 https://doi.org/10.24191/jeesr.v29i1.013

Evaluating Deep Learning Models for EEG-Based Emotion Recognition Using the SEED-IV Dataset

Ali Ahmed Hasan Bahbry and Swee Kheng Eng*

***Abstract*—Emotion recognition using electroencephalography (EEG) has attracted growing interest in affective computing due to its ability to capture neural responses associated with emotional states. However, the variability and complexity of EEG signals pose significant challenges for accurate classification. This study evaluates the performance of three deep learning models—Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM), Temporal Convolutional Network (TCN), and Bidirectional Long Short-Term Memory (BiLSTM)—for EEG-based emotion recognition using the publicly available SEED-IV dataset. Differential Entropy (DE) and Power Spectral Density (PSD) features were extracted from preprocessed EEG signals and used to train and evaluate the models under identical conditions. Experimental results demonstrated that the BiLSTM model delivered superior performance, achieving the highest testing accuracies of 90.38% (DE) and 88.46% (PSD). The TCN model also showed strong generalization and competitive results, securing testing accuracies of 88.03% (DE) and 85.03% (PSD), whereas the CNN-LSTM model yielded the lowest comparative results at 82.85% (DE) and 80.85% (PSD). The findings highlight the effectiveness of temporal deep learning models for capturing dynamic EEG patterns and confirm the suitability of DE and PSD features for emotion recognition tasks.**

***Index Terms*—BiLSTM, CNN-LSTM, deep learning, differential entropy, EEG, emotion recognition, power spectral density, SEED-IV dataset, TCN.**

I. Introduction

Electroencephalography (EEG) is a widely used non-invasive technique for measuring the brain's electrical activity, offering high temporal resolution and direct access to neural signals. It has become an essential tool in emotion recognition systems, especially in affective computing, due to its ability to capture rapid changes in brain activity in response to emotional stimuli [1]. EEG signals reflect the neural dynamics associated with various affective states, such as happiness, sadness, fear, and neutrality, across different frequency bands including delta, theta, alpha, beta, and gamma [2], [3]. Despite its potential, EEG-based emotion recognition remains challenging due to the non-stationary nature of EEG signals, inter-subject variability, and the influence of noise and artifacts [2]. As such, effective preprocessing and meaningful feature extraction are essential to enhance signal quality and model performance. Two commonly used features are Differential Entropy (DE), which captures the complexity of the EEG signal in a statistical sense, and Power Spectral Density (PSD), which reflects the distribution of signal energy across frequency bands [3], [4].

Recent advances in deep learning have enabled more accurate and automated classification of emotional states from EEG signals. Deep neural architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Bidirectional LSTMs (BiLSTMs) have shown strong capabilities in learning spatial and temporal dependencies within EEG data [1],[5]. Temporal Convolutional Networks (TCNs) have also been explored as an efficient alternative for sequence modeling [6].

This manuscript is submitted on January 22, 2026, revised on March 23, 2026, accepted on May 8, 2026 and published on October 31, 2026.

Ali Ahmed Hasan Bahbry was a student of the Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Pauh Putra Campus 02600 Arau, Perlis, Malaysia (e-mail: elewa.a17@gmail.com).

Eng Swee Kheng is with Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis and a fellow at the Centre of Excellence for Intelligent Robotics and Autonomous System (CIRAS), Universiti Malaysia Perlis (e-mail: kheng@unimap.edu.my).

*Corresponding author
Email address: kheng@unimap.edu.my

1985-5389/© 2026 The Authors. Published by UiTM Press. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

II. Literature Review

EEG-based emotion recognition has attracted increasing attention due to its potential applications in human-computer interaction, mental health monitoring, and affective feedback systems. Over the past decade, several studies have focused on improving the accuracy and efficiency of emotion classification from EEG signals using both traditional and deep learning models. Differential Entropy and Power Spectral Density have emerged as some of the most frequently used features. Gao et al. [7] employed a feature fusion network that integrated DE and PSD for robust EEG classification, showing that combining complementary features enhances generalization.

Deep learning models have been particularly effective in this domain. Liu et al. [5] developed a hybrid approach using Sparse Autoencoders and Convolutional Neural Networks, achieving over 96% accuracy on the SEED dataset [8]. Similarly, Subasi et al. [9] combined Tunable Q Wavelet Transform with a Rotation Forest Ensemble classifier, attaining strong results while maintaining computational efficiency. Their model outperformed conventional machine learning methods when tested on the SEED dataset [8].

Zhang et al. [4] investigated various deep learning configurations and emphasized the advantages of hybrid architectures that integrate both convolutional and recurrent layers. Their findings indicated that models like CNN-LSTM were particularly effective for capturing both spatial and temporal aspects of EEG signals. In another notable work, Alhagry et al. [1] utilized an LSTM Recurrent Neural Network,

achieving high classification accuracy using DE features extracted from the SEED dataset [8]. Their study reaffirmed the effectiveness of temporal models in decoding EEG signals for emotion classification.

Despite these advancements, comparative studies across models and feature types remain limited. Differences in datasets, preprocessing steps, and feature extraction pipelines often make it difficult to benchmark architectures on a consistent basis. To bridge this gap, this study conducts a direct comparison of CNN-LSTM, TCN, and BiLSTM using the same preprocessing steps and EEG feature sets derived from the SEED-IV dataset [8].

III. Methodology

This study follows a structured methodology as shown in Fig. 1. which begins with acquiring EEG data from a public dataset, followed by signal preprocessing and feature extraction. The extracted features are used to train and evaluate multiple deep learning models for emotion classification. The process includes splitting the data for training, validation, and testing, ensuring a fair comparison of model performance and selection of the most accurate approach.

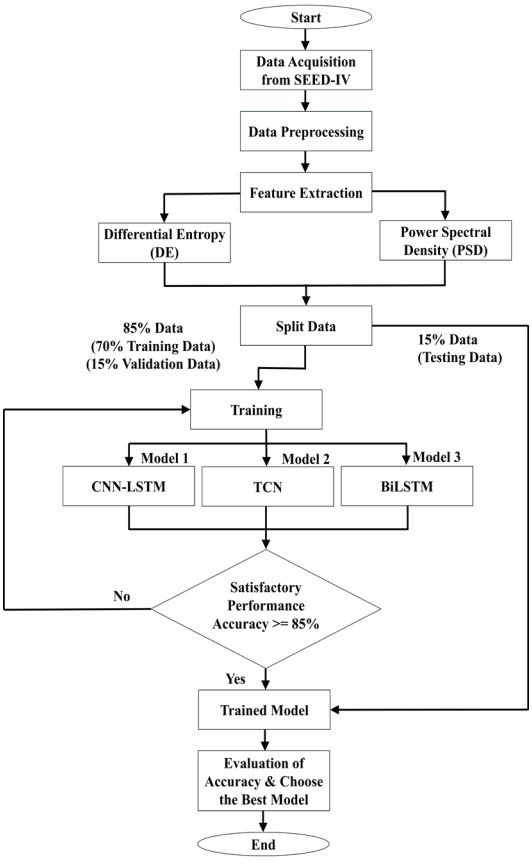


Fig. 1. Methodology Flowchart.

A. Dataset and Preprocessing

The EEG data used in this study were obtained from the publicly available SEED-IV dataset [8]. This dataset contains EEG recordings collected from 15 participants while they watched video clips designed to evoke four distinct emotional states: happiness, sadness, fear, and neutrality. The recordings were made using a 62-channel EEG headset at a sampling rate of 1000 Hz as shown in Fig. 2. Each subject completed multiple sessions, with 24 labeled trials per session. These data were used without any additional data collection, ensuring the study remained entirely computational.

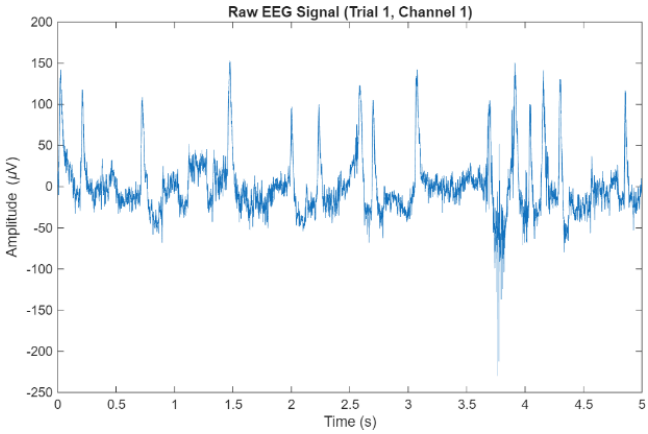


Fig. 2. Sample of dataset.

To enhance signal quality and reduce noise, several preprocessing steps were applied. First, a band-pass filter with a passband of 1 Hz to 75 Hz was used to retain the frequency components relevant to neural activity while eliminating low-frequency drifts and high-frequency noise. Next, the EEG signals were down sampled from 1000 Hz to 200 Hz as shown in Fig. 3. to reduce computational complexity while preserving essential signal characteristics.

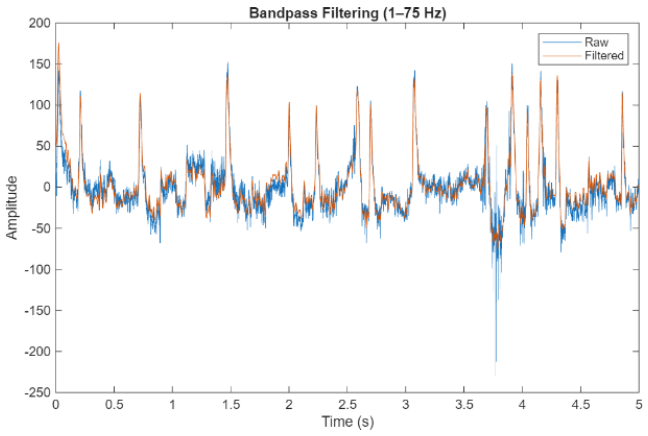


Fig. 3. Sample of filtered dataset.

Following this, z-score normalization was applied on a per-channel basis for each trial as shown in Fig. 4. This normalization step ensured that each channel's values had zero mean and unit variance, helping to standardize the data and reduce variability due to subject-specific and session-specific factors.

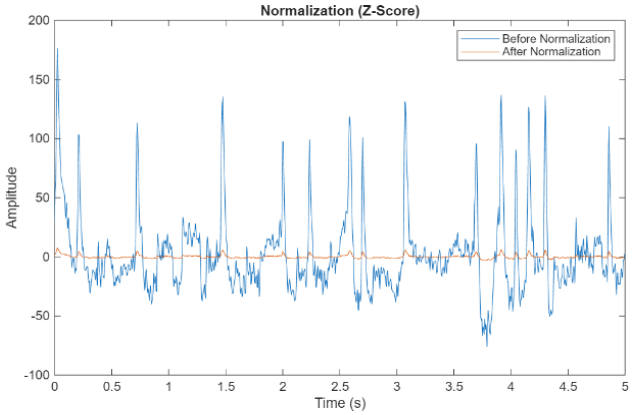


Fig. 4. Dataset after normalization.

B. Feature Extraction

Two types of features were extracted from the preprocessed EEG data: Differential Entropy (DE) and Power Spectral Density (PSD). The EEG signals were segmented using a 2-second sliding window with 1-second overlap. For each segment, the signal was filtered into five frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz) as shown in Fig. 5. For DE features, log-variance was calculated to represent the signal's power distribution in each frequency band. For PSD features, Welch's method was used to compute the energy content within each band. These feature sets were stored as 3D arrays consisting of channels, time windows, and frequency bands, forming the input data for training.

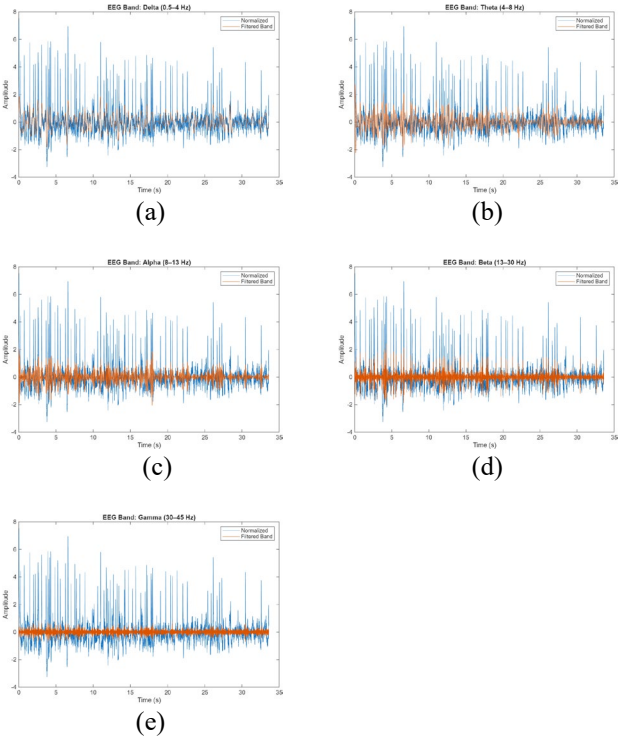


Fig. 5. Five Frequency bands (a) Delta (b) Theta (c) Alpha (d) Beta (e) Gamma.

C. Training and Evaluation

To effectively address the challenges of temporal EEG modeling and decode the complex dynamic patterns inherent in brain signals, this study selected three distinct deep learning architectures: a hybrid CNN-LSTM, a Temporal Convolutional Network (TCN), and a Bidirectional LSTM (BiLSTM). The CNN-LSTM model was chosen because EEG data inherently consists of both rapid, short-term neural oscillations and slower, evolving emotional states, a hybrid approach that has shown high robustness in recent EEG studies [4]. In this hybrid architecture, the convolutional layers extract localized short-term temporal features and reduce noise, while the subsequent LSTM layer captures the global sequential dependencies, effectively modeling how these emotional EEG patterns evolve over the entire time series. The TCN was introduced as a highly efficient alternative to standard recurrent networks specifically for temporal modeling [10]. By utilizing dilated causal convolutions, TCNs can effectively capture long-range, multi-scale temporal dependencies without increasing computational complexity or accessing future time steps. Finally, the BiLSTM was selected to maximize temporal modeling capabilities beyond unidirectional approaches, which has been proven highly effective for sequential EEG emotion classification [11]. Because emotional EEG patterns often depend on both preceding and succeeding neural events, the BiLSTM's bidirectional structure is highly advantageous, allowing it to process temporal information from both past and future contexts simultaneously.

To ensure a rigorous and unbiased evaluation, the entire dataset was partitioned using a trial-wise, stratified splitting strategy. Rather than separating the data by subject (subject-wise), all preprocessed EEG trials were pooled and then split into training (70%), validation (15%), and testing (15%) sets. All three models, CNN-LSTM, Temporal Convolutional Network (TCN), and Bidirectional LSTM (BiLSTM)—were trained using 70% of the dataset, with 15% allocated for validation and the remaining 15% for testing. The training process was conducted over 50 epochs using the Adam optimizer with a mini-batch size of 16. Throughout the training, accuracy and loss values were continuously monitored to ensure proper convergence and to detect any potential signs of overfitting. The performance of each model was evaluated using accuracy of the model, In addition, confusion matrices were generated to examine how each model distributed its predictions across the four target emotional classes. Each model-feature combination-whether using Differential Entropy (DE) or Power Spectral Density (PSD)-was tested independently under the same training conditions to ensure a fair and consistent comparison.

IV. Results and Discussion

The final training and validation performance of each model was evaluated for both DE and PSD features, as shown in Table I and II. For the CNN-LSTM model using DE features, the training accuracy plateaued at 82.14%, accompanied by a validation accuracy of 85.71%. When trained on PSD features,

the model slightly improved its training accuracy to 82.94%, though validation performance dropped to 71.43%, suggesting potential overfitting or sensitivity to PSD representations.

The TCN model displayed excellent training behavior for both DE and PSD; their training accuracies reached 92.46% and 90.41%, respectively, with stable validation accuracies of 90.41% for DE and 90.17% for PSD. These results confirm the TCN's ability to generalize well across feature types, owing to its dilated convolutional architecture that efficiently captures multi-scale temporal dependencies.

BiLSTM achieved strong, consistent learning across both feature sets. With DE, it attained a training accuracy of 90.87% and a validation accuracy of 90.43%. The PSD feature set led to a slightly higher training accuracy of 92.06%, while maintaining a strong validation accuracy of 92.86%. These trends highlight BiLSTM's ability to learn complex temporal relationships in EEG signals, benefiting from its bidirectional structure that processes information from both past and future contexts.

TABLE I. RESULTS OF TRAINING FOR THREE MODELS

	Training		
	CNN-LSTM	**TCN**	**BiLSTM**
DE	82.14%	92.46%	90.87%
PSD	82.94%	91.07%	92.06%

TABLE II. RESULTS OF VALIDATION FOR THREE MODELS

	Validation		
	CNN-LSTM	**TCN**	**BiLSTM**
DE	85.71%	90.41%	90.43%
PSD	71.43%	90.17%	92.86%

Table III. summarizes the testing performance of each model using both DE and PSD features. The BiLSTM model demonstrated superior performance, achieving the highest testing accuracies of 90.38% with DE features and 88.46% with PSD features. The TCN model also showed strong generalization, securing the second-highest results with 88.03% for DE and 85.03% for PSD. In contrast, the CNN-LSTM model yielded the lowest comparative results, with 82.85% accuracy for DE and 80.85% for PSD. Across all architectures, the DE feature set consistently outperformed PSD, indicating that Differential Entropy provides more discriminative information for this specific classification task.

TABLE III. RESULTS OF TESTING FOR THREE MODELS

	Testing		
Model	**CNN-LSTM**	**TCN**	**BiLSTM**
DE	82.85%	88.03%	90.38%
PSD	80.85%	85.03%	88.46%

A. Discussion on Feature Representations (DE vs. PSD)

The interaction between the deep learning architectures and the feature representations significantly influenced classification performance, with the Differential Entropy (DE) feature set consistently supporting stronger generalization compared to Power Spectral Density (PSD). The BiLSTM model achieved the highest testing accuracy when paired with DE features. Its bidirectional structure, which processes information from both past and future contexts, perfectly complemented the complex temporal relationships inherent in the DE signals. Furthermore, the TCN model demonstrated robust generalization across both DE and PSD feature types. This adaptability is attributed to its dilated convolutional architecture, which efficiently captures multi-scale temporal dependencies, allowing its deep receptive field to adapt equally well to spectral and entropy-based representations. Conversely, the CNN-LSTM model yielded the lowest comparative results, particularly with PSD features. The model's localized convolutional layers struggled to exploit long-sequence frequency cues from PSD, leading to potential overfitting, whereas it generalized moderately better with the time-domain robustness of DE.

B. Comparison with Previous Studies

To place these findings in context, the obtained results compare favorably with previous benchmark studies evaluated on the SEED-IV dataset. For example, prior works using Differential Entropy (DE) features have reported highest average emotion recognition accuracy of 81.90% using Semi-Supervised Regression with Adaptive Graph Learning [12] and 84.21% using Bidirectional Long Short-Term Memory (BiLSTM) networks [11]. By comparison, the BiLSTM framework implemented in this study achieved a superior testing accuracy of 90.38% utilizing DE features. This improvement highlights that the standardized preprocessing, feature extraction, and temporal modeling pipeline employed in this study is highly competitive for decoding EEG-based emotional states.

V. CONCLUSION

This study presented a comparative analysis of three deep learning models—CNN-LSTM, Temporal Convolutional Network (TCN), and Bidirectional LSTM (BiLSTM)—for EEG-based emotion recognition using the SEED-IV dataset [8]. The models were evaluated using two feature extraction techniques: Differential Entropy (DE) and Power Spectral Density (PSD), both derived from preprocessed EEG signals.

All models were trained and tested under standardized conditions to ensure fair comparison across feature types and architectures.

The results demonstrated that both TCN and BiLSTM achieved strong and consistent performance across DE and PSD features, with BiLSTM reaching the highest test accuracy of 90.38% using DE. CNN-LSTM demonstrated moderate generalization, with test accuracies of 82.85% for DE and 80.85% for PSD features. These findings indicate that temporal modeling, particularly through bidirectional and dilated convolutional structures, plays a critical role in accurately decoding emotional states from EEG signals.

The outcomes highlight the effectiveness of using deep learning models for automated emotion classification and reinforce the potential of EEG-based systems in real-world affective computing applications. Future efforts may focus on expanding the dataset, integrating multimodal signals, and optimizing models for real-time deployment.

ACKNOWLEDGMENT

The authors would like to acknowledge the Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis for the continuous support for this research.

REFERENCES

[1] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, “Emotion Recognition based on EEG using LSTM Recurrent Neural Network,” 2017.

[2] A. N. N. M. Yosi, K. A. Sidek, H. S. Yaacob, M. Othman, and A. Z. Jusoh, “Emotion recognition using electroencephalogram signal,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 2, pp. 786–793, Aug. 2019, doi: 10.11591/ijeecs.v15.i2.pp786-793.

[3] N. S. Suhaimi, J. Mountstephens, and J. Teo, “EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities,” 2020, *Hindawi Limited*. doi: 10.1155/2020/8875426.[4] Y. Zhang *et al.*, “An Investigation of Deep Learning Models for EEG-Based Emotion Recognition,” *Front Neurosci*, vol. 14, Dec. 2020, doi: 10.3389/fnins.2020.622759.

[5] J. Liu *et al.*, “EEG-Based Emotion Classification Using a Deep Neural Network and Sparse Autoencoder,” *Front Syst Neurosci*, vol. 14, Sep. 2020, doi: 10.3389/fnsys.2020.00043.

[6] M. Yanagimoto and C. Sugimoto, “Convolutional neural networks using supervised pre-training for EEG-based emotion recognition,” 2016.

[7] Q. Gao, Y. Yang, Q. Kang, Z. Tian, and Y. Song, “EEG-based Emotion Recognition with Feature Fusion Networks,” International Journal of Machine Learning and Cybernetics, vol. 13, no. 2, pp. 421–429, Feb. 2022, doi: 10.1007/s13042 021-01414-5.

[8] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki, EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. IEEE Transactions on Cybernetics, Volume: 49, Issue: 3, March 2019, Pages: 1110-1122, DOI: 10.1109/TCYB.2018.2797176.

[9] A. Subasi, T. Tuncer, S. Dogan, D. Tanko, and U. Sakoglu, “EEG-based emotion recognition using tunable Q wavelet transform and rotation forest ensemble classifier,” Biomed Signal Process Control, vol. 68, Jul. 2021, doi: 10.1016/j.bspc.2021.102648.

[10] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint arXiv:1803.01271, 2018.

[11] J. Yang, X. Huang, H. Wu, and X. Yang, “EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network,” in Procedia Computer Science, Elsevier B.V., 2020, pp. 491–504. doi: 10.1016/j.procs.2020.06.117.

[12] T. Sha, Y. Zhang, Y. Peng, and W. Kong, “Semi-supervised regression with adaptive graph learning for EEG-based emotion recognition,” Mathematical Biosciences and Engineering, vol. 20, no. 6, pp. 11379–11402, 2023, doi: 10.3934/mbe.2023505.