EVALUATING THE PERFORMANCE OF SOME DEEP LEARNING ARCHITECTURES IN THE PROBLEM OF EMOTION RECOGNITION FROM EEG SIGNALS

Thuy Pho Duc [1]

[1] *Head of the Equipment Department, Military Hospital 91, Thai Nguyen, Vietnam*

ABSTRACT

The problem of emotion recognition from EEG signals is a topic of great interest in research worldwide. Effectively solving these problems will provide an efficient solution in building advanced intelligent HCI systems, applicable in various fields such as healthcare, education, entertainment, military, and society. The biggest challenge is finding suitable deep learning architectures that yield high recognition accuracy. In this paper, the author evaluates the effectiveness of three deep learning architectures (CNN, LSTM, a combination of CNN and LSTM) in recognizing emotions from the internationally recognized DEAP dataset. Experimental results show that the combined CNN and LSTM architecture achieves the highest efficiency with an accuracy of 92.26%, a loss of 0.1703, and an F1 score of 0.9055. This confirms the potential application of the combined CNN and LSTM deep learning architecture in real-world emotion recognition from EEG signals.

Keyword: *Emotion Recognition, EEG, CNN, LSTM, DEAP Dataset*

1. INTRODUCTION

Human emotions are important not only in expressing psychological states and physical health but also in interaction, recognition, and decision-making [1]. Therefore, the ability to accurately recognize emotions can enhance the intelligence of human-computer interaction systems, better serving human needs [2].

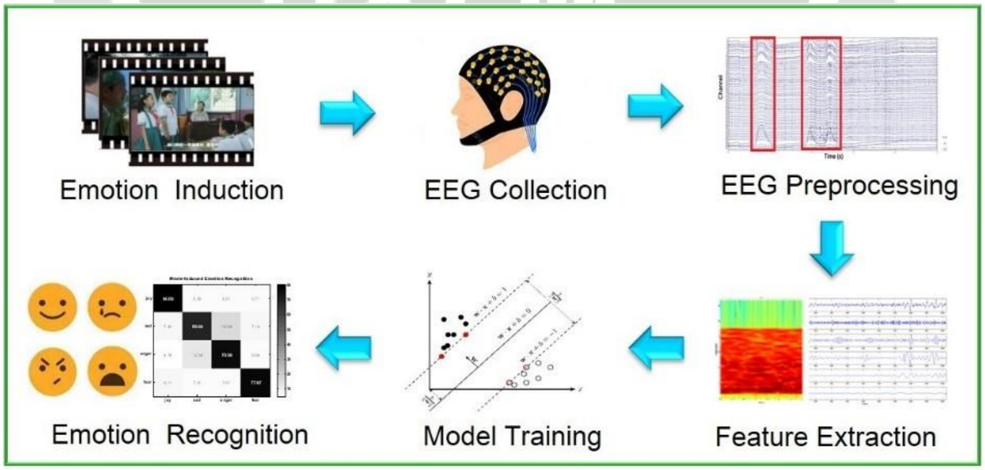


Fig -1: Steps in the Emotion Recognition Process [5]

Based on this reality, many techniques have been introduced to solve the emotion recognition problem. Some methods use non-physiological signals, such as voice, facial expressions, and body posture. However, the accuracy of these methods can be affected by various factors, including mental state, gender, education level, age, and the ability to conceal emotions. Due to these subjective factors, evaluating a person's emotions can become challenging and inaccurate [3][4].

In addition, some studies have shown that physiological signals and emotional expressions are closely related. Therefore, methods using physiological signals such as heart rate, skin impedance, respiration, and EEG have been proposed to determine emotional states. Among these, the study of emotional cognition mechanisms and recognizing emotional states using EEG signals is particularly important, especially for those who cannot express emotions through natural speech, facial expressions, or body posture [3].

As depicted in Figure 1, emotion recognition based on EEG signals primarily includes the steps of emotion stimulation, EEG signal collection, EEG signal preprocessing, extraction and analysis of emotion-related EEG features, building and training emotion computational models, and detecting and recognizing emotional states. Emotion recognition techniques from EEG signals are mainly developed from two approaches: traditional machine learning and deep learning [4].

Deep learning techniques, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have shown their superiority in detection and classification in fields like natural language processing, audio, and image processing [6]. Therefore, applying these techniques to emotion recognition from EEG signals promises notable results ([3] [6]). Due to these advantages and potentials, many deep learning architectures have recently been proposed by the scientific community to find an effective model for developing practical applications based on EEG signal recognition ([6] -[12]).

In this paper, to find suitable deep learning network architectures, the author proposes three different deep learning architectures (CNN, LSTM, a combination of CNN and LSTM) and evaluates their recognition effectiveness on a publicly recognized EEG dataset (DEAP). Based on that, the remainder of the paper is structured as follows. In Section 2, the author describes the selection of the Framework for data preparation and feature extraction. Then, in Section 3, the author analyzes and proposes the network architectures. In Section 4, the paper presents some experimental results on the recognition effectiveness of the three proposed network architectures. Section 5 will provide some conclusions.

2. DATA PREPARATION, FEATURE EXTRACTION, AND FRAMEWORK SELECTION

2.1. Framework Selection

To effectively and time-efficiently train deep learning architectures, it requires the trainer to use high-performance hardware. Given the research conditions, the author chose to conduct experiments using the TensorFlow library. All deep learning models were trained on Google Colab Pro with High RAM and GPUT4. The author used the Early Stopping technique to halt training when the loss value on the validation set did not improve after a certain number of epochs.

2.2. Data Preparation

To recognize emotions, having a source of data is essential. However, many studies are limited by conditions and cannot independently establish a standard experimental environment. Most researchers, when benchmarking results against similar studies, often rely on widely accepted standard datasets. Therefore, creating open-source EEG emotion datasets becomes crucial to support emotion recognition from EEG signals.

In this paper, the author chose to use the DEAP dataset [7]. As shown in Figure 2, this dataset includes physiological signals from 32 participants (16 males and 16 females) collected through a 40-channel system, with 32 channels from EEG electrodes and 8 other channels from peripheral physiological signals. Participants were asked to watch 40 video clips, each 60 seconds long. After viewing each clip, they needed to rate their emotions on a scale of 1 to 9 for four types of emotions: Arousal, Valence, Dominance, and Liking. These ratings are used as labels for the corresponding signal collection process. Once the signals and ratings are verified to be consistent, the signal collection proceeds with the next video.

The author used the preprocessed version of the DEAP dataset, provided by the authors. This version is divided into two main parts: data and labels. The data part includes a matrix of dimensions 40×40×8064 (number of samples × number of videos × number of channels). Regarding the labels, it contains a 40×4 matrix (rating level × type of emotion), where each column represents a specific type of emotion such as Arousal, Valence, Dominance, and Liking. This preprocessed version has a reduced sampling rate of 128 Hz, and the signals are filtered with a bandpass filter with frequencies from 4Hz to 40.5Hz. The data collected in the first 3 seconds were discarded. The author created two datasets. The first dataset uses the 32 EEG channels in the preprocessed dataset because the remaining channels (channels 33 to 40) are used for measuring EOG signals and other parameters such as temperature and blood pressure, which are not relevant to this study. For the second dataset, the author only used data from electrodes equivalent to a

14-channel device like the Emotiv Epoc. Both datasets were randomly divided into three sets: training set (65%), validation set (15%), and test set (20%).

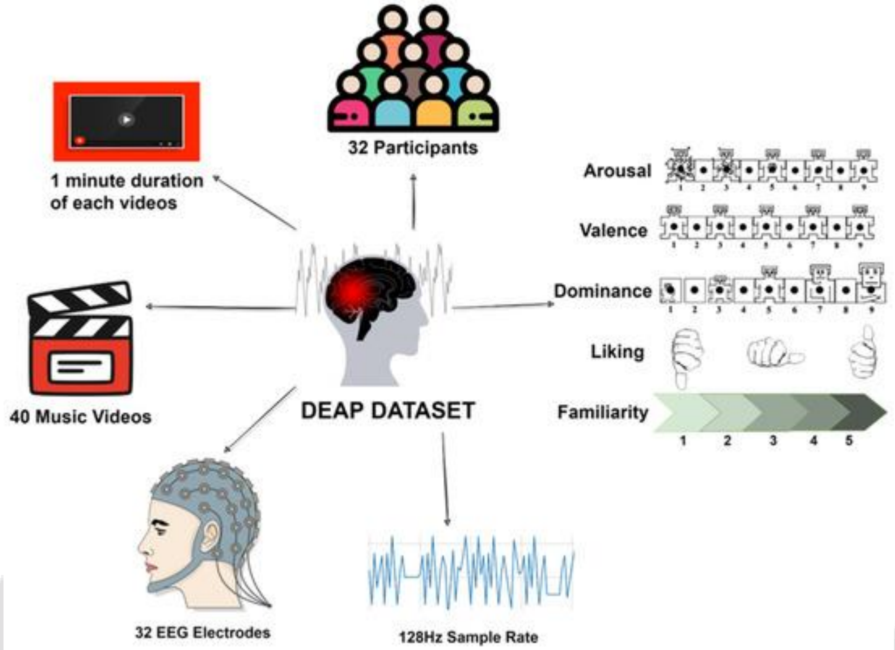


Fig -2: Introduction to the DEAP Dataset [7]

2.3. Feature Extraction from EEG Signals

Although deep learning architectures allow us to train recognition models from raw data, the accuracy of emotion recognition mainly depends on the extracted features. Therefore, extracting important EEG features related to emotional states is crucial. EEG feature analysis is often performed in the time domain, frequency domain, or both. However, frequency domain features have shown better performance in emotion recognition using EEG signals compared to time domain features ([1] [3]). Frequency domain analysis methods transform EEG signals from the time domain to the frequency domain to evaluate and extract frequency domain features. Typically, EEG signals are divided into five different frequency bands. Parameters such as power spectral density (PSD), logarithmic energy spectrum, higher-order spectra (HOS), and differential entropy (DE) from each frequency band are used for analysis. The method chosen in this paper for frequency analysis is Fast Fourier Transform (FFT) with the following steps [9]:

1) Initialize parameters: Including the number of EEG channels (32, 14, 5), FFT window size (256 samples, equivalent to 2 seconds), step size to move the window (16), list of EEG frequency bands to be analyzed (delta, theta, alpha, beta, gamma).

2) Divide the time-domain data into short segments. Each segment is 256 samples long (window_size) to optimize computational efficiency and move it through the time-domain data with a step size of 16.

3) For each segment, we use FFT to transform the data from the time domain to the frequency domain. The result of this transformation is a sequence of complex numbers, each representing a different "fundamental frequency" in the original time-domain data. The real part of the complex number represents the amplitude of that frequency, while the imaginary part represents the phase (time shift) of that frequency.

The result of the FFT, represented as a spectral plot, shows the strength of each fundamental frequency in the data. This is very useful in many applications, including audio analysis, signal processing, and EEG data analysis, as it allows differentiation of the signal's various frequency components.

3. PROPOSED NETWORK ARCHITECTURES

The recognition effectiveness of an emotion recognition system, besides depending on the type of data and input features, also depends on the chosen deep learning network architecture.

With the DEAP dataset, which is a multi-dimensional EEG dataset (data from 32 channels), reflecting the activity of different brain regions, this creates important spatial features. CNN architecture is well-suited to process these spatial features.

Moreover, EEG signals contain specific frequency waveforms, such as Alpha, Beta, Gamma, Delta, Theta waves, which may relate to the user's emotions. Transforming the signal from the time domain to the frequency domain will help reveal these frequency features.

When EEG signals are divided into small segments (to find frequency features), each signal segment has a temporal relationship with other segments. LSTM architecture can learn patterns and temporal changes from these signal segments, enhancing performance in many tasks.

Therefore, the author chose to experiment with three types of network architectures (CNN, LSTM, and a combination of CNN and LSTM). The parameters of these network architectures were determined based on experimental results combined with hyperparameter tuning in the Google Colab Pro environment [13].

3.1. CNN Network Architecture Selection

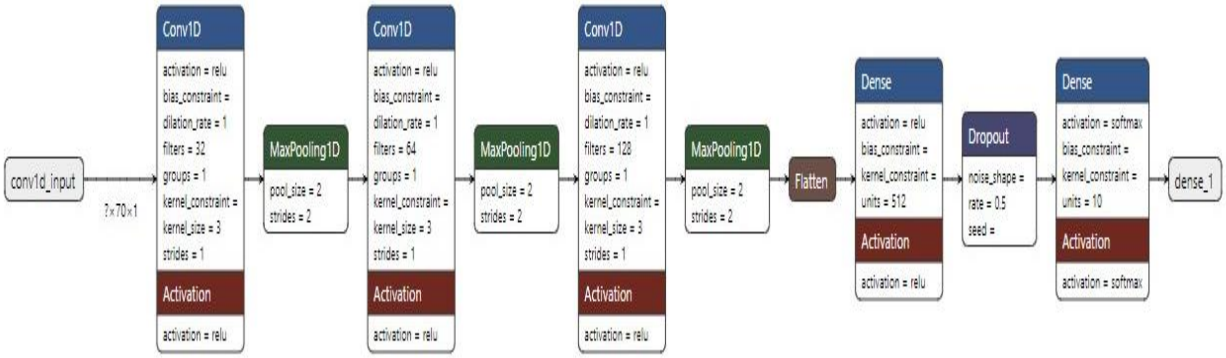


Fig -3: CNN Architecture

Figure 3 describes the CNN architecture for 32-channel input data chosen by the author. This architecture includes:

First Convolutional Layer, *Conv1D (32 filters, kernel size 3)*: Extracts spatial features from the EEG signals. This is the first layer to process the signal, helping to recognize characteristic patterns in small spatial regions. MaxPooling1D (pool size=2): Reduces the spatial dimension of the data, decreases computational load, and helps the model focus on more important features.

Second Convolutional Layer, *Conv1D (64 filters, kernel size 3*): Extracts more complex features from the previous features. This layer helps the model learn more features from the data. MaxPooling1D (pool size=2): Continues to reduce the spatial dimension, helps reduce the number of parameters, and prevents overfitting.

Third Convolutional Layer, *Conv1D (128 filters, kernel size 3)*: Extracts additional high-level features from the data. This layer helps the model recognize the most important features. MaxPooling1D (pool size=2): Further reduces the spatial dimension, helping the model focus on the most important features.

Flatten Layer: Converts the 3D tensor into a 1D vector to prepare for the Dense layers. This is a necessary step before feeding the data into the fully connected layers.

First Dense Layer, *Dense (512 units, activation='relu')*: Learns non-linear features from the extracted data. This layer has a large number of units, helping the model learn many complex features.

Dropout Layer, *Dropout (rate=0.5)*: Reduces overfitting by randomly dropping some nodes during training. This increases the model's generalization ability.

Second Dense Layer (Output layer), *Dense (num_classes, activation='softmax')*: Predicts the probability of each emotion class (10 classes). This layer uses the softmax activation function to convert the output into probabilities for each class.

The CNN model is initialized with the following parameters:

- Batch size: This is the size of each data batch fed into the model during each training iteration. Here, the batch size is 256.
- Number of classes (num_classes): This is the number of output classes of the model, corresponding to the number of labels the model will predict. Here, the number of classes is 10.
- Number of epochs: This is the number of times the entire training data will be passed through the model. Here, the number of epochs is 200.
- Input shape: The input data has a shape of (160, 1), meaning each data sample has 160 features and 1 channel.

Additionally, the author used the Early Stopping technique to stop training early if the loss value on the validation set does not decrease after a certain number of epochs, to avoid overfitting during training.

3.2. LSTM Network Architecture Selection

EEG signals are sequential data, meaning their value depends on previous values. LSTM has the ability to remember previous information in the sequence, helping the model understand the temporal relationships of EEG signals. This is especially important in EEG signal analysis because important features may appear at different time intervals. LSTM helps the model learn sequential features from EEG signals, such as changes in amplitude or frequency over time, which Conv1D layers may not fully capture. Figure 4 describes the chosen LSTM network architecture. This architecture includes:

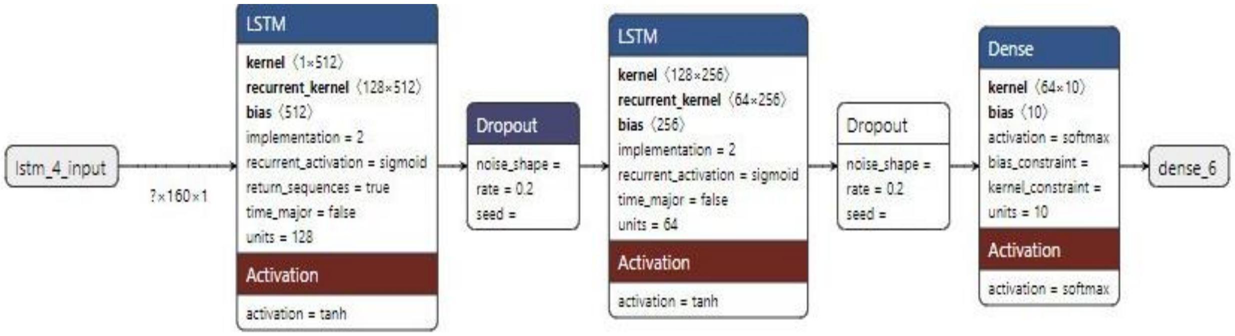


Fig -4: LSTM Architecture

First LSTM Layer, *LSTM (128 units, return_sequences=True)*: Learns sequential features from EEG signals. This layer can remember long-term and short-term sequential information, suitable for the nature of EEG data.

Dropout Layer (0.2): Reduces overfitting by randomly dropping some units during training, helping to improve the model's generalization ability.

Second LSTM Layer, *LSTM (64 units)*: Learns higher-level sequential features from the previous LSTM layer, helping the model recognize more complex sequential patterns.

Dropout Layer (0.2): Continues to reduce overfitting and improve the model's generalization ability.

Final Dense Layer, *Dense (num_classes, activation='softmax')*: Predicts the probability of each emotion class (10 classes). This layer uses the softmax activation function to convert the output into probabilities for each class, suitable for multi-class classification problems.

3.3. Combined CNN and LSTM Network Architecture Selection

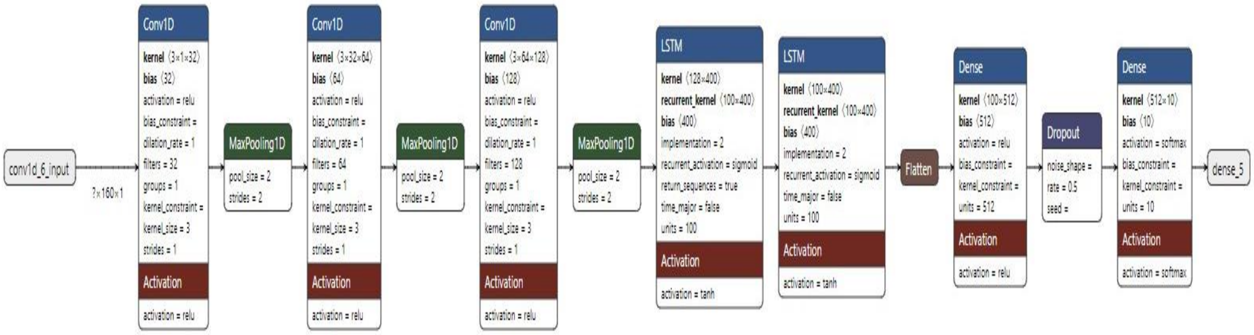


Fig -5: Combined CNN and LSTM Network Architecture

Figure 5 describes the combined CNN and LSTM network architecture for 32-channel input data. In this architecture, the initial Conv1D layers help extract spatial features from the EEG signals, such as amplitude variations at specific positions in the signal sequence. The two subsequent LSTM layers (added after the third Conv1D layer) help combine these spatial features with temporal features, creating a more comprehensive model for classifying emotional states from EEG signals. The first LSTM layer with “return_sequences=True” retains the entire output sequence to be fed into the next LSTM layer. The second LSTM layer (100 units) functions to learn higher-level sequential features from the previous LSTM layer. This layer helps the model recognize more complex sequential patterns.

4. EXPERIMENTAL RESULTS

4.1. For the CNN Network Architecture

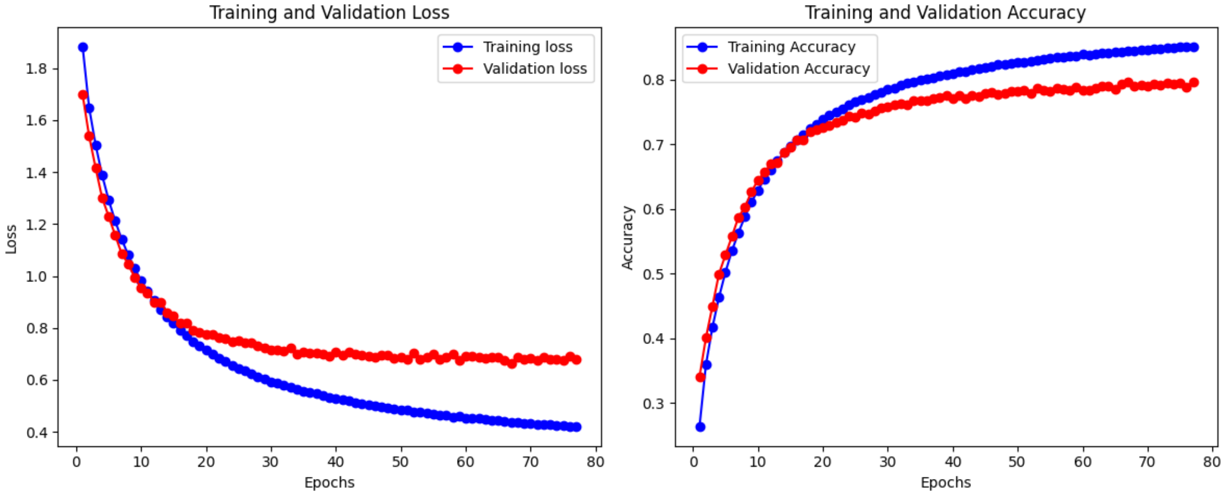
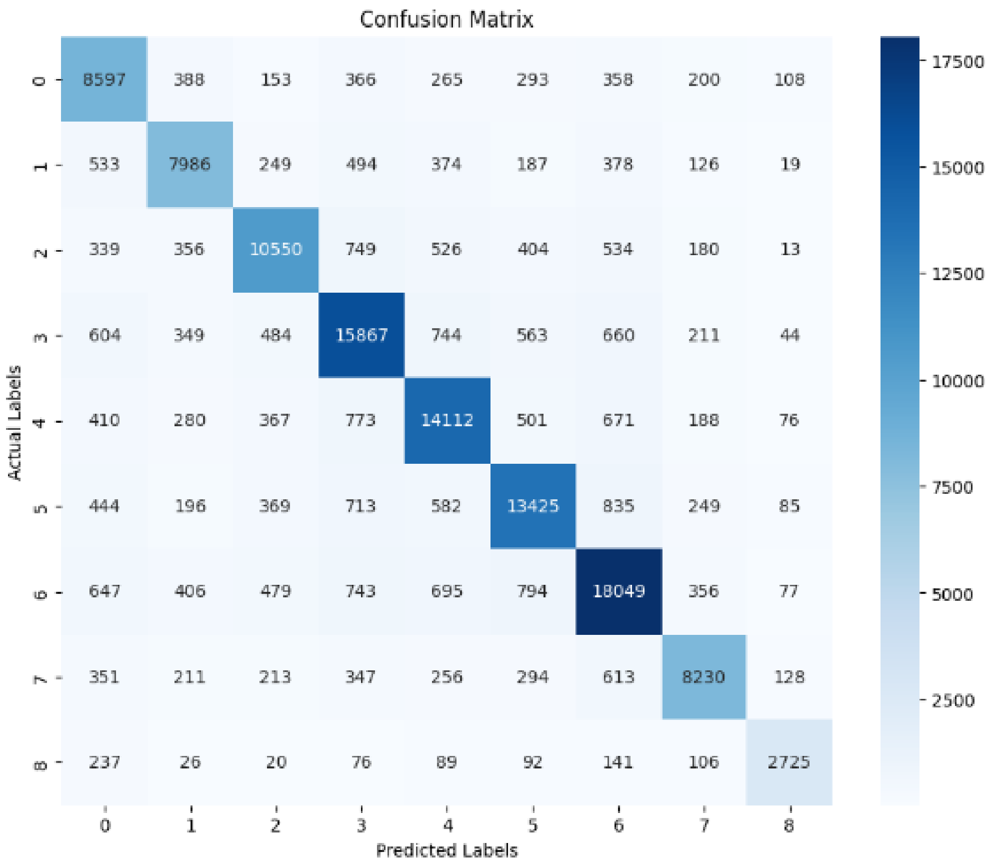


Fig -6: Learning Curve of the CNN Network



	0	1	2	3	4	5	6	7	8
0	8597	388	153	366	265	293	358	200	108
1	533	7986	249	494	374	187	378	126	19
2	339	356	10550	749	526	404	534	180	13
3	604	349	484	15867	744	563	660	211	44
4	410	280	367	773	14112	501	671	188	76
5	444	196	369	713	582	13425	835	249	85
6	647	406	479	743	695	794	18049	356	77
7	351	211	213	347	256	294	613	8230	128
8	237	26	20	76	89	92	141	106	2725

Fig -7: Confusion Matrix of the CNN Architecture

Figure 6 shows the training results of the CNN network with 32-channel input data. The model stopped training at epoch 77 due to the Early Stopping mechanism, indicating good convergence and avoidance of overfitting. The accuracy on the test set is 91.68%, demonstrating high accuracy and the ability to accurately classify emotional states.

Additionally, the loss value on the test set is 0.2143, showing that the model has learned well and generalized the data. The Training and Validation Loss graph shows that both the training and validation losses decrease steadily over the epochs. The validation loss decreases more slowly and begins to stabilize around epoch 30, with no significant improvement afterward, indicating the model's optimization capability is limited.

In the Training and Validation Accuracy graph, the accuracy on both the training and validation sets increases steadily over the epochs. The accuracy on the validation set is slightly higher than the training set initially, but the gap between the two gradually widens after about epoch 30, indicating signs of slight overfitting.

The results shown in the confusion matrix in Figure 7 indicate that the labels have good accuracy, but there is still some confusion between closely related labels. Specifically, labels 3 and 6 have a high correct prediction rate, demonstrating good recognition ability for these emotional states. On the other hand, some labels such as 0, 1, and 2 show a significant degree of confusion with other labels. This may be due to EEG signals containing a lot of noise or having indistinguishable features.

From these results, it can be seen that the CNN model achieves high accuracy with low loss values, demonstrating effective emotion classification from EEG signals. The training and validation graphs illustrate the model's good convergence and the ability to avoid overfitting due to the Early Stopping mechanism. Although there is still some confusion between labels, these results are very promising and demonstrate the potential practical application of the CNN model in emotion analysis.

4.2. For the LSTM Network Architecture

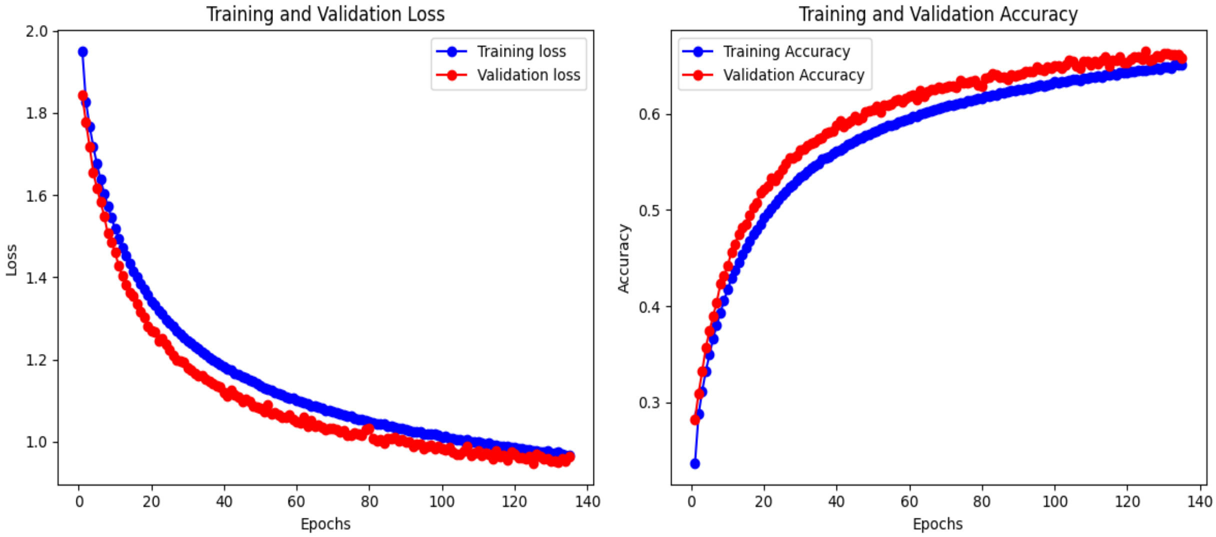


Fig -8: Learning Curve of the LSTM Network

The recognition performance of the LSTM deep learning model is shown in Figure 8. Thanks to the Early Stopping technique, the model stopped training at epoch 135. This indicates that the model has achieved convergence and avoided overfitting.

The accuracy on the test set reached 74.43%, indicating that the LSTM model's ability to classify emotional states is not as high as expected. The loss value on the test set is 0.59865, which is relatively high, suggesting that the model has not learned well and may not have fully captured important features from the EEG data. The Training and Validation Loss graph shows that both the training and validation losses decrease steadily over the epochs. The validation loss begins to be higher than the training loss from around epoch 40, with no significant improvement afterward, indicating the model's limited optimization capability.

From the Training and Validation Accuracy graph, we can see that the accuracy on both the training and test sets increases steadily over the epochs. The test accuracy is slightly higher than the training accuracy initially, but the gap between the two gradually widens after around epoch 40, indicating signs of overfitting.

From the confusion matrix in Figure 9, we can see that the labels have average accuracy but still significant confusion between closely related labels. The performance of the LSTM model is lower compared to the CNN model. This leads to the conclusion that while the LSTM model can learn sequential features from EEG signals, its performance is not high, and there are signs of overfitting. This indicates that using LSTM alone is not sufficient to achieve good performance in the task of emotion classification from EEG signals. Further improvements and adjustments are needed to enhance the model's performance.

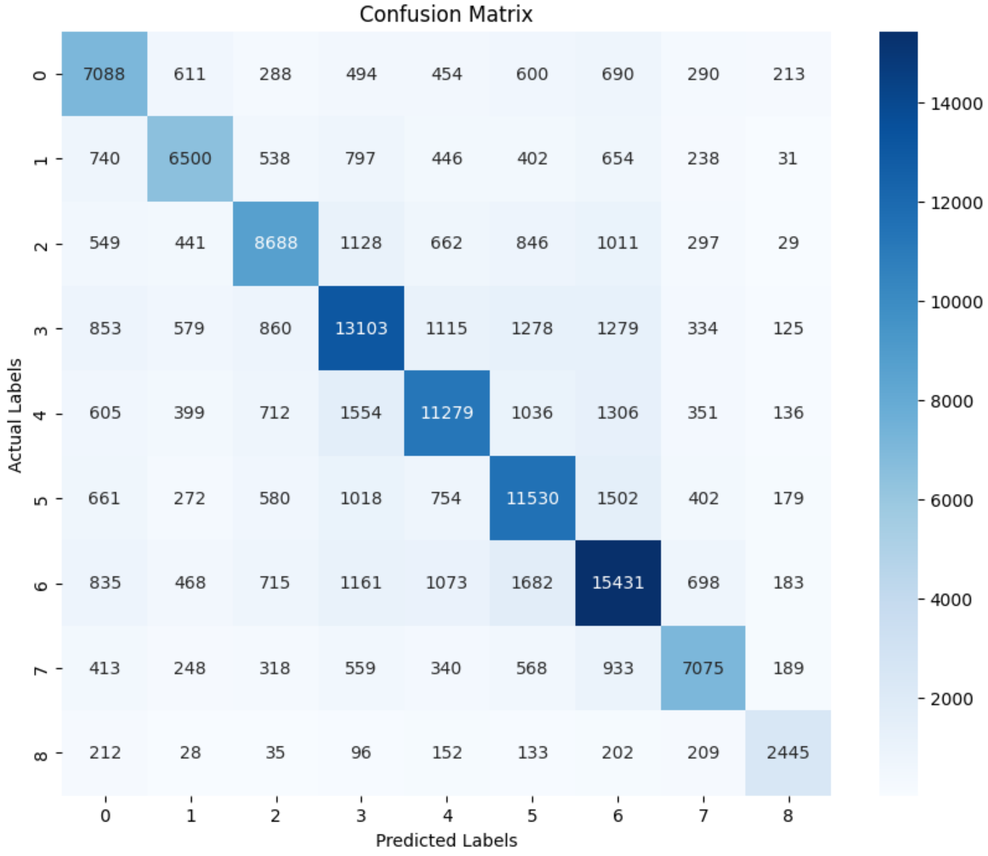


Fig -9: Confusion Matrix of the LSTM Architecture

4.3. For the Combined CNN and LSTM Network Architecture

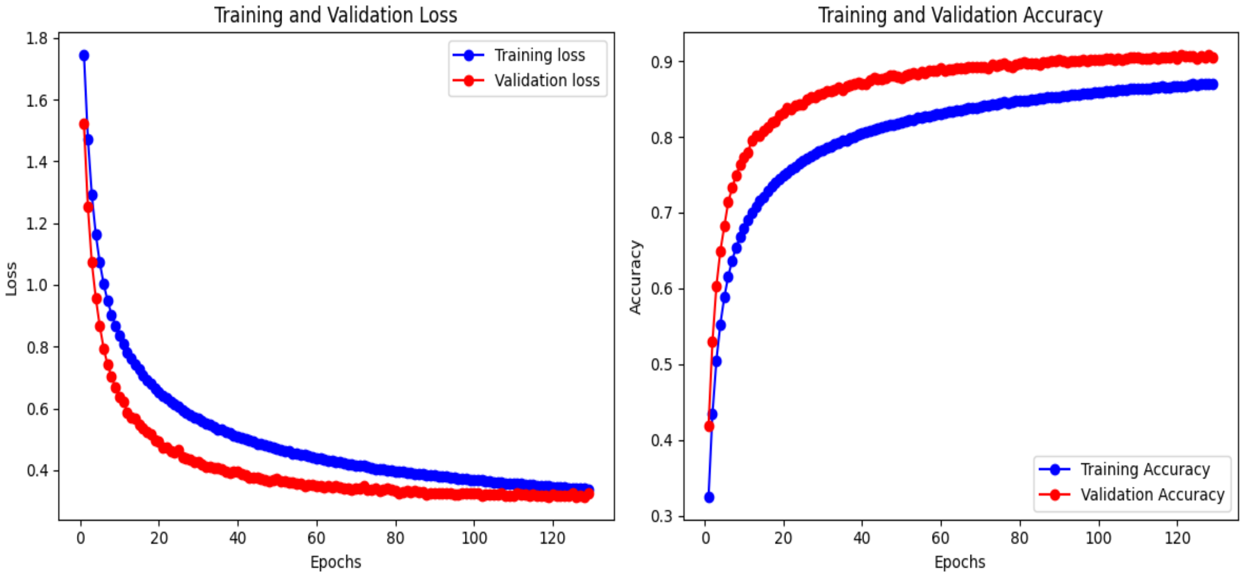


Fig -10: Learning Curve of the Combined CNN and LSTM Network

Figure 10 shows the training results of the combined CNN and LSTM network with 32-channel input data. The model stopped training at epoch 129 due to the Early Stopping mechanism, indicating good convergence and avoidance of overfitting. The accuracy on the test set is 92.26%, demonstrating high accuracy and the ability to accurately classify emotional states. Additionally, the loss value on the test set is 0.1702, showing that the model has learned well and generalized the data. The Training and Validation Loss graph shows that both the training and validation losses decrease steadily over the epochs. The validation loss starts to be higher than the training loss after a few initial epochs, indicating the beginning of overfitting, but Early Stopping helped stop training in time. The loss decreased from around 1.8 to about 0.3, indicating that the model learned well and significantly reduced errors during training. In the Training and Validation Accuracy graph, the accuracy on both the training and test sets increases steadily over the epochs. The test accuracy is slightly higher than the training accuracy, demonstrating the model's generalization ability.

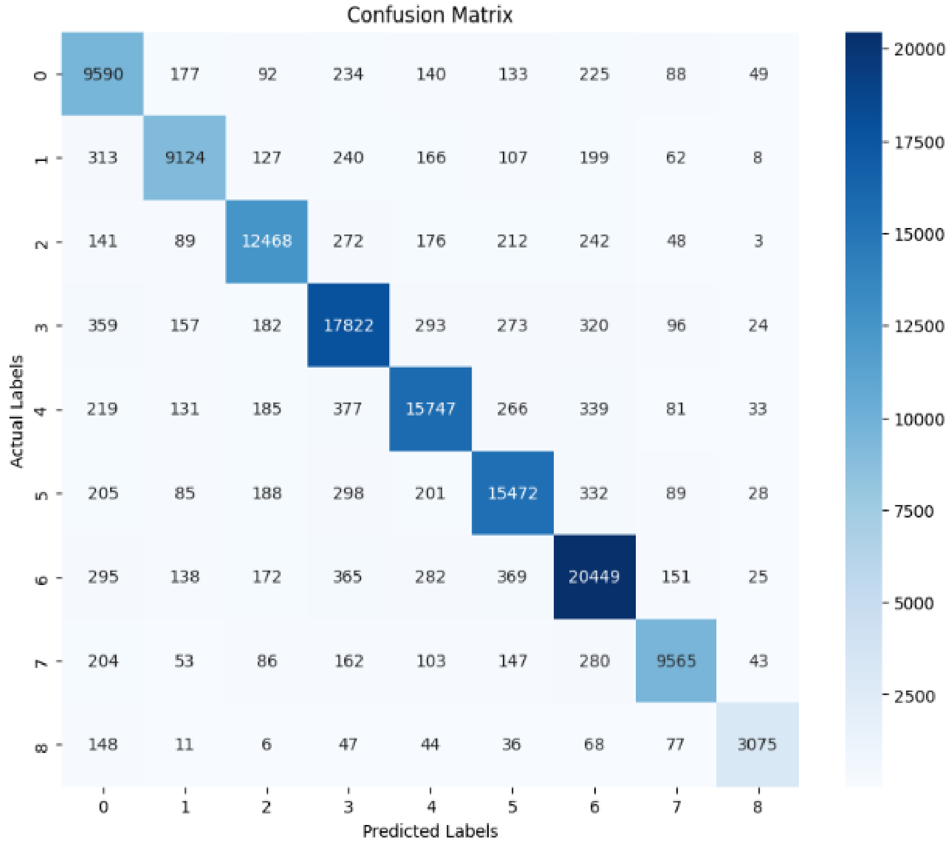


Fig -11: Confusion Matrix of the Combined CNN and LSTM Network Architecture

The results shown in the confusion matrix (Figure 11) indicate that most labels are predicted accurately, demonstrating the high performance of the model. Labels 3, 6, and 7 have high correct prediction rates, showing the model's strong recognition ability for these labels. However, some labels, such as 0, 1, and 2, exhibit a certain degree of confusion, especially among labels with similar characteristics.

From these results, it can be seen that the Combined CNN and LSTM model achieves high accuracy with low loss values, demonstrating effective emotion classification from EEG signals. The training and validation graphs illustrate the model's good convergence and the ability to avoid overfitting due to the Early Stopping mechanism. Although there is still some confusion between labels, these results are very promising and show the potential of combining spatial and temporal features from EEG signals.

Table 1 shows detailed results illustrating the recognition performance of the deep learning architectures for each emotion label measured by parameters such as accuracy, loss, and F1 score. It can be seen that the combined CNN and LSTM network architecture achieved the highest recognition performance as expected.

Table -1: Recognition Results of Various Deep Learning Algorithms Name of the Table

Model	Label	Acc	Loss	F1 Score
CNN	Valence	0.907	0.2126	0.8908
	Arousal	0.912	0.2054	0.8755
	Dominance	0.9256	0.2254	0.8635
	Like	0.9227	0.2141	0.8845
Average		**0.916825**	**0.214375**	**0.878575**
LSTM	Valence	0.6655	0.6468	0.7145
	Arousal	0.7414	0.6006	0.6998
	Dominance	0.7762	0.5984	0.7001
	Like	0.7944	0.5488	0.7546
Average		**0.744375**	**0.59865**	**0.71725**
CNN+LSTM	Valence	0.9168	0.1625	0.8994
	Arousal	0.9212	0.1724	0.9016
	Dominance	0.9245	0.1847	0.9115
	Like	0.9278	0.1614	0.9097
Average		**0.922575**	**0.17025**	**0.90555**

5. CONCLUSIONS

Emotion recognition based on EEG signals is a fascinating research field due to its potential applications in many real-world problems. This is a complex problem, but it can be addressed by applying research achievements in fields such as digital signal processing and artificial intelligence. In particular, using deep learning architectures in recognition has yielded encouraging results.

With EEG data being multidimensional (data from multiple electrode channels), it reflects the activity of different brain regions, creating important spatial features. The CNN architecture is well-suited to process these spatial features. On the other hand, each signal segment has a temporal relationship with other segments. LSTM networks can learn patterns and temporal changes from these signal segments, meaning they consider both past and future information of each data point, enhancing performance in many tasks. Therefore, constructing a combined CNN and LSTM network allows for improved effectiveness in emotion recognition from EEG signals.

By using hyperparameter tuning techniques on Google Colab, in this paper, the author evaluated the recognition performance of several deep learning architectures (CNN, LSTM, and a combination of CNN and LSTM) with the same FFT input features on DEAP data. The evaluation results show that, as theoretically predicted, the combined CNN and LSTM network architecture provided good recognition performance with an accuracy of 92.26%, a loss of 0.1703, and an F1 score of 0.9055. These results partially confirm the potential application of EEG signal-based emotion recognition solutions in various real-life problems.

6. REFERENCES

[1] . Lin, Wenqian, and Chao Li. 2023. "Review of Studies on Emotion Recognition and Judgment Based on Physiological Signals" Applied Sciences 13, no. 4: 2573. https://doi.org/10.3390/app13042573

[2] . Vempati, Raveendrababu & Sharma, Lakhan. (2023). A Systematic Review on Automated Human Emotion Recognition using Electroencephalogram Signals and Artificial Intelligence. Results in Engineering. 18. 101027. 10.1016/j.rineng.2023.101027.

[3] . Wang X, Ren Y, Luo Z, He W, Hong J and Huang Y (2023) Deep learning-based EEG emotion recognition: Current trends and future perspectives. Front. Psychol. 14:1126994. doi: 10.3389/fpsyg.2023.1126994

[4] . Liu, Haoran & Zhang, Ying & Li, Yujun & Kong, Xiangyi. (2021). Review on Emotion Recognition Based on Electroencephalography. Frontiers in Computational Neuroscience. 15. 10.3389/fncom.2021.758212.

[5] . He, Zhongyang, Ning Zhuang, Guangcheng Bao, Ying Zeng, and Bin Yan. 2022. "Cross-Day EEG-Based Emotion Recognition Using Transfer Component Analysis" Electronics 11, no. 4: 651. https://doi.org/10.3390/electronics11040651

[6] . Akter, Sumya, Rumman Ahmed Prodhan, Tanmoy Sarkar Pias, David Eisenberg, and Jorge Fresneda Fernandez. 2022. "M1M2: Deep-Learning-Based Real-Time Emotion Recognition from Neural Activity" Sensors 22, no. 21: 8467. https://doi.org/10.3390/s22218467

[7] . Koelstra S, et al. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing. 2011;3(1):18-31; Available from: https://doi.org/10.1109/ T-AFFC.2011.15

[8] . Zhang Y, Chen J, Tan JH, Chen Y, Chen Y, Li D, Yang L, Su J, Huang X and Che W (2020) An Investigation of Deep Learning Models for EEG-Based Emotion Recognition. Front. Neurosci. 14:622759. doi: 10.3389/fnins.2020.622759

[9] . Mai, T.D.T., Phung, TN. (2023). Evaluating the Performance of Some Deep Learning Model for the Problem of Emotion Recognition Based on EEG Signal. In: Nghia, P.T., Thai, V.D., Thuy, N.T., Son, L.H., Huynh, VN. (eds) Advances in Information and Communication Technology. ICTA 2023. Lecture Notes in Networks and Systems, vol 847. Springer, Cham. https://doi.org/10.1007/978-3-031-49529-8_19

[10] .Garg, Neha, and Kamlesh Sharma. 2023. ‘Feature Extraction for Emotion Recognition: A Review’. Emotion Recognition - Recent Advances, New Perspectives and Applications. IntechOpen. doi:10.5772/intechopen.109740.

[11] .Abdulrahman, Awf and Muhammet Baykara. “A Comprehensive Review for Emotion Detection Based on EEG Signals: Challenges, Applications, and Open Issues.” Traitement du Signal 38 (2021): 1189-1200.

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

[13] .Samia Mezzah, Abdelmalek Tari, Practical hyperparameters tuning of convolutional neural networks for EEG emotional features classification, Intelligent Systems with Applications, Volume 18, 2023, 200212, ISSN 2667-3053, https://doi.org/10.1016/j.iswa.2023.200212.