

Harnessing Deep Learning for EEG Emotion Recognition: A Hybrid Approach with Attention Mechanisms

Ali H. Abdulwahhab ¹, Alaa Hussein Abdulaal ², Ali M. Jasim ³, Riyam Ali Yassin ⁴, Morteza Valizadeh ⁵, Ahmed Nidham Qasim ⁶, A. F. M. Shahen Shah ⁷, Mehdi Chehel Amirani ⁸

¹ Department of Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey
Email: ahabdulwahhab@gmail.com

² Department of Electrical Engineering, Urmia University, West Azerbaijan
Email: Engineerlaahusseini@gmail.com

³ Department of Space Technology and Non-conventional Energy Sources, National Aerospace University, Kharkiv, Ukraine
Email: drali7819@gmail.com

⁴ Department of Electrical Engineering, Urmia University, West Azerbaijan
Email: RiyamAliYassin@gmail.com

⁵ Department of Electrical Engineering, Urmia University, West Azerbaijan
Email: mo.valizade@urmia.ac.ir

⁶ Department of Electronic and Control Engineering, Northern Technical University, Mosul, Iraq
Email: Ahmednidham23@ntu.edu.iq

⁷ Department of Electronics and Communication Engineering, Yildiz Technical University, Istanbul, Turkey
Email: shah@yildiz.edu.tr

⁸ Department of Electrical Engineering, Urmia University, West Azerbaijan
Email: m.amirani@urmia.ac.ir

Article History

Received: Aug. 04, 2025

Revised: Oct. 13, 2025

Accepted: Oct. 18, 2025

Abstract

Emotion recognition from EEG signals has emerged as a pivotal area of research, driven by its transformative potential in healthcare, brain-computer interfaces, and affective computing systems. However, the intrinsic complexity, non-linearity, and susceptibility to noise in EEG data present significant challenges to accurate emotional state classification. This study proposes a robust and interpretable hybrid deep learning model for EEG-based emotion recognition. The architecture integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and attention mechanisms, together with advanced signal processing techniques such as Continuous Wavelet Transform (CWT) and Power Spectral Density (PSD). This integrated approach facilitates the extraction of comprehensive spatial, temporal, and spectral features from EEG signals, enhancing the model's ability to capture intricate patterns associated with emotional states. Experimental evaluations on the SEED-IV dataset, encompassing four emotional categories—Neutral, Happy, Sad, and Fear—demonstrated the model's exceptional performance, achieving a macro-average F1-score of 93% and an area under the ROC curve (AUC) of 0.94. These results validate the model's effectiveness in accurately distinguishing complex emotional patterns, even under noisy conditions and inter-class ambiguities. Overall, this research advances the domain of EEG-based emotion recognition by introducing a high-performing, interpretable framework suitable for real-world applications while laying the foundation for future developments in adaptive neurofeedback systems and emotion-aware brain-computer interfaces.

Keywords- Attention Mechanisms, Deep Learning, EEG Emotion Recognition, Feature Extraction, Hybrid Model.

I. INTRODUCTION

Electric brain activity measured by EEG provides copious and direct information about the associated neural activity of emotion and, hence, is a far better choice than the other modalities [1]. Due to its high temporal resolution, EEG is especially apt for emotion detection in real time since the rapid changes in brain activity that characterize emotional processing are detectable [2]. Detection and decoding of emotional cues can have revolutionary implications in mental health, where psychological disorders like depression or worry could be detected at the earliest and timely intervention. In brain-computer systems, emotion detection can be used to tailor system attitude in response to the emotional condition of the user, enriching the human-computer interaction in communication devices, gaming, and cognitive enhancement systems [3]. In neurofeedback, emotion monitoring in real-time can be utilized to train the user in regulating emotional reactions, hence achieving improved general well-being.

However, emotion detection accurately from EEG data is a highly demanding task since emotional responses are a nonlinear, highly variable, and subject-specific phenomenon. Various subjects might feel and express their emotion differently, and it might not always be possible to choose common brain activity patterns under a specific emotion in different subjects. Such subject-to-subject variation imposes a tremendous task on the emotion recognition problem. Raw EEG signals, as shown in Figure 1, also gets corrupted by noise and artifacts of various types, say, muscle activity blinks or even environmental noise, which distorts the signal's quality and adversely affects the classifier's performance. Hence, such problems need to be addressed effectively using robust preprocessor and feature extraction techniques so that the proper emotional signals can effectively be captured and analyzed.

In recent years, the combination of deep learning models [4-5] and advanced signal processing techniques has been found to have great potential to enhance emotion detection systems. Methods such as the Continuous Wavelet Transform (CWT) [6] and Short-Time Fourier Transform (STFT) are commonly used to extract time-frequency representations of EEG signals, which are beneficial for capturing temporal and spectral information. These techniques convert the raw EEG data to a representation in which the content at any given instant is emphasized, thereby presenting a more informative brain activity image [7]. Time-frequency representations, through the use of the techniques of CWT and STFT, can present helpful information on the dynamic emotional conditions of subjects, and more accurate emotion detection can be achieved. However, efficient use of the features in enhancing the detection of emotion involves building advanced models that can operate on high-dimensional and complex data. Advanced models such as Convolutional Neural Networks (CNNs) [8] and Recurrent Neural Networks (RNNs) [9] are highly effective in the extraction of meaningful time-frequency representation patterns. However, it is not a simple task in real-world applications to build models that utilize both the temporal and spectral cues and handle individuality and noise simultaneously. Accordingly, advanced hybrid models using many deep models and signal processing techniques are needed to build more robust and accurate emotion detection systems.

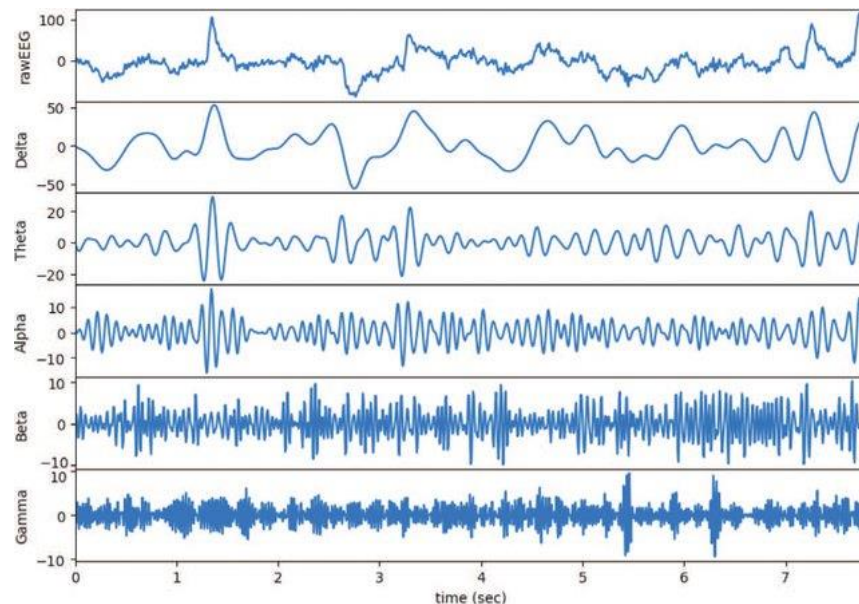


Figure 1. The signal channel raw EEG signal [10]

A novel hybrid architecture composed of Convolutional Neural Networks (CNNs) and an Attention mechanism is presented, using the CWT and the PSD as parallel input to address this task. The CNN component can directly extract spatial cues from the time-frequency representations naturally, and the Attention mechanism allows the classifier to zero in on the most helpful input data parts and, therefore, increase its discriminative emotional cue extraction capabilities. We anticipate the hybrid architecture will utilize the feature-learning capabilities of CNNs and the context-aware capabilities of attention mechanisms to achieve more accurate emotion classification.

II. RELATED WORK

The junction of deep learning methodologies and biomedical signal processing has progressed significantly over the last two years, enabling remarkable improvements in the automatic classification of intricate physiological signals.

Liu et al. published a CNN-BiLSTM hybrid deep learning architecture augmented by DWT-based features for fetal acidosis detection [11]. Their proposed method, integrating spatial feature extraction by the CNN and temporal pattern detection by bi-directional long short-term memory (BiLSTM) networks, showed outstanding diagnostic performances compared to classical models. Integration of DWT in the feature pipeline also contributed significantly towards overcoming the issue of overfitting and enhancing the discriminatory power of the extracted features. This work emphasized the capabilities of hybrid models in efficiently removing the intrinsic coupled nonlinear processes in physiological time series, and they showed robust performances in the context of fetal monitoring.

In 2023, a new theoretical work by Farag applied the DSP viewpoint to explain the traditional convolutional layers of the networks [12]. By conceptualizing the operations of the convolutions in the context of FIR filtering and continuous wavelet transforms (CWT), Farag established the mathematics that guides CNN action. This enhanced the explainability of deep learning models and opened the door to developing self-contained, computation-lean architectures tuned to get deployed in edge devices. It combined classical signal processing paradigms and contemporary deep learning models and thus enhanced the flexibility of CNN-based models in various use domains.

The study's progress was also seen in the research work of Abdulwahhab et al. in 2024, where they presented a fusion of CNN and recurrent neural network (RNN) architecture in detecting epileptic seizures from EEG data [13]. Their two-pathway approach applied both raw sequences of the EEG and time-frequency representation of the sequences by utilizing STFT and CWT transformations. Using the multi-scale feature representation, the model realized high-performing classifier accuracies of more than 99%, indicating the robustness of integrating spectral and temporal analysis in a single framework. The paper emphasized the revolutionary contributions of deep feature fusion techniques in enhancing the validity of the diagnosis of neurological diseases.

Du et al. went one step beyond by introducing a light feature extraction network, the CWT-AM-CNN, for spectral aero-engine infrared categorization [14]. With the utilization of a peak-finding mechanism in the continuous wave transformation domain and self-attention mechanism, the method of Du et al. learned discriminative spectral features of diverse engines in a time-efficient manner. The proposed approach ensured both high accuracy and computation efficiency, and thus, it holds high practicability in real-world industrial scenarios, where quick inference is of top priority.

Similarly, Sun et al. proposed a new multi-feature fusion attentional neural network to recognize epileptic EEG signals [15]. They combined the Hilbert spectrum (HS) analysis and the grayscale recurrence plots (GRP) to effectively capture the time-frequency and the nonlinear dynamic behavior of the EEG signals. The subsequent use of a self-attention mechanism enabled dynamic weighting of diverse sets of features, once more enhancing the expressive power and discriminability of the model. Sun et al.'s findings witnessed the use of multi-domain feature extraction capabilities and advanced attention mechanisms despite heterogeneity and noise in the EEG data.

Masad et al. also examined the stage classification of sleep using a CNN-based method that converted the epochs of the EEG into time-frequency images before classifying them [16]. Their method recorded outstanding levels of accuracy above 99%, supporting the use of convolutional architectures in the task of EEG-based pattern recognition. Of interest was that their method proved robust and reliable across various EEG channels, providing flexibility and robustness for use in clinical sleep medicine scenarios.

Expanding the scope of cardiovascular diagnostics, Berrahou et al. (2025) proposed a hybrid AI arrhythmia forecast model based on Higuchi's fractal dimension analysis, RR-interval properties, and attention-based CNN [17]. This hybrid approach efficiently combined the morphological and temporal properties of the ECG signal in a way that allowed the model to produce high precision in the complex scenario of the inter-patient classification problem. Their study emphasized the removal of class imbalances and dataset heterogeneity, and the derived models generalize and operate effectively in a broad set of clinical situations.

These studies collectively indicate more advanced hybrid models, which integrate the complementary capabilities of CNNs, RNNs, and attention mechanisms. While contemporary models achieve high levels of accuracy under controlled conditions, a series of challenges persist, including narrow generalizability to noisy real-world data and high-complexity architecture computational overheads. Overdependence on preprocessing pipelines can also hamper the solution's scalability. Building on the above insights, research in the current study extends a next-generation hybrid architecture, blending CNN, LSTM, attention mechanisms, and dual CWT and STFT feature extraction streams. By probing the context of adaptive feature fusion and optimization in accuracy and computational cost, the research seeks to bridge present gaps and facilitate the practical applicability of EEG-based classifier systems in emotional state recognition.

III. METHODOLOGY

A. SEED-IV Dataset

SEED-IV is a multimodal dataset established by Shanghai Jiao Tong University's Brain and Cognitive Science Lab (BCMI) to study emotion recognition [18]. It includes 15 subjects and three sessions per subject, 24 trials per session, and 1,080 trials. Trials are evenly distributed over four emotion categories: happy, sad, fearful, and neutral, as shown in Figure. 2. In a trial, subjects viewed movie clips selected to evoke specific emotions. At the same time, a 62-channel ESI NeuroScan System captured the EEG data at a sampling rate

of 1000 and the eye movement data using SMI eye-tracking glasses. SEED-IV is a valuable contribution to developing practical computing and brain-computer interface research.

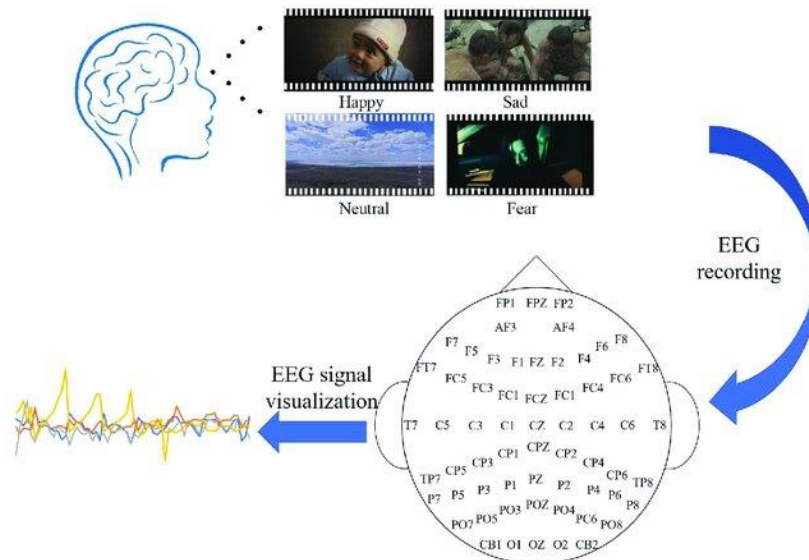


Figure 2. EEG acquisition experimental process [19]

B. Proposed Model

1. Overview

The method proposed in Figure 3 is a deep learning approach incorporating ResNet-18, LSTM, and attention mechanisms to boost feature representation and classification accuracy. It begins with two primary pipelines of feature extraction: ResNet 18-Att, which uses a CNN coupled with an attention mechanism to highlight useful spatial features, and LSTM-SR, which processes ordering data by passing it to the LSTM network. These are then boosted by two specialized modules, i.e., LSTM-SE, which uses a Squeeze-and-Excitation (SE) mechanism to normalize the feature weights, and ResNet18-Att, which boosts spatial features using attention mechanisms. The two pipelines are feature-fused using the Adaptive Feature Fusion method, allowing the model to merge the salient information before final classification.

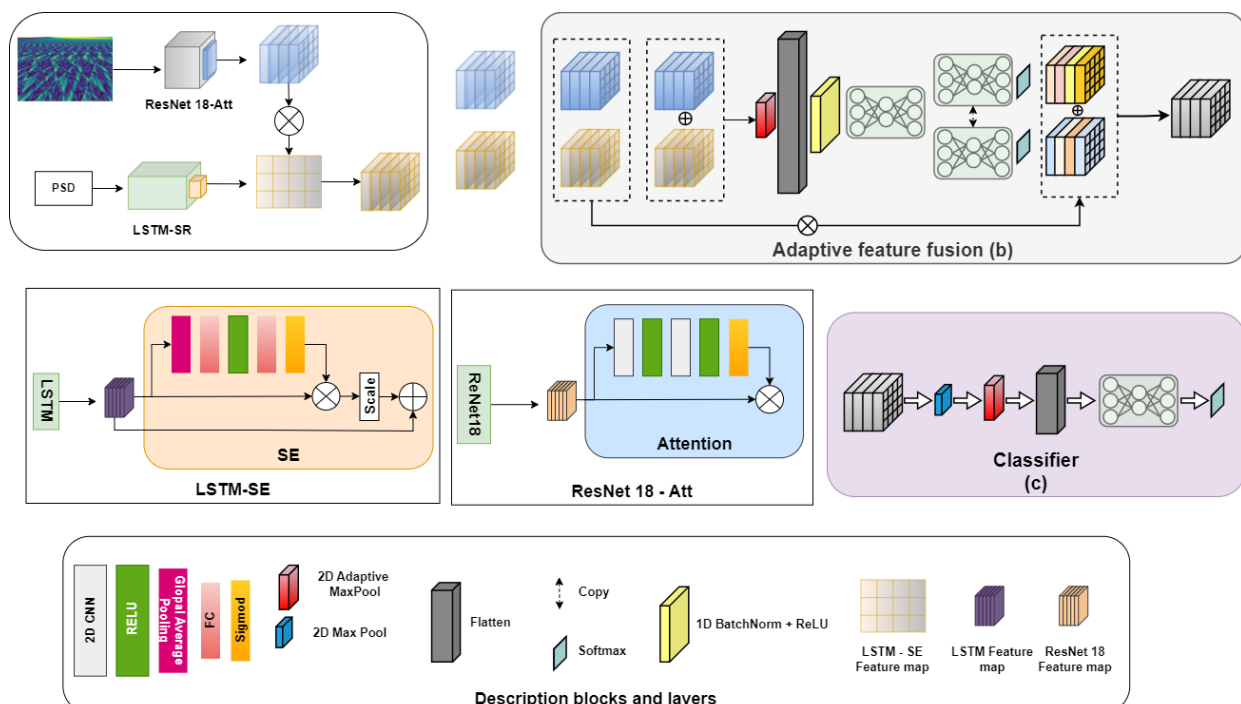


Figure 3. Proposed Model

2. Pre-processing

For pre-processing of the SEED-IV dataset, the raw EEG data of the 3 selected channels, Fp1, Fp2, and Fz, have been used since they are of maximum use in emotional processing. These data were initially band-pass filtered in the 1-75 Hz frequency band to remove unwanted artifacts and noise, preserving the brainwave frequencies responsible for the emotional responses. These trials, which were 240 seconds or 4 minutes, were partitioned into 4 overlapping pieces with 60 seconds each. Such segmentation helps process the brain's activity in shorter durations to detect dynamic changes due to emotional stimuli. Overlapping pieces (50% overlap) will enable temporal continuity and sufficient data points available to detect the features by extraction. After pre-processing, the data will be examined using feature extraction techniques, viz., Continuous Wavelet Transform (CWT) and Power Spectral Density (PSD), to detect spectral-temporal features and power distribution in different frequency bands, respectively.

The four emotion labels in the dataset are Neutral, Happiness, Sadness, and Fear. Each label has 270 trials, each corresponding to an emotional stimulus. There are 3 channels and 4 overlapping parts per trial, so the parts per label are 3,240. In total, the four labels contain 12,960 samples or segments. This pre-processing prepares the dataset for analysis and the subsequent building of the model.

C. Feature extraction

1. Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) is applied to the segmented EEG data to capture spectral and temporal properties [20]. CWT provides a time-frequency representation of the signal wherein the transient and the brain's oscillatory activity can be examined, which is important in emotional response analysis. CWT is applied by convolving the signal using a wavelet function that scales and shifts in the data. Converting the signal's frequency components over time details the evolving frequency bands as they progress along the trial.

CWT is computed in the preprocessing pipeline on each 60-second chunk of the EEG signal. A sliding window approach is used, where a window of one second is slid over the trial chunk, obtaining a total of 60 windows per trial. These windows have 1,000 samples based on the 1,000 Hz per second sampling rate. Morlet wavelet is typically favored, as it offers a good compromise between time and frequency localization. The CWT formula is provided mathematically as:

$$CWT_x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

Where a and b are scale and translation, respectively, $x(t)$ is the EEG signal, and ψ is the mother wavelet

The result of the CWT is the scalogram, a time-frequency representation of a 2D plot with time along the x-axis, frequency or scale along the y-axis, and amplitude of wavelet coefficient as color value. One scalogram result, one per epoch of the time period of 60 sec and of dimension 77×60 (60-time points and 77 frequency bands), is produced, enabling a high-resolution view of the spectral-temporal features of the EEG signal. This scalogram is a visual representation of the EEG signal in the guise of an image and is readily available to use in the subsequent feature extraction and classification analysis.

2. Power Spectral Density

Power Spectral Density (PSD) is a transformation method applied to the segmented EEG data to study the power distribution in the signal across the different bands of frequencies [21]. PSD provides details on the power distribution of the EEG signal in the different frequency components under study, and it helps identify the emotional state or cognitive process in question. PSD, in contrast to the Continuous Wavelet Transform (CWT), which provides time-frequency graphical plots, PSD is more focused on the measurement of the power in the specific bands of frequencies, i.e., delta, theta, alpha, beta, and gamma, commonly applied in the cognitive process and EEG analysis.

The PSD is calculated per 60-second segmented EEG signal during the preprocessing pipeline. One of the steps is to apply the Fourier Transform to every one of the segments to convert the time-domain signal to the frequency domain. This deconstructs the signal to its respective frequencies. Subsequently, the Welch method is typically applied to estimate the power spectral density. This involves partitioning the signal into overlapping sections, using a windowing function (in the present case, the Hamming window), and averaging periodograms to keep the variance low and produce a smooth spectral power estimate.

The formula of PSD is mathematically given as:

$$PSD(f) = \lim_{T \rightarrow \infty} \frac{1}{T} \left| \int_0^T x(t) e^{-j2\pi ft} dt \right|^2 \quad (2)$$

The derived PSD thus presents a 1D representation of the signal's power distribution as a function of the different frequencies. Power distribution is typically plotted across the traditional frequency bands (i.e., delta, theta, alpha, and beta) and can be used to demarcate the brain activity's intensity at different frequencies, typically referring to emotional/affective or cognitive workload states.

Following the computation of the PSD of all segments, the frequency domain features thus extracted are applied in the analysis and the classification. The outcome of the transformation is a vector of power levels in different frequency bands. Power levels are valuable features in the ensuing classification models in detecting emotional states or cognitive processes.

D. ResNet 18

ResNet-18, a deep residual family member, is a strong deep convolutional neural network architecture that is extensively found to work efficiently in the task of hierarchical feature extraction of highly intricate biomedical signals, like in the situation of EEG. ResNet architectures, which were initially put forward by He et al., address the notorious problem of the vanishing gradient standard in deep networks by providing identity-based skip connections, hence facilitating the backpropagation of the gradient easily in deep layers. In the present study, ResNet-18 serves as a building block in the feature extraction module so that the model can learn the discriminative representation of the time-frequency images formed through Continuous Wavelet Transform (CWT).

ResNet-18 comprises 18 weighted layers structured in convolutional blocks alternated by batch normalization and rectified linear unit (ReLU) activations [22]. Convolutional blocks consist of a residual connection that bypasses the transformation layers so the network can learn residual mappings, not direct feature transformations. This addresses the degradation problem of multi-layer networks; whereby additional layers tend to cause the performance to saturate or decrease [23].

For emotion recognition using EEG, ResNet-18 retains several distinct advantages. First, its relatively lightweight framework balances computation efficiency. It does not compromise the depth of abstraction of the features due to the high-dimensional spectrogram and scalogram outputs of the EEG data. Second, the residual connections facilitate the ease of learning low- and high-level abstractions and the detection of minute neurophysiological characteristics of diverse emotional states by the model. Third, pre-training the network using large-scale databases and fine-tuning the EEG dataset yields the benefits of transfer learning, including improved convergence and generalization.

In this work, ResNet-18 was implemented carefully in the presented hybrid architecture, which receives time-frequency representations and yields spatial features necessary for further classification. Subsequently, the feature maps generated by ResNet-18 are fed through attention modules and sequential models such as LSTM layers to capture both the spatial hierarchies and the temporal relations of the EEG data. Integration in this manner ensures that the ResNet-18 module facilitates robust feature learning, general explain ability, and the system's performance.

Besides, extensive empirical experiments verified the efficacy of ResNet-18 in the hybrid system. Its architecture outperformed in discriminative feature extraction of various emotional classes, thereby verifying its core role in the system. Its ability to efficiently handle the input of the CWT validates its versatility for use in different spectral representations, making it a core building block in the proposed emotion recognition pipeline.

E. Attention Module (Att)

Integrating the attention mechanisms into our deep learning architecture further enhanced the ability of the model to capture and utilize EEG signal representations of emotion. Building on the high quality of feature extraction of ResNet18, the proposed Att module intensifies these representations selectively to highlight emotion-relevant features in the EEG signal, as shown in Figure. 4. In our hybrid architecture, the Att module follows directly after the ResNet18 feature extractor, operating on the dense frequential and temporal representations of ResNet18's learned CWT-based scalogram. The figures show that the ResNet18 feature maps are passed directly to the Att module to sharpen its representations.

The self-attention design is replaced by using the convolutional in the attention module. Specifically, the module includes two 2D CNN layers, followed by ReLU activation functions. Using this convolutional architecture, the module learns spatial features in the emotion-relevant feature maps of the input data. These CNN-ReLU layers' outputs are passed through a sigmoid activation function, generating attention weights ranging from 0 to 1 for every element in the feature map. These attention weights are element-wise multiplied by the initial ResNet18 feature map, effectively weighing the different spatial locations in the representation by scaling

their importance. This multiplication step emphasizes the salient areas of the representation and under-emphasizes the less informative areas, generating a more enhanced feature representation that better captures the implicit emotional states in the given EEG data.

This selective weighting is of particular utility in emotion recognition using EEG, in which the discriminative signal features tend to be irregularly distributed both in time and the frequency domain. By weighing informative segments more and minimizing noise and artifacts characteristic of the EEG data, the attention module outputs more discriminative representations. In addition to the gain in performance, the attention module greatly helps interpret the model. By visualizing learned attention weights, we can identify areas of the feature maps that are informative to the predictions and distill informative insights mapping computational outcomes to the neurophysiological models of emotion processing.

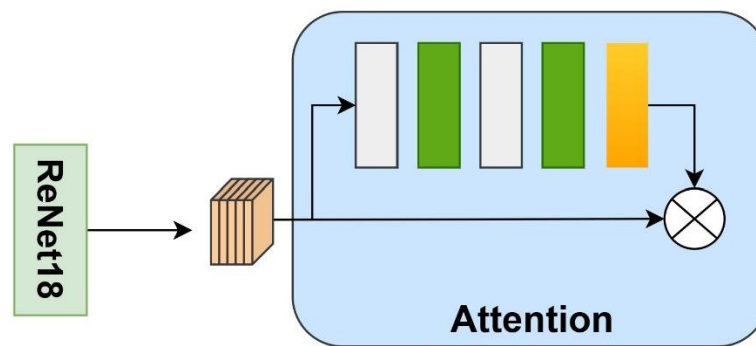


Figure 4. ResNet 18 - Attention Module

F. LSTM

Long Short-Term Memory (LSTM), a variant of the recurrent neural network (RNN), has proved highly effective in modeling temporal relations in sequence data [24]. Initially proposed by Hochreiter and Schmidhuber in response to the exploding and vanishing gradient problems of the traditional RNN, the LSTM networks utilize a novel memory cell design and the use of gating mechanisms enabling the selective forgetting and remembering of data during the progression of long-term time intervals. These attributes have made LSTM highly applicable as a method of processing medical, non-linear, and non-stationary signals, such as the EEG, in which time dynamics are critical in distinguishing different emotional or cognitive states.

In the proposed hybrid architecture, the LSTM module plays the base temporal feature extractor role to identify the dynamic patterns in the sequential EEG data. After the ResNet-18 module extracts the spatial features of the time-frequency representations, the spatial features are reshaped to temporal sequences and fed to the LSTM layer. Due to the temporal modeling ability of LSTM, the network is in a position to capture the nuanced temporal relationships typical of transient emotional expressions in the EEG signal.

Structurally, there are three primary gates in the LSTM module: the input gate, the forget gate, and the output gate. The input gate regulates the flow of new data into the memory cell, the forget gate determines data to discard in the cell state, and the output gate regulates the data flow to the subsequent layers. This gating provides the LSTM the ability to weigh the preservation of useful past information and learning new temporal information and also aids the generalizing ability of the model in different time scales.

Furthermore, the use of LSTM in this context is not restricted to a one-way architecture. To capture a more comprehensive temporal context, the architecture employs a bidirectional LSTM (BiLSTM) setup, in which the sequence data is fed in both the forward and the backward directions. This double-pronged method ensures that the network will capture past and future relations in the EEG signal, portraying a more accurate picture of the temporal processes underpinning emotional state.

Briefly, the LSTM module deepens the temporal dimension of the feature representations and narrows the gap between static spatial features and dynamic temporal patterns. Its integration into the hybrid system assures the system's handling of complex EEG data, resulting in more accurate and reliable emotion recognition results.

G. Sequence-extraction Module (SE)

The Sequence-extraction Module (SE) is one of the primary elements of our proposed hybrid architecture [25]. It is specifically tailored to refine the model's ability to extract and structure sequential dependencies in the EEG data. Because the EEG data are time series, and emotional processes emerge dynamically in time, effective extraction and representation of sequential dependencies are of the highest priority to ensure accurate and reliable emotion detection.

As indicated in the architecture illustration, the SE module receives the input features of the LSTM network and processes them in a thoughtfully structured method, as shown in Figure. 5. The module involves a highly advanced processing pipeline composed of five significant components operating in a stepwise manner:

Global Adaptive Pooling: The initial block (in magenta in the figure above) performs global adaptive pooling on the LSTM outputs, compressing temporal data while preserving the primary details along the time axis. It effectively reduces the dimensionality while maintaining the necessary emotional content of the signal.

FC Layer: Following the pooling, the features pass through a fully connected layer (light pink) that transforms the pooled features into a shape suitable for attention weighting.

ReLU Activation: In the third block (green), we apply the Rectified Linear Unit (ReLU) activation, which provides the flow of processing with a degree of non-linearity and makes the system more competent in learning complicated data patterns.

Second FC Layer: This second fully connected layer (light yellow) further processes the feature representation and transforms it into a space to compute the attention weights.

Sigmoid Activation: The final block (yellow) employs the sigmoid activation function to map the outputs to attention weights between 0 and 1. These weights indicate the importance of the different features in the original LSTM's output.

Attention is implemented by the multiplication operation ($\otimes$) in the above figure, in which the sigmoid-activated weights are element-wise multiplied by the initial LSTM features. This weighting enables the model to highlight the EEG signal's more informative regions and ignore the signal's redundant parts.

The SE module contains a "Scale" component following attention weighting, which scales the attention-weighted features. It finishes by summing the scaled attention-weighted and input features using an element-wise addition operation ($\oplus$). It uses a residual connection to enable the info flow and the backdrop during the training.

Empirical experiments in the present study revealed that including the Sequence-extraction Module contributed considerably to enhancing the hybrid model's performance. Experiments in comparative settings revealed models utilizing SE performing better than models not using the module under conditions where emotional alterations are subtle or in the presence of noisy signal conditions. The module's ability to extract and highlight time-related features was crucial in achieving robust classification outcomes across various emotional states.

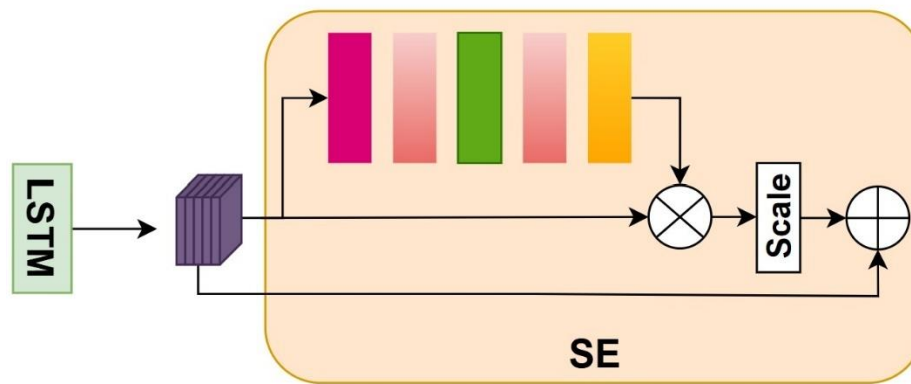


Figure 5. LSTM-Sequence-extraction Module

In addition, the SE module makes the system more interpretable by defining the time intervals most influential in emotion recognition. Analyzing the attention weights generated by the sigmoid activation makes it easier to understand the temporal patterns given more attention, revealing the emotional processing processes modeled by the model.

In summary, the Sequence-extraction Module is a crucial component of the temporal processing pipeline of the proposed framework. By its sequential Global Adaptive Pooling, FC layers, and ReLU and Sigmoid activations, the SE module achieves a highly effective attention mechanism that allows the model to extract the most informative temporal features of the EEG signals to detect emotions. Such a capability is essential in achieving enhanced accuracy, robustness, and explain ability of the EEG-based emotion recognition systems and moving the model towards practical use in affective computing systems.

IV. RESULTS

The hybrid model's performance is extensively displayed using three pertinent performance measures: the classification report table, confusion matrix, ROC curve, and AUC measure. These measures present essential information about the model's potential for classifying emotional states as Natural, Sad, Fear, and Happy with high precision and reliability.

The performance of the proposed hybrid model is presented in Table I and Figure. 6. The four emotion classes exhibit robust and uniform performance in the classification table. In particular, Natural enjoys the highest precision of 95% and recall of 93%, for an F1-score of 94%, demonstrating the model being highly consistent and confident in classifying the natural expressions. Sad ranks a close second at a precision of 94%, recall of 93%, and an F1-score of 94%, demonstrating that the model is somewhat less precise in identifying sadness but is consistent in the same measure. Fear and Happiness register precision and recall of approximately 91–93%,

and the F1-scores register at 92%, demonstrating somewhat more variation in the classification than the other sentiments. Overall, when the average of the entire set of classes is calculated, the macro-average precision is at 93%, the recall is at 92.5%, and the F1-score is at 93%, demonstrating a stable and generalized performance in the absence of any skewed tendencies toward a specific emotion.

These indicate the accuracy and balance at which the model classifies correct and incorrect positives and negatives—important in emotion recognition experiments in which misclassifications give rise to mistyping of affective states.

TABLE I. EVALUATION MEASURES

labels	Precision	Recall	F1-score	Accuracy
Natural	95%	93%	94%	93%
Sad	94%	93%	94%	
Fear	92%	92%	92%	
Happy	91%	93%	92%	
All	93%	92.5%	93%	

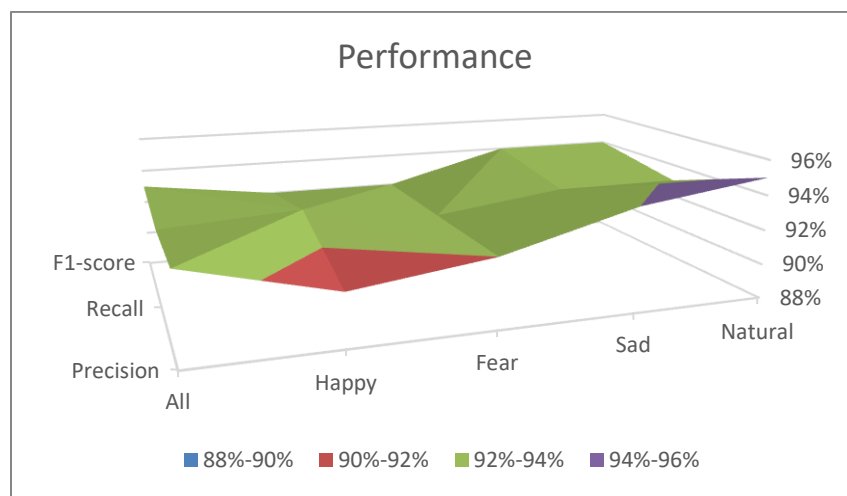


Figure 6. Performance analysis

Figure 7 Confusion matrix provides visibility of the per-class distribution of the classification results. It shows good diagonal dominance, where 298 Natural, 301 Sad, 300 Fear, and 302 Happy are classified correctly out of 324 per class. Misclassifications are minimal and typically occur amongst similar or next-in-category emotion types. Natural, for example, is likely to be misclassified as Sad (15 errors), perhaps due to the similarities in facial or physiological cues. Fear is sometimes mistaken as Sad or Neutral (10 errors in both cases), perhaps reflecting the ambiguity of the emotional cues in the real-world situation. Happy, in all its relative discriminability, is misclassified in 8–9 errors in other categories—still a narrow margin in the context of the overall correctness of the system. This matrix provides not only the statistical high-performing status of the system but also its correct intuition of emotional similarities, and hence, it is sufficient for use in real-time emotional analysis in real-world scenarios.

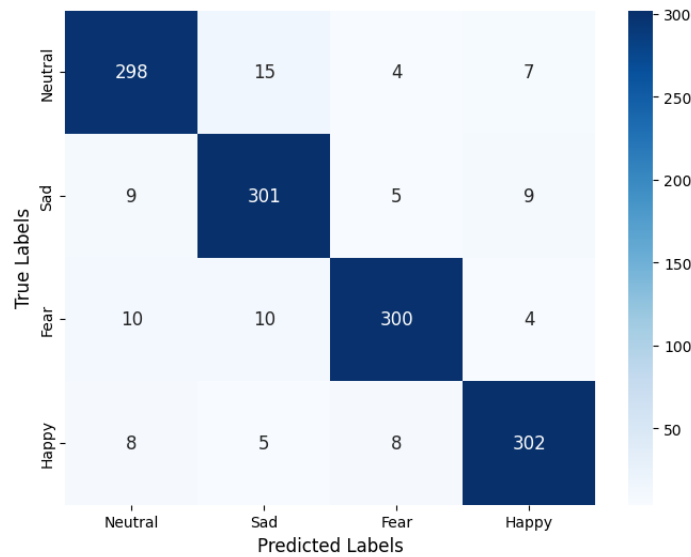


Figure 7. Confusion matrix

Figure 8 presents the AUC of the hybrid model. The hybrid model's ROC curve is indicated by a high AUC value of 0.94, demonstrating the high discriminatory power of the model. An AUC measures the model's discriminatory power in separating the two classes—values close to 1 indicate high performance. An AUC of 0.94 implies the ability of the model to rank the positive instances above the negative instances consistently, regardless of the emotion class, as needed in use scenarios where the decision boundary might vary. This high AUC also reflects the model's resilience under scenarios of imbalanced data or different noise levels, common in affective computing paradigms (e.g., EEG or image-based emotion detection).

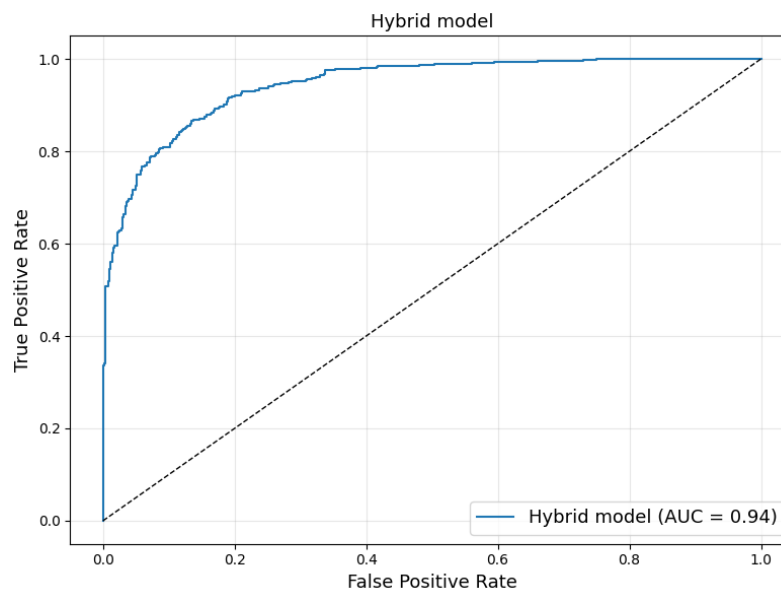


Figure 8. ROC of the hybrid model

V. CONCLUSION

This study presents a novel hybrid deep learning framework for EEG-based emotion recognition, which integrates ResNet-18, LSTM, and attention mechanisms, complemented by parallel feature extraction pathways utilizing Continuous Wavelet Transform (CWT) and Power Spectral Density (PSD). By systematically combining spatial feature extraction through ResNet-18, temporal dynamics modeling with LSTM, and selective emphasis on salient information via attention modules, the architecture captures the complex, non-stationary patterns inherent in EEG signals related to emotional states. The incorporation of the Sequence-extraction Module (SE) further refines temporal dependencies, allowing for accurate classification of transient and evolving emotional cues.

Empirical results from the SEED-IV dataset demonstrate the efficacy of the proposed model, achieving high precision, recall, and F1 Scores across four emotional categories: Natural, Sad, Fear, and Happy. The hybrid model reached an overall macro-average F1 Score of 93% and an AUC of 0.94, indicating robust generalization even amidst inter-class ambiguities and signal noise. Validation through the confusion matrix revealed minimal misclassifications and high accuracy across all emotional classes.

In addition to quantitative performance, the architecture provides significant interpretability benefits. Visualization of attention weights and sequence extraction activations offers insights into the temporal and spectral regions of EEG signals most indicative of emotional states. This interpretability enhances model transparency and aligns computational findings with neuroscientific theories of emotion processing, fostering confidence in real-world applications.

While substantial advancements have been achieved, opportunities for future exploration remain. Expanding the model to incorporate multi-modal inputs, such as facial expressions or physiological signals, could enhance emotion recognition accuracy. Additionally, exploring domain adaptation techniques would improve generalizability across diverse populations and recording conditions, which would be instrumental for clinical applications. Optimizing the model for real-time processing also presents the potential for deployment in practical brain-computer interface systems and affective computing platforms.

In conclusion, the proposed hybrid model significantly advances EEG-based emotion recognition, offering a robust, interpretable, and high-performing solution. Its effective integration of multi-scale feature extraction and advanced deep learning techniques makes it a promising candidate for future research and applications in affective computing and neurotechnology.

REFERENCES

- [1] R. Ranjan, B. Chandra Sahana, and A. Kumar Bhandari, "Deep Learning Models for Diagnosis of Schizophrenia Using EEG Signals: Emerging Trends, Challenges, and Prospects," *Archives of Computational Methods in Engineering*, vol. 31, no. 4, pp. 2345–2384, Jan. 2024, doi: <https://doi.org/10.1007/s11831-023-10047-6>.
- [2] A. H. Abdulwahhab, Indrit Myderrizi, and Muhammet Mustafa Yurdakul, "PAFWF-EEGC Net: parallel adaptive feature weight fusion based on EEG-dynamic characteristics using channels neural network for driver drowsiness detection," *Signal Image and Video Processing*, vol. 19, no. 7, May 2025, doi: <https://doi.org/10.1007/s11760-025-04102-x>.
- [3] A. H. Abdulwahhab, Indrit Myderrizi, and M. K. Mahmood, "Drone Movement Control by Electroencephalography Signals Based on BCI System," *Advances in Electrical and Electronic Engineering*, vol. 20, no. 2, Jun. 2022, doi: <https://doi.org/10.15598/aeec.v20i2.4413>.
- [4] A. H. Abdulaal, M. Valizadeh, M. C. Amirani, and A. F. M. Shahan Shah, "A self-learning Deep Neural Network for Classification of Breast Histopathological Images," *Biomedical Signal Processing and Control*, vol. 87, no. B, p. 105418, Jan. 2024, doi: <https://doi.org/10.1016/j.bspc.2023.105418>.
- [5] A. H. Abdulaal et al., "Unsupervised Histopathological Sub-Image Analysis for Breast Cancer Diagnosis Using Variational Autoencoders, Clustering, and Supervised Learning," *Journal of Engineering and Sustainable Development*, vol. 28, no. 6, pp. 729–744, Nov. 2024, doi: <https://doi.org/10.31272/jeasd.28.6.6>.
- [6] H. S. Mansour, M. Valizadeh, A. H. Abdulaal, and M. C. Amirani, "Enhanced Electrocardiogram Arrhythmia Diagnosis with Deep Learning and Selective Attention Mechanism," *Advances in Technology Innovation*, 2025, doi: <https://doi.org/10.46604/aiti.2024.14034>.
- [7] C. J. O. Echeverri, S. Salazar-Colores, and G. Hernández-Nava, "A Study of the Relationship of Wavelet Transform Parameters and Their Impact on EEG Classification Performance," *Brain-Computer Interfaces*, pp. 115–130, Apr. 2025, doi: <https://doi.org/10.1016/B978-0-323-95439-6.00012-0>.
- [8] A. H. Abdulaal et al., "Deep Learning-based Signal Identification in Wireless Communication Systems: a Comparative Analysis on 3G, LTE, and 5G Standards," *Al-Iraqia Journal for Scientific Engineering Research*, vol. 3, no. 3, pp. 60–70, 2024, doi: <https://doi.org/10.58564/IJSER.3.3.2024.224>.
- [9] A. H. Abdulaal et al., "Hybrid CNN and RNN Model for Histopathological Sub-Image Classification in Breast Cancer Analysis Using Self-Learning," *Journal of Engineering and Sustainable Development*, vol. 29, no. 3, pp. 310–320, May 2025, doi: <https://doi.org/10.31272/jeasd.2746>.
- [10] N. Bajaj, "Wavelets for EEG Analysis," *Wavelet Theory*, Feb. 2021, doi: <https://doi.org/10.5772/intechopen.94398>.
- [11] M. Liu, Y. Lu, S. Long, J. Bai, and W. Lian, "An attention-based CNN-BiLSTM Hybrid Neural Network Enhanced with Features of Discrete Wavelet Transformation for Fetal Acidosis Classification," *Expert Systems with Applications*, vol. 186, pp. 115714–115714, Dec. 2021, doi: <https://doi.org/10.1016/j.eswa.2021.115714>.
- [12] M. M. Farag, "Design and Analysis of Convolutional Neural Layers: a Signal Processing Perspective," *IEEE Access*, vol. 11, pp. 27641–27661, 2023, doi: <https://doi.org/10.1109/access.2023.3258399>.
- [13] A. H. Abdulwahhab, A. H. Abdulaal, Assad, A. A. Mohammed, and M. Valizadeh, "Detection of Epileptic Seizure Using EEG Signals Analysis Based on Deep Learning Techniques," *Chaos, Solitons & fractals/Chaos, Solitons and Fractals*, vol. 181, pp. 114700–114700, Apr. 2024,

- doi: <https://doi.org/10.1016/j.chaos.2024.114700>.
- [14] S. Du, W. Han, Z. Kang, X. Lu, Y. Liao, and Z. Li, "Continuous Wavelet Transform Peak-Seeking Attention Mechanism Conventional Neural Network: a Lightweight Feature Extraction Network with Attention Mechanism Based on the Continuous Wave Transform Peak-Seeking Method for Aero-Engine Hot Jet Fourier Transform Infrared Classification," *Remote Sensing*, vol. 16, no. 16, p. 3097, Aug. 2024,
doi: <https://doi.org/10.3390/rs16163097>.
- [15] C. Sun, C. Xu, H. Li, H. Bo, L. Ma, and H. Li, "A Novel multi-feature Fusion Attention Neural Network for the Recognition of Epileptic EEG Signals," *Frontiers in Computational Neuroscience*, vol. 18, Jun. 2024,
doi: <https://doi.org/10.3389/fncom.2024.1393122>.
- [16] I. S. Masad, A. Alqudah, and S. Qazan, "Automatic Classification of Sleep Stages Using EEG Signals and Convolutional Neural Networks," *PLOS ONE*, vol. 19, no. 1, pp. e0297582–e0297582, Jan. 2024,
doi: <https://doi.org/10.1371/journal.pone.0297582>.
- [17] N. Berrahou, A. E. Alami, R. E. Alami, and H. Qjidaa, "Synergistic Approaches for Accurate Arrhythmia Prediction: a Hybrid AI Model Integrating Higuchi Dimensional Fractal, RR-intervals and Attention-based Convolutional Neural Network in ECG Signal Analysis," *Statistics Optimization & Information Computing*, vol. 13, no. 2, pp. 547–567, Mar. 2024,
doi: <https://doi.org/10.19139/soic-2310-5070-2091>.
- [18] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: a Multimodal Framework for Recognizing Human Emotions," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019,
doi: <https://doi.org/10.1109/tcyb.2018.2797176>.
- [19] Z. Huang, Y. Ma, R. Wang, W. Li, and Y. Dai, "A Model for EEG-Based Emotion Recognition: CNN-Bi-LSTM with Attention Mechanism," *Electronics*, vol. 12, no. 14, pp. 3188–3188, Jul. 2023,
doi: <https://doi.org/10.3390/electronics12143188>.
- [20] M. R. Khan, A. A. Tania, and M. Ahmad, "A Comparative Study of Time-frequency Features Based spatio-temporal Analysis with Varying Multiscale Kernels for Emotion Recognition from EEG," *Biomedical Signal Processing and Control*, vol. 107, p. 107826, Sep. 2025,
doi: <https://doi.org/10.1016/j.bspc.2025.107826>.
- [21] K. F. Walters et al., "Resting-State EEG Power Spectral Density Analysis between Healthy and Cognitively Impaired Subjects," *Brain Sciences*, vol. 15, no. 2, p. 173, Feb. 2025,
doi: <https://doi.org/10.3390/brainsci15020173>.
- [22] A. H. Abdulaal et al., "Cutting-edge CNN Approaches for Breast Histopathological Classification: the Impact of Spatial Attention Mechanisms," *ShodhAI: Journal of Artificial Intelligence*, vol. 1, no. 1, Oct. 2024,
doi: <https://doi.org/10.29121/shodhai.v1.i1.2024.14>.
- [23] A. H. Abdulwahhab, O. Bayat, and A. A. Ibrahim, "HAFMAB-Net: hierarchical adaptive fusion based on multilevel attention-enhanced bottleneck neural network for breast histopathological cancer classification," *Signal Image and Video Processing*, vol. 19, no. 5, Mar. 2025,
doi: <https://doi.org/10.1007/s11760-025-04001-1>.
- [24] F. Rivas, J. Enrique Sierra-Garcia, and J. M. Camara, "Comparison of LSTM- and GRU-Type RNN Networks for Attention and Meditation Prediction on Raw EEG Data from Low-Cost Headsets," *Electronics*, vol. 14, no. 4, pp. 707–707, Feb. 2025,
doi: <https://doi.org/10.3390/electronics14040707>.
- [25] M. A. Mulkey, H. Huang, T. Albanese, S. Kim, and B. Yang, "Supervised Deep Learning with Vision Transformer Predicts Delirium Using Limited Lead EEG," *Scientific Reports*, vol. 13, no. 1, p. 7890, May 2023,
doi: <https://doi.org/10.1038/s41598-023-35004-y>.