RESEARCH PAPER
*Future Research in AI & IoT, 2026, 2(2)*
*Homepage: https://frai.journals.umz.ac.ir/*
*DOI: 10.22080/frai.2026.30818.1040*

# Five-channel EEG-based emotion recognition using CNN and LSTM networks

**Mahsa Akhbari[1], Amirhossein Pournorouz[2]**

[1]*Department of Biomedical Engineering, SR. C., Islamic Azad University, Tehran, Iran*

[2]*Department of Biomedical Engineering, SR. C., Islamic Azad University, Tehran, Iran*

**Article Info**
Received 16 Dec 2026
Accepted 05 Jun 2026
Available online 05 Jun 2026

**Keywords:**

Emotion Recognition; Electroencephalogram (EEG); Deep Learning; Convolutional Neural Network (CNN); Long Short-Term Memory (LSTM); Time-Series.

**Abstract:**

This study presents an emotion recognition approach using five-channel electroencephalogram (EEG) signals analyzed through convolutional neural networks (CNN) and long short-term memory (LSTM) networks. The goal is to achieve robust binary classification (positive vs. negative emotion) using minimal electrode input to reduce computational cost. We employed five electrodes positioned in the prefrontal and temporal regions of the brain to reduce computational complexity and enhance classification accuracy. Using the SEED dataset (SJTU Emotion EEG Dataset), we applied continuous wavelet transform (CWT) to generate 2-D images for the CNN model and wavelet packet transform (WPT) for gamma band extraction in the LSTM model. Our results show an average classification accuracy of 94.57±0.06% for CNN and 74.6% for LSTM. We emphasize that, despite using basic architectures, effective performance was achieved with minimal electrodes. Training time for the CNN model was 3 minutes, and for the LSTM model was 31 minutes, which indicates the robustness of this methodology. This lightweight system has practical applications in emotion-aware technologies, mental health tools, and human-computer interaction.



***Corresponding Author:** akhbari.mahsa@iau.ac.ir

**Supplementary information:** Supplementary information for this article is available at https://frai.journals.umz.ac.ir/

**Please cite this paper as:**

## 1. Introduction

Over the past several decades, emotion recognition has garnered significant attention from researchers due to its broad applicability and potential benefits. Human emotion is a complex phenomenon involving a cascade of physical and chemical processes within the body, particularly in the brain [1]. Physiologists typically employ two primary models to characterize emotions: the discrete (basic) emotion model and the dimensional model [2]. Discrete emotions, which were interpreted by Tomkins [2] comprise nine basic emotions: interest-excitement, surprise-startle, enjoyment-joy, distress-anguish, dismal, fear-terror, anger-rage, contempt-disgust, and shame-humiliation. It is believed that these nine basic emotions play an important role in optimal mental health [2]. Due to the inherent complexities of affective states, some researchers argue that the discrete model has limitations. Consequently, dimensional models, such as Russell's 2D emotional model (Figure 1), are more frequently used. This model delineates four major regions: Region 1 encompasses high arousal positive valence (HAPV) emotions, ranging from feelings of pleasure to excitement; Region 2 covers high arousal negative valence (HANV) emotions, extending from nervousness to annoyance; Region 3 involves low arousal negative valence (LANV) emotions; and Region 4 includes low arousal positive valence (LAPV) emotions, such as calmness and relaxation [2].

Physiological signals, which are responses to external stimuli originating from the central and autonomic nervous systems (CNS and ANS), offer valuable insights into the underlying mechanisms of emotion [3]. Among these signals, brain signals are particularly useful for emotion recognition, as they provide a direct window into brain activity. These signals contain distinct features that can be analyzed in both the time and frequency domains.



*ISSN 3092-7552*

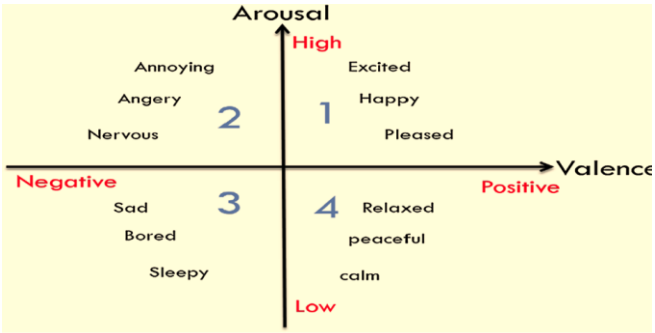


**Figure 1. Russell's 2D emotional model [2], containing four major regions. Region 1: high arousal positive valence (HAPV) emotions. Region 2: high arousal negative valence (HANV) emotions. Region 3: low arousal negative valence (LANV), and Region 4: low arousal positive valence (LAPV) emotions.**

By extracting these features and applying deep neural networks, brain-computer interfaces (BCIs) can engage more intelligently with users, taking their emotional states into account [3].

Traditionally, affective computing [4] has relied on more conventional modalities such as facial expressions, vocal intonations, and body gestures to infer physical and emotional states. For instance, some researchers [5] have used dynamic features like Electrocardiogram (ECG), Electrodermal Activity (EDA), and Electromyography (EMG), and eye-blink signals for fatigue detection. Some researchers use speech spectrograms as an input feature for speech emotion recognition. These 2-D representations of speech signals are very suitable for extracting important features of speech signals [6]. Facial expressions, analyzed through computer vision techniques, provide an intuitive and non-invasive method for detecting emotions. For instance, specific facial action units are linked to different emotional expressions, such as smiles for happiness or frowns for sadness. Similarly, audio signals, including pitch, tone, and speech rhythm, have been extensively used to classify emotional states. Also, EEG has been used for the classification of the BCI multi-class motor imagery task [7] and epileptic seizure detection [8]. While these methods have shown considerable success in controlled environments, they are often limited by variations in lighting, occlusions, or cultural differences in facial expressions, as well as background noise in audio recordings. Consequently, there has been a growing interest in using more direct physiological signals, such as EEG, to overcome these challenges and enhance the robustness of emotion recognition systems [8].

Various methods have been proposed to monitor brain activity during emotional episodes, with the electroencephalogram (EEG) being one of the most commonly employed techniques in brain-based emotion recognition due to its effectiveness, low-cost setup, and high temporal resolution [8]. Several approaches have been developed for classifying emotions using EEG signals across time, frequency, and time-frequency domains. Research has shown that analyzing EEG signals in the time-frequency domain is particularly effective for capturing dynamic changes in neural spectra. For instance, Zhang et al. [9] utilized time-frequency techniques to distinguish between genuine and fake emotional expressions, employing k-nearest neighbor (KNN), support vector machine (SVM), and artificial neural network (ANN) algorithms, with ANN and KNN classifiers yielding the best results. Lou et al. [10] implemented a spiking neural network to classify emotions based on features extracted using discrete wavelet transform (DWT), variance, and Fast Fourier Transform (FFT), finding that variance features were more effective for processing raw EEG data.

Bagherzadeh et al. [11] employed 2 effective connectivity measures for recognizing five emotional states, utilizing pre-trained deep learning algorithms, including ResNet-50, which achieved the highest accuracy among other previous works. During emotional episodes, specific brain lobes exhibit heightened activity relative to other regions. Zheng et al. [12] demonstrated that the lateral temporal areas show increased activation for positive emotions compared to negative ones, particularly in the beta and gamma sub-bands. While many studies [13-15] have highlighted the robustness of convolutional neural networks (CNNs) in automatically extracting features for emotion recognition, whether using pretrained networks or constructing novel network structures, some research, such as that by Zhang et al. [17], has explored hierarchical feature extraction. They constructed various network hierarchies with CNNs to classify emotional states more effectively. In their work, both EEG signals and Peripheral Physiological signals (PPS) of the DEAP dataset are used, and features of both signals are extracted. Hwang et al. [13] proposed a method to estimate 3 class emotional states (positive, negative, neutral) from the 62 channels of the EEG SEED dataset. Their methods consist of two steps. In the first step, topology-preserving differential entropy features are generated. And in the second step, a CNN network is used for classification. Yang et al. [14] took the constructed 3D EEG cube as input and used a continuous CNN for emotion prediction from the DEAP dataset.

Mahmoud et al. [16] Leverage CNNs to learn discriminative cues directly from raw EEG signals, simplifying the preprocessing pipeline and extracting more informative features. They evaluate their approach on two benchmark emotion datasets, DEAP and SEED. The study reports state-of-the-art performance, achieving accuracies of 96.32% on the SEED dataset and 92.54% on the DEAP dataset.

In this work, we employed five electrodes positioned in the prefrontal and temporal regions of the brain to reduce computational complexity and enhance classification accuracy. In the proposed framework, continuous wavelet transform (CWT) was utilized as a time-frequency analysis technique on EEG signals to generate 2-D images for CNN-based feature extraction, while wavelet packet transform was applied for gamma band analysis using the Long Short-Term Memory (LSTM) network.

The paper is organized as follows: Database, methods, and proposed networks are described in Section 2. Experiments and results are comprehensively illustrated in Section 3, and finally, the conclusion and discussion are handled in Section 4.

## 2. Methods and materials

### 2.1. Database

EEG data from nearby time points have certain correlations. On the other hand, EEG activity may undergo coherent subtle changes over time. Therefore, using convolutional neural networks (CNNs) or recurrent neural networks (RNNs) such as long short-term memory (LSTM) is a reasonable choice [17]. In this study, we utilized the SJTU[1] Emotion EEG Dataset (SEED[2]), which is a public database provided by the BCMI laboratory [18]. This database provides EEG signals collected from 15 participants while they watched 15 different video clips, each designed to evoke a distinct emotional response: positive, neutral, or negative. Each participant completed the experiment three times over the course of three weeks, with a one-week interval between sessions. EEG signals were recorded using a 62-channel device at a sampling rate of 1000 Hz. For the preprocessed data, all signals were downsampled to 200 Hz, and a bandpass filter with a range of 0–75 Hz was applied. The dataset for each participant comprises 16 arrays: 15 arrays containing the preprocessed signals and one label array, where labels are represented as 1 (positive), -1 (negative), and 0 (neutral).

Our proposed methodology focuses on binary classification of emotions using a significantly reduced number of electrodes. As previously mentioned, the dataset includes 45 sets of EEG signals (15 participants × 3 experiments) in both raw and preprocessed forms. To ensure a fair comparison with existing studies, we used the preprocessed data. For each participant, we excluded the EEG signals associated with neutral emotions because we want to do binary classification. We manually selected 10 EEG recordings corresponding to positive and negative emotions based on the labels provided in the dataset. These selected signals were used for both CNN and LSTM classifications.

Since it has been proved that the prefrontal and temporal lobes of the brain have a strong correlation with positive and negative emotional states [19-21], we utilized the following 5 electrodes: FP1, FPz, FP2, T7, and T8 for the classification task.

The block diagram of the proposed methodology is depicted in Figure 2. In the proposed framework, continuous wavelet transform (CWT) was utilized as a time-frequency analysis technique on EEG signals to generate 2-D images for CNN-based feature extraction, while wavelet packet transform was applied for gamma band analysis using the LSTM network.

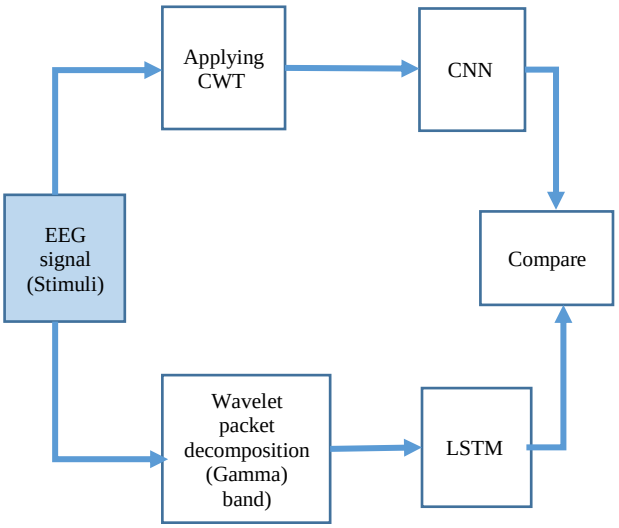


**Figure 2. The block diagram of the proposed methodology: 1) EEG acquisition as external stimuli, 2) Processing the pre-processed EEG data using the CWT and WPD to generate distinct representations (image and time-series representations), 3) Training CNN (with image representation as input data) and LSTM (with time-series representation as input data) networks with processed data. 4) Evaluating the performance of both models to determine which architecture yields superior results.**

### 2.2. Wavelet transform

As previously noted, information loss during the processing and feature extraction of EEG signals is often inevitable. One of the primary advantages of time-frequency analysis is its ability to preserve both temporal

[1] Shanghai Jiao Tong University

[2] Brain-like Computing and Machine Intelligence (https://bcmi.sjtu.edu.cn/home/seed/)

and frequency domain features. This dual preservation capability allows for accurate time resolution at high frequencies and effective frequency resolution at low frequencies [2].

The Continuous Wavelet Transform (CWT) was employed to convert the 1-D EEG time-series signals into 2-D time-frequency images, which served as the input for our Convolutional Neural Network (CNN). This transformation allows the CNN to leverage its spatial feature extraction capabilities by treating the time-frequency representations as image data. The continuous wavelet transform (CWT) of a given signal $x(t)$ is defined as the integral function of $x(t)$ with a family of wavelet functions $\psi_{a,b}(t)$ [22]:

$$\mathrm{CWT(a,b)} = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}\psi_{ab}\left(\frac{t-b}{a}\right)\cdot x(t)dt \quad (1)$$

The CWT is defined as the sum of the signal multiplied by scaled and shifted versions of the wavelet function $\psi$ [22].

$$CWT(scale, position) = \int_{-\infty}^{\infty} x(t) \cdot \psi(scale, position, t)dt \quad (2)$$

The function $\psi(t)$ is known as the mother wavelet, and the components of functions $\psi_{a,b}(t)$ are called daughter wavelets. The daughter wavelets are simply obtained by scaling and shifting the mother wavelet. The scale factor *a* represents the scaling of the function $\psi(t)$, while the shift factor *b* represents the temporal translation of the function.

For the Long Short-Term Memory (LSTM) network, we used Wavelet Packet Decomposition (WPD) to extract specific frequency bands, particularly the gamma band (30-75 Hz), from the EEG signals. WPD is a generalization of the Discrete Wavelet Transform (DWT) that provides a richer analysis by decomposing both the approximation and detail coefficients at each level, allowing for a more precise isolation of desired frequency components. The discrete wavelet transform (DWT) will be obtained as follows:

$$\mathrm{DWT(j,k)} = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{|2^j|}}\psi\left(\frac{t-2^j k}{2^j}\right)dt. \quad (3)$$

Where *j* is the scale parameter, and *k* is the shift parameter. WPD applies this decomposition iteratively to segment the signal's frequency spectrum into distinct sub-bands. By isolating the gamma band through WPD, we obtain a focused time-series feature set for the LSTM, enabling it to model temporal dependencies within this specific frequency range relevant to emotional processing.

### 2.3. Convolutional neural network

A Convolutional Neural Network (CNN) is a specialized deep learning model designed to process and analyze visual data, such as images. The core concept of CNNs lies in their use of convolutional layers, which implement local filtering using small kernels on the input data. These filters slide systematically over the data, extracting essential features and generating feature maps that represent the learned characteristics of the input. The output of each convolutional layer is a set of feature maps that encapsulate these learned features. To reduce the spatial dimensions of these feature maps and enhance computational efficiency, CNNs often incorporate pooling layers. The most common type of pooling is max pooling, which selects the maximum value within a localized region of the feature map. This operation helps in down-sampling the data while retaining critical information.

After passing through several convolutional layers and pooling layers, the output is flattened and directed into one or more fully connected layers, ultimately feeding into an output layer for classification or regression tasks. In our proposed methodology, we designed a CNN with six convolutional layers using the *tanh* activation function, complemented by four max-pooling layers (Figure 3). The default β1, β2, and ϵ values for the Adam optimizer (learning rate = 0.0001) as implemented in the TensorFlow/Keras framework (e.g., β1=0.9, β2=0.999, ϵ=10−7) were used. The batch size was not fixed. A total of 2200 images were generated by applying CWT to 1-D EEG signals, which served as input for the CNN. Unlike most existing studies that employ 32 or 62 channels for emotion recognition, our methodology utilizes a limited number of electrodes (only 5 electrodes), achieving high accuracy while significantly reducing computational complexity.

Through multiple experimental iterations, we discovered that placing max-pooling layers between the first three convolutional layers resulted in a substantial decline in validation accuracy. This decrease is likely due to the information loss associated with max-pooling, despite its advantages in dimensionality reduction. Max-pooling can cause critical feature loss, which is detrimental in scenarios requiring detailed feature preservation. To mitigate this issue and retain useful features, we did not include pooling layers between the first three convolutional layers. A hyperparameter is implemented, which is called the dropout rate. The dropout rate refers to the proportion of neurons randomly deactivated, or "dropped out," during a forward pass in training. This regularization technique is applied to prevent overfitting by reducing the network's reliance on specific neurons, thereby encouraging the model to develop more

generalized representations. Typically, the dropout rate indicates the fraction of neurons set to zero (e.g., a 0.5 dropout rate means half of the neurons are omitted during training). In our proposed CNN structure, the dropout rate was set to 0.6, and the stride was configured to (1,1). Stride refers to the number of steps the convolutional filter moves (or slides) across the input feature map during the convolution operation. A flatten layer was subsequently used to convert the multidimensional feature maps into a single continuous linear vector, preparing data to be input into the dense layers, which require 1D input. The final classification was performed using a dense layer. The dense layer combines and interprets the features from the convolutional and pooling layers to make predictions. The detailed input and output shapes of each layer are illustrated in Figure 4.

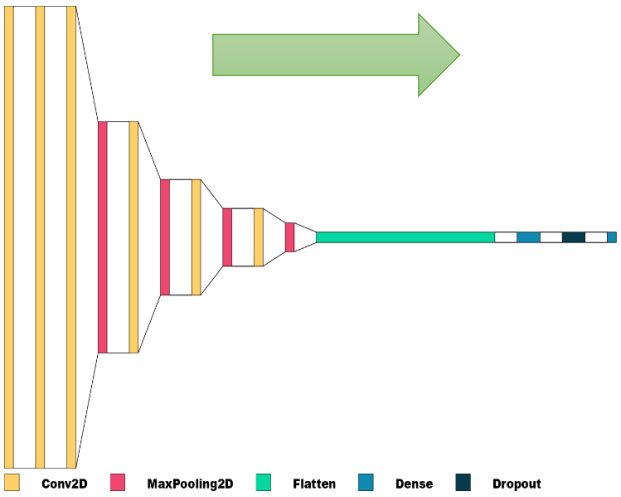


**Figure 3. Architecture of the proposed CNN. We discovered that placing max-pooling layers between the first three convolutional layers resulted in a substantial decline in validation accuracy. Thus, no max-pooling layer was used between these layers.**

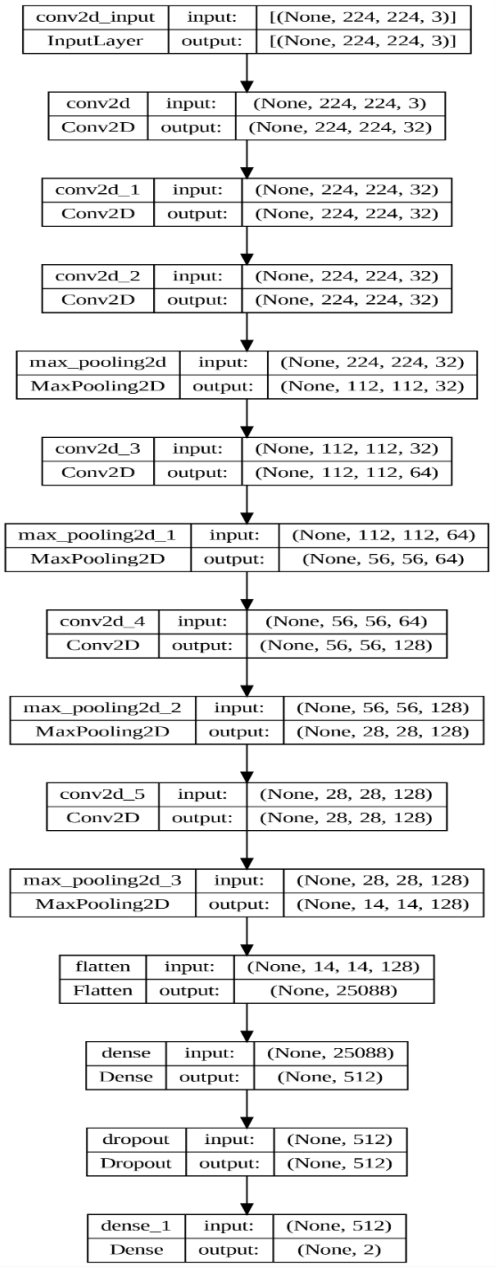


**Figure 4. Detailed input and output shape of each layer of the proposed CNN Architecture. In the first layer, "None" represents the batch size. Since we didn't use a fixed batch size here, we show this parameter as "None". "224" represents the height and width of the input/output images, and "3" corresponds to an RGB image, where the 3 channels represent red, green, and blue color intensities for each pixel. For each layer, the name, input size, and number of filters are presented. The data in the other rows have the same meaning.**

### 2.4. LSTM network

LSTM network is a type of recurrent neural network (RNN) designed to address the vanishing gradient problem [23]. It has the ability to capture long-term dependencies in sequential data, making it well-suited for tasks like language modeling, speech recognition, and time series prediction. LSTMs use memory cells with gates to control information flow, enabling them to retain relevant information for extended periods and forget irrelevant information [24]. Figure 5 shows the LSTM structure that is used in this paper. We used a subject-independent split: data from the first trial of each subject were used for training, and data from the second trial were

used for testing. The third trial was omitted due to computational constraints. This approach helps ensure generalization across sessions and avoids potential data leakage. For LSTM, we extracted gamma-band time series using wavelet packet decomposition (Daubechies-4, 4 levels). There was no data overlap between the training and testing sets.

Due to the various lengths of signals, all signals were padded in the first layer in order to have the same length as the longest time series. A bidirectional LSTM (BiLSTM) layer with 100 neurons is applied to the sequences for detecting time dependencies. BiLSTMs have the feature of additional training by training the data in two different directions (from left to right and from right to the left), which increases the performance and accuracy of the model. Also, BiLSTM works effectively to solve sequence prediction issues and time series [25]. The activation function used for updating cells and hidden neurons is the *tanh*. In addition, we consider the Adam function with a rate of 0.001 for loss calculations. After all, a fully connected layer, a SoftMax layer, and a classification layer are used to classify time-related patterns, respectively.

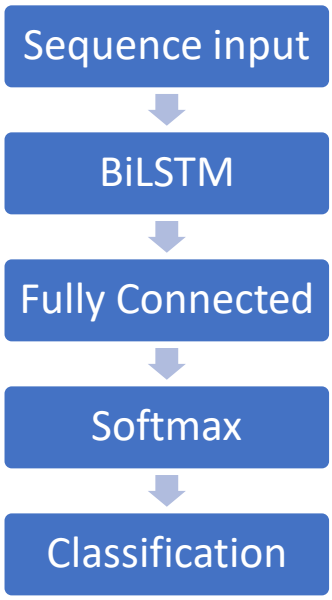


**Figure 5. The block diagram of using LSTM: 1) Padding the input signals to match the length of the longest time series, 2) Using a bidirectional LSTM (BiLSTM) layer with 100 neurons to capture time dependencies in both forward and backward directions, 3) Employing a fully connected layer, SoftMax layer, and classification layer for time-related pattern classification.**

## 3. Evaluation and Results

The evaluation of the classification results is done using the confusion matrix, whose entries contain the percentage of classification results, as shown in Table 1. Based on the confusion matrix, one usually defines 5 main metrics for result evaluation, which are accuracy, specificity, sensitivity, precision, and F-measure [26].

**Table 1. Definition of a confusion matrix**

| | | Predicted Value | |
|---|---|---|---|
| | | Positive | Negative |
| Actual Value | Positive | True Positive (TP) | False Negative (FN) |
| | Negative | False Positive (FP) | True Negative (TN) |

**Accuracy:** This metric measures how many cases are correctly classified. It works well if the classes are balanced. It is defined by equation (4):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FN} + \text{FP} + \text{TN})} \tag{4}$$

**Sensitivity**: It is also called recall. It measures how often a classifier correctly classifies a positive result. It is defined by equation (5):

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{5}$$

**Specificity**: It is also known as True Negative Rate (TNR). A highly specific test means that there are more true negative results. It is calculated by equation (6):

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \tag{6}$$

**Precision**: Precision is a measure of statistical variability. This metric denotes the percentage of correct classifications. It is defined by equation (7):

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{7}$$

**F-Measure**: It represents the harmonic mean of precision and sensitivity values. It is important because if the precision increases, the sensitivity would decrease instead. It is defined by equation (8):

$$\text{F} = 2 \times \frac{(\text{Precision} \times \text{Sensitivity})}{(\text{Precision} + \text{Sensitivity})} \tag{8}$$

### 3.1. Model Evaluation

#### 3.1.1 CNN model

The model evaluation has been done on the SEED database. A balanced database consisting of 2200 images was obtained, of which 1100 images within the database were labeled as positive and 1100 were labeled negative. These images were pooled into a common dataset. We trained the CNN with 60 epochs using 80 percent of the images for training, and 20 percent were used for the test set. As shown in Figure 6 and Figure 7The validation accuracy reached a stable state after certain epochs. The average accuracy of the proposed methodology after 10 runs is 94.57±0.06 % (standard deviation ≈ 0.064). For each run, only the training data has been shuffled. The confusion matrix obtained from the model evaluation can be seen in Figure 8. Values of test data evaluation metrics for our proposed CNN model are presented in two distinct classes, positive and negative, which can be seen in Table 2. The reason for presenting these metrics separately is that high precision for one class and low for the other might indicate class imbalance or a bias in the predictions. Similarly, the specificity might indicate the model's ability to avoid false positives for one class versus the other. These values do not vary too much, which is an indication of our model's robustness in correctly classifying two positive and negative emotions.

**Table 2. Values of test data evaluation metrics**

| | Precision | Recall | F1-score | Total |
|---|---|---|---|---|
| Positive Class | 94.5% | 93.9% | 94.5% | |
| Negative Class | 94.4% | 94.3% | 94.4% | |
| Average Accuracy | | | | 94.57 ±0.06 % |

We did a comparison with other studies using the same dataset in Table 3, which shows that our proposed CNN structure is highly capable of classifying emotions with only five electrodes, compared to the conventional 32 or 64-electrode configurations. This reduction in the number of electrodes offers several key advantages. First, it significantly enhances the system's practicality by reducing setup complexity, which is critical for real-world applications such as wearable devices or clinical settings. Second, it lowers the computational cost and data processing requirements, facilitating faster analysis and enabling the deployment of resource-efficient models. As in our proposed methodology, the CNN model's training time was 3 minutes, which means that utilizing GPU-based classification using 5 electrodes is a promising method for building a reliable emotion recognition system. Finally, minimizing the number of electrodes improves user comfort and reduces potential obstructions, making the system more user-friendly and scalable for diverse populations. Despite the reduced sensor count, the proposed method demonstrates competitive performance, underscoring its efficacy and potential for broader adoption in affective computing and neuroscience research. As we can see in Table 3, to our knowledge, there is only one work ([29]) that uses a small number of EEG electrodes from the SEED dataset for emotion recognition, like our work, but the results of their work are less than our obtained results. They are also two works that use the DEAP dataset for emotion recognition, and they select EEG electrodes that have more influence on emotion recognition. Zhang et al. [30] use the mRMR and Relieff algorithms for finding the best EEG electrodes that show the emotion. They consider different cases, and in the best situation, they obtained 90% accuracy by selecting the 12 EEG electrodes from all 32 electrodes of the DEAP dataset. In another work, Prodhan et al. [31] achieved 95% accuracy for emotion classification by using 12 EEG electrodes of the DEAP dataset. These results show that finding suitable electrodes that could represent the emotion features is not easy. In our work, we can find suitable electrodes, and we can obtain 95% accuracy only by using 5 EEG electrodes, which is very valuable.


**Figure 6. Training and validation loss vs. number of epochs for the proposed CNN model**

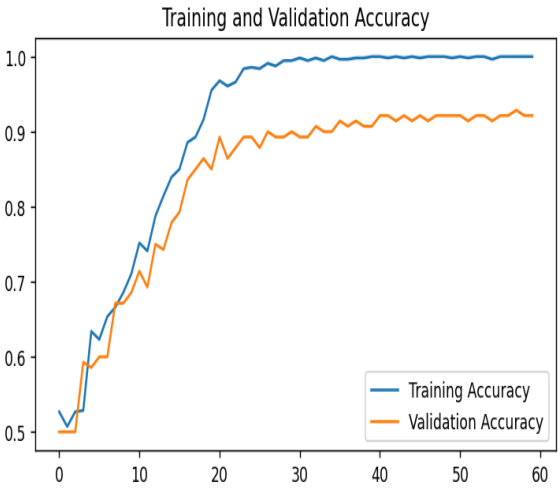


**Figure 7. Training and validation accuracy (Y-axis) vs. number of epochs for the proposed CNN model**

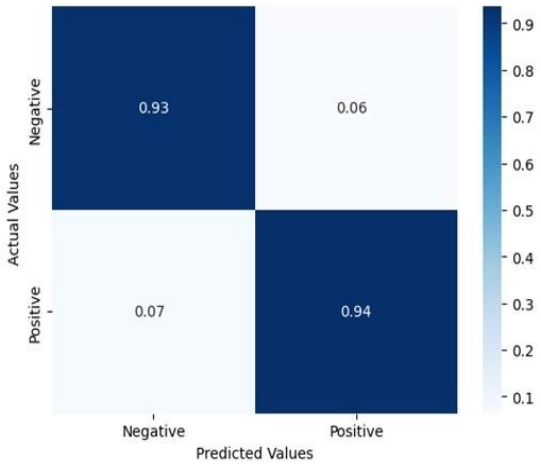


**Figure 8. Confusion matrix of the CNN model (Based on average accuracy percentage). The Y-axis shows actual values, and the X-axis shows the predicted output of the model.**

### 3.1.2 LSTM Model Evaluation

We utilized wavelet packet transform with 4 levels of decomposition using Daubechies 4 as the mother wavelet to extract the gamma band from the signals.

**Table 3. Comparison of the results of our proposed models with other studies**

| Data | Accuracy (%) | Classifier | No. of electrodes | Study |
|---|---|---|---|---|
| SEED | 94.57(±0.06) | CNN-SoftMax | 5 | Proposed methodology |
| SEED | 74.6 | LSTM-SoftMax | 5 | Proposed methodology |
| SEED | 86.03 | DBN | 12 | Zheng et al. [18] |
| SEED | 87.26 | K-NN | 22 | Li et al. [23] |
| SEED | 97.16 | CNN+LSTM (stack) | 62 | Iyer et al. [27] |
| SEED | 90.41<br>3 classes (Positive, negative, neutral) | Extracting differential entropy features+CNN | 62 | Hwang et al. [15] |
| SEED | 91.67<br>93<br>94.33<br>95.67 | Improved CNN | 4<br>6<br>9<br>12 | Ramar et al. [29] |

We used MATLAB 2022, which offers a wide range of deep learning tools; therefore, due to its robustness, we implemented this program first to extract the gamma band and then used an LSTM neural network for emotion classification. The dataset consisted of three trials, each comprising film clips with a duration of approximately 4 minutes. The first and second trials were used for training and testing, respectively, while the third trial was excluded to reduce computational costs and address hardware constraints. The proposed LSTM structure achieved an accuracy of 74.6% after a single training run. The training process, illustrated in Figure 9, reveals oscillations in accuracy over epochs. These fluctuations, observed on the Y-axis, are a common phenomenon in neural network training due to the stochastic nature of optimization processes. Contributing factors include the randomness of weight initialization, the use of mini-batches in stochastic gradient descent, and learning rate variations. Such oscillations reflect the model's navigation through complex loss landscapes, often encountering local minima or saddle points during training.

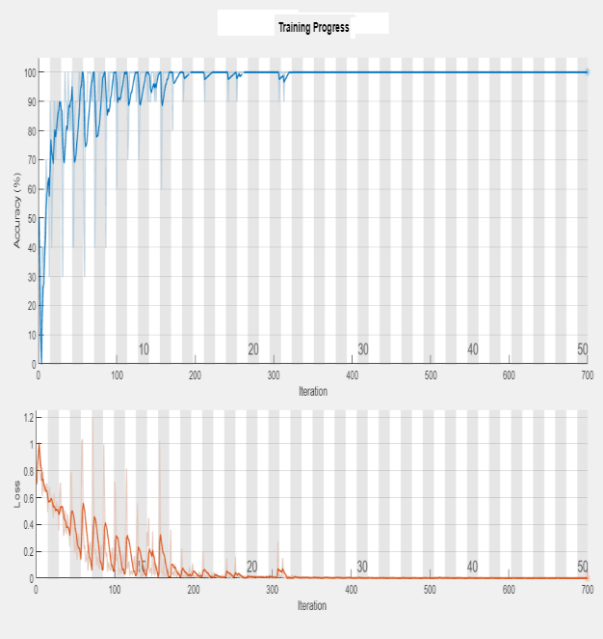


**Figure 9. Training Accuracy and Loss for the proposed LSTM model. The Y-axis and X-axis depict accuracy and number of iterations, respectively. This plot was illustrated by MATLAB.**

## 4. Discussion and Conclusion

In this study, we employed two deep neural network architectures, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, to classify emotional states from EEG signals (positive and negative classes). Our primary objective was to develop a robust methodology for binary emotion classification using only five strategically selected electrodes, aiming to reduce computational complexity and enhance practical applicability.

Our CNN model achieved a remarkable average classification accuracy of (94.57±0.06) % after 10 runs. This high performance underscores its effectiveness in distinguishing emotional states from EEG data even with a minimal number of electrodes, which is significant for real-world applications where computational resources and user comfort are critical. The success of the CNN can be largely attributed to its inherent ability to effectively process the 2-D images generated by the Continuous Wavelet Transform (CWT) from the 1-D EEG signals. CWT transforms the temporal EEG data into a time-frequency representation, creating rich spatial patterns (like spectrograms) that CNNs are exceptionally well-suited to learn and extract features from. This transformation effectively leverages the CNN's strength in identifying hierarchical patterns and local dependencies within these image-like representations, suggesting that the spatial and spectral characteristics captured by CWT are highly discriminative for emotion classification in our setup. The CNN model's training time was approximately 3 minutes, further demonstrating its efficiency.

In contrast, our LSTM network, which analyzed gamma-band time-series features extracted using Wavelet Packet Decomposition (WPD), achieved an accuracy of 74.6% after a single training run. While this result demonstrates the LSTM's capability to process sequential EEG data for emotion recognition, its performance was notably lower than that of the CNN. This discrepancy merits a rigorous examination. Several factors likely contributed to the LSTM's relatively modest accuracy. Firstly, the LSTM focused solely on the gamma band, which, while associated with cognitive and emotional processes, might not encapsulate the full spectrum of temporal dependencies crucial for robust emotion differentiation. Integrating features from a broader range of frequency bands (e.g., alpha, beta, and theta) or more complex time-frequency representations (similar to the CWT used for the CNN) could provide richer temporal context for the LSTM. Secondly, the LSTM's hyperparameters were not as extensively optimized as the CNN's; the architecture was relatively basic, and further exploration of different LSTM layer configurations, neuron counts, dropout rates, and learning rate schedules could significantly enhance its performance. Thirdly, due to computational constraints, the third trial of the SEED dataset was omitted for the LSTM training, potentially limiting the amount of temporal data available for the model to learn from compared to the CNN, which utilized 2200 generated images. Lastly, while LSTMs are excellent at capturing long-term dependencies in time-series data, the specific nature of EEG emotional patterns might benefit more from the multi-scale spatial feature extraction offered by CNNs, particularly when signals are pre-processed into a 2D format like that from CWT.

When comparing our results with prior studies using the SEED dataset (Table 3), the strengths and areas for future improvement become clearer. Our CNN's 94.57% accuracy with only five electrodes is highly competitive, and in many cases superior, to methodologies using a significantly larger number of electrodes (e.g., Zheng et al. [20] with 12 electrodes at 86.03%, Li et al. [25] with 22 electrodes at 87.26%). This highlights our CNN's efficacy and the practical advantage of a reduced electrode setup, which is crucial for building lightweight, comfortable, and resource-efficient emotion recognition systems. While Iyer et al. [32] achieved a higher accuracy (97.16%) using a CNN+LSTM (stack) approach, it is important to note that their methodology utilized 62 electrodes, emphasizing the trade-off between electrode count and potential accuracy gains. Our work demonstrates that high performance can still be achieved with significantly fewer sensors. Regarding our LSTM, its 74.6 accuracy, while demonstrating potential, is lower

than that of other reported LSTM-based approaches or hybrid models. This reinforces the need for further investigation into optimal feature engineering for temporal models and more extensive hyperparameter tuning, as discussed above.

Our findings indicate that different neural network architectures offer distinct advantages depending on the data representation. The CNN's proficiency in extracting spatial features from CWT-transformed EEG signals proved more effective in our current setup. However, the LSTM's capability to model temporal sequences, despite its current performance, remains highly relevant for analyzing time-series data like EEG. This suggests that a hybrid approach, or a more tailored feature extraction method for the LSTM, could provide a more comprehensive understanding of the emotional dynamics encoded in brain signals. Future research could explore combining these architectures with a smaller number of electrodes to create a more agile detection system and further reduce computation cost, or incorporating advanced techniques such as attention mechanisms to enhance the model's ability to focus on critical regions of the data.

Overall, this study contributes to the growing body of work in affective computing by demonstrating the potential of deep learning models for emotion recognition using a reduced number of electrodes. By refining these models and exploring new hybrid approaches, there is significant potential to develop more accurate and efficient emotion recognition systems for brain-computer interface (BCI) systems and real-world applications, such as mental health monitoring, human-computer interaction (HCI), and adaptive learning environments.

Looking forward, future work could expand on these findings by integrating multi-modal data, such as combining EEG with other physiological signals or contextual information, to further enhance emotion classification accuracy. Evaluation on another database (DEAP) was intended, but access to the DEAP dataset was not possible due to the unavailability of the dataset's website at the time. The study, conducted on a limited dataset of 15 participants from a single country, highlights the necessity of validating the results on a larger, more diverse dataset. This would include participants of varying ages, cultural backgrounds, and nationalities to ascertain the generalizability of EEG-based methods across these factors. Ultimately, the ambition is to develop emotion recognition systems that are not only more accurate but also capable of real-time application, paving the way for more empathetic and adaptive human-computer interactions.

## Acknowledgment

The authors would like to thank Prof. Christian Jutten for his helpful comments on the paper.

## AI-Assisted Technology Declaration

We have used AI-assisted tools (common large language models) for polishing the text of this paper. The use of this AI is limited only to text polishing.

## References

[1] Stajić, T., Jovanović, J., Jovanović, N., & Janković, M. M. (2021, November). Emotion recognition based on deap database physiological signals. *In 2021 29th telecommunications forum (TELFOR)* (pp. 1-4). IEEE. doi.org/10.1109/TELFOR52709.2021.9653286

[2] Houssein, E. H., Hammad, A., & Ali, A. A. (2022). Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications,* 34(15), 12527-12557. doi.org/10.1007/s00521-022-07292-4

[3] W. B. Cannon. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology.* 39, 106-124. doi.org/10.2307/1415404

[4] Schuller, Björn & Picard, Rosalind & Andre, Elisabeth & Gratch, Jonathan & Tao, Jianhua. (2021). Intelligent Signal Processing for Affective Computing [From the Guest Editors]. *IEEE Signal Processing Magazine.* 38, 9-11. doi/org/10.1109/MSP.2021.3096415.

[5] S. A. Zendehbad, J. Ghasemi, and F. Samsami Khodadad. (2025). FatigueNet: A hybrid graph neural network and transformer framework for real-time multimodal fatigue detection. *Scientific Reports*. 15, 33781. doi.org/10.1038/s41598-025-00640-z

[6] N. Senthilkumar, S. Karpakam, M. G. Devi, R. Balakumaresan, and P. Dhilipkumar. (2022). Speech emotion recognition based on Bi-directional LSTM architecture and deep belief networks. *Materials Today: Proceedings*, 57, 2180-2184. doi.org/ 10.1016/j.matpr.2021.12.246

[7] S. Razi, M. R. K. Mollaei, and J. Ghasemi. (2019). A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory. *Information Sciences,* 484, 14-26. doi.org/10.1016/j.ins.2019.01.053

[8] P. Sarma and S. Barma. (2021). Emotion recognition by distinguishing appropriate EEG segments based on random matrix theory. *Biomedical Signal Processing and Control*, 70, 102991. doi.org/10.1016/j.bspc.2021.102991

[9] M. Alex, U. Tariq, F. Al-Shargie, H. S. Mir, and H. Al Nashash. (2020). Discrimination of genuine and acted emotional expressions using EEG signal and machine learning. *IEEE Access*, 8, 191080-191089. doi.org/10.1109/ACCESS.2020.3032380

[10] Luo, Y., Fu, Q., Xie, J., Qin, Y., Wu, G., Liu, J., ... & Ding, X. (2020). EEG-based emotion classification using spiking neural networks. *IEEE Access*, 8, 46007-46016. doi.org/10.1109/ACCESS.2020.2978163

[11] S. Bagherzadeh, K. Maghooli, A. Shalbaf, and A. Maghsoudi. (2022). Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals. *Cognitive Neurodynamics*, 16, 1087-1106. doi.org/10.1007/s11571-021-09756-0

[12] Zheng, W. L., Zhu, J. Y., & Lu, B. L. (2017). Identifying stable patterns over time for emotion recognition from EEG. *IEEE transactions on affective computing,* 10(3), 417-429. doi.org/10.1109/TAFFC.2017.2712143

[13] Cimtay, Y., & Ekmekcioglu, E. (2020). Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. *Sensors*, 20(7), 2034. doi.org/10.3390/s20072034

[14] Hwang, S., Hong, K., Son, G., & Byun, H. (2020). Learning CNN features from DE features for EEG-based emotion recognition. *Pattern Analysis and Applications,* 23(3), 1323-1335. doi.org/10.1007/s10044-019-00860-w

[15] Yang, Y., Wu, Q., Fu, Y., & Chen, X. (2018, November). Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In *International conference on neural information processing* (pp. 433-443). Cham: Springer International Publishing. doi.org/10.1007/978-3-030-04239-4_39

[16] Mahmoud, A., Amin, K., Al Rahhal, M. M., Elkilani, W. S., Mekhalfi, M. L., & Ibrahim, M. (2023). *A CNN approach for emotion recognition via EEG. Symmetry,* 15(10), 1822. doi.org/10.3390/sym15101822

[17] Wang, Y., Zhang, B., & Di, L. (2024). Research progress of EEG-based emotion recognition: A survey. *ACM Computing Surveys,* 56(11), 1-49. doi.org/10.1145/3666002

[18] Zheng, W. L., & Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development,* 7(3), 162-175. doi.org/10.1109/TAMD.2015.2431497

[19] Li, J., Zhang, Z., & He, H. (2018). Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cognitive Computation,* 10(2), 368-380. doi.org/10.1007/s12559-017-9533-x

[20] Maithri, M., Raghavendra, U., Gudigar, A., Samanth, J., Barua, P. D., Murugappan, M., ... & Acharya, U. R. (2022). Automated emotion recognition: Current trends and future perspectives. *Computer methods and programs in biomedicine,* 215, 106646. doi.org/10.1016/j.cmpb.2022.106646

[21] Zhang, H., Jolfaei, A., & Alazab, M. (2019). A face emotion recognition method using convolutional neural network and image edge computing. *IEEE Access,* 7, 159081-159089. doi.org/10.1109/ACCESS.2019.2949741

[22] Al Rahhal, M. M., Bazi, Y., Al Zuair, M., Othman, E., & BenJdira, B. (2018). Convolutional neural networks for electrocardiogram classification. *Journal of Medical and Biological Engineering,* 38(6), 1014-1025. doi.org/10.1007/s40846-018-0389-7

[23] Li, J. W., Barma, S., Pun, S. H., Vai, M. I., & Mak, P. U. (2022). Emotion recognition based on EEG brain rhythm sequencing technique. *IEEE Transactions on Cognitive and Developmental Systems,* 15(1), 163-174. doi.org/10.1109/TCDS.2022.3149953

[24] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems,* 28(10), 2222-2232. doi.org/10.1109/TNNLS.2016.2582924

[25] Algarni, M., Saeed, F., Al-Hadhrami, T., Ghabban, F., & Al-Sarem, M. (2022). Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM). *Sensors,* 22(8), 2976. doi.org/10.3390/s22082976

[26] Aslan, M. (2022). CNN based efficient approach for emotion recognition. *Journal of King Saud University-Computer and Information Sciences,* 34(9), 7335-7346. doi.org/10.1016/j.jksuci.2021.08.021

[27] Iyer, A., Das, S. S., Teotia, R., Maheshwari, S., & Sharma, R. R. (2023). CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings. *Multimedia Tools and Applications,* 82(4), 4883-4896. doi.org/10.1007/s11042-022-12310-7