, vol. 58, p. 101756, Apr. 2020, doi: 10.1016/j.bspc.2019.101756.

[42] M. Khateeb, S. M. Anwar, and M. Alnowami, "Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset," *IEEE Access*, vol. 9, pp. 12134–12142, 2021, doi: 10.1109/ACCESS.2021.3051281.

[43] D. W. Prabowo, N. A. Setiawan, J. Debayle, and H. A. Nugroho, "Multi-Representation Convolutional Neural Network to Recognize Human Emotion," in *2023 8th International Conference on Information Technology and Digital Applications (ICITDA)*, Nov. 2023, pp. 1–6. doi: 10.1109/ICITDA60835.2023.10427131.

[44] N. A. Setiawan, D. W. Prabowo, and H. A. Nugroho, "Benchmarking of feature selection techniques for coronary artery disease diagnosis," *Proc. - 2014 6th Int. Conf. Inf. Technol. Electr. Eng. Leveraging Res. Technol. Through Univ. Collab. ICITEE 2014*, pp. 1–5, 2014, doi: 10.1109/ICITEED.2014.7007898.

[45] F. Wang, S. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data Augmentation for EEG-Based Emotion Recognition with Deep Convolutional Neural Networks," in *Lecture Notes in Computer Science*, vol. 10705 LNCS, Shenzhen: Springer, Cham, 2018, pp. 82–93. doi: 10.1007/978-3-319-73600-6_8.

[46] Y. Luo and B.-L. Lu, "EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2018, vol. 2018-July, pp. 2535–2538. doi: 10.1109/EMBC.2018.8512865.

[47] Y. Luo, L.-Z. Zhu, and B.-L. Lu, "A GAN-Based Data Augmentation Method for Multimodal Emotion Recognition," in *Advances in Neural Networks – ISNN 2019. ISNN 2019. Lecture Notes in Computer Science*, vol. 11554, Shanghai: Springer, Cham, 2019, pp. 141–150. doi: 10.1007/978-3-030-22796-8_16.

[48] A. Zhang, L. Su, Y. Zhang, Y. Fu, L. Wu, and S. Liang, "EEG data augmentation for emotion recognition with a multiple generator conditional Wasserstein GAN," *Complex Intell. Syst.*, vol. 8, no. 4, pp. 3059–3071, Aug. 2022, doi: 10.1007/s40747-021-00336-7.