International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

AI-Powered Emotion Detection from EEG Signals: An Experimental Comparison of CNN, LSTM, and GRU Architectures

Naveen Saini

Department of Computer Science and Engineering

Abstract:
Recognising emotion from brain activity is an increasingly active area of affective computing, since electroencephalography (EEG) reflects an internal physiological state that is far harder to disguise than a facial expression or a spoken word. This paper reports a controlled experimental comparison of three deep learning architectures — a one-dimensional Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) network, and a Gated Recurrent Unit (GRU) network — trained under an identical configuration to classify EEG-derived features into negative, neutral, and positive emotional states. All three models were trained and tested on the same 2,132-sample Kaggle EEG Brainwave dataset, recorded with a 14-channel Emotiv EPOC headset, so that the resulting accuracy figures reflect the architecture itself rather than differing experimental conditions. Evaluated on accuracy, precision, recall, F1-score, and confusion-matrix analysis, the GRU model reached the highest test accuracy (98.1%), followed by LSTM (96.2%) and CNN (95.1%), with all three models converging smoothly and generalising well to the held-out test set. These findings suggest that gated recurrent networks model the temporal structure of EEG signals more effectively than a purely convolutional model, while remaining substantially lighter than the graph- and Transformer-based architectures increasingly reported in this field.

Keywords: EEG, Emotion Recognition, Deep Learning, CNN, LSTM, GRU, Affective Computing, Comparative Study.

1. Introduction

Human emotion is difficult to measure directly, so most computing systems infer it indirectly — from a smile, a raised voice, or a typed sentence — channels a person can just as easily suppress, exaggerate, or express differently depending on context and culture. A system that relies only on these outward cues inherits their unreliability.

Brain activity offers a more direct alternative. Scalp-recorded EEG captures electrical activity generated by cortical neurons; because the signal arises involuntarily, it cannot be consciously edited the way a facial expression can, and it is far less expensive to record than functional imaging techniques such as fMRI. Practical EEG emotion-recognition systems typically classify power or statistical features drawn from the delta, theta, alpha, beta, and gamma bands, and the resulting technology now supports

International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

applications from mental-health monitoring to adaptive interfaces, gaming, tutoring systems, and neurofeedback (Fig. 1).

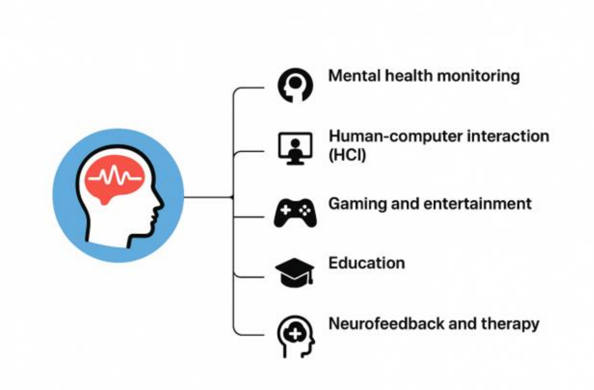

Fig. 1. Representative applications of EEG-based emotion recognition

1.1 Related Work

Convolutional and hybrid CNN-recurrent designs form one major thread of this literature. Li et al. [1] combined a convolutional front end with an LSTM back end so that spatial correlations across electrodes and their evolution over a trial could be learned within one pipeline, a template later extended by Yang et al.'s parallel CNN-RNN design [2], which processes convolutional and recurrent branches separately before merging them for classification. Salama et al. [3] reshaped multi-channel EEG into a 3-D volume for a 3-D CNN, while Hosseini et al. [4] paired a similar CNN-LSTM pipeline with personality features to push accuracy above 95% on a denser electrode set. At the lighter end of this design space, Aldawsari et al. [5] showed that a 1-D CNN operating directly on single-channel time series can trade a little peak accuracy for a substantially smaller model.

A second thread compares recurrent architectures more directly. Chowdary et al. [6] used a plain RNN as a lower-complexity baseline, and Sailaja and Kumar Rao [7] benchmarked LSTM, GRU, and a plain DNN against several CNN-hybrid and graph-based alternatives, finding that the simpler gated models reached accuracy within a few points of far more elaborate options. Davarzani et al. [8] reached a similar conclusion comparing shallow classifiers against CNN and RNN-LSTM models, Nguyen's Echo-GRU [9] paired an Echo State Network with a GRU for a wearable, low-electrode setting, and Pamungkas et al. [10] showed through a systematic hyperparameter search that a well-tuned GRU can match or exceed LSTM accuracy while training faster.

A third thread models electrode relationships or attention explicitly. Song et al.'s Dynamical Graph Convolutional Network [11] treats EEG channels as nodes in a learned graph, Setiawan et al. [12] compared several graph-based variants on the harder five-class SEED-V benchmark, Devarajan et al. [13] deployed a hybrid CNN-Transformer model on edge hardware, and Cheng et al.'s gated-Transformer MSDCGTNet [14] combines a multi-scale CNN with attention for very high reported accuracy — at a consistently higher computational cost than a single recurrent branch. Classical machine-learning baselines remain a useful reference point: Anagha Prakash et al. [15] found that none of ten classical classifiers matched deep-learning accuracy on comparable features, and Shashi Kumar et al. [16] reached competitive accuracy only after restricting analysis to a small set of frontal electrodes.

Despite this activity, few studies train CNN-only, LSTM-only, and GRU-only branches under identical preprocessing and hyperparameters purely to isolate the architecture's own effect on accuracy, and

International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

computational cost is rarely reported alongside it. This paper addresses that gap by training and evaluating all three architectures under one shared configuration, comparing them with standard classification metrics and confusion-matrix analysis, to identify which offers the best accuracy for a lightweight, potentially real-time emotion-recognition system.

2. Methodology

2.1 Dataset and Preprocessing

This study uses the publicly available "EEG Brainwave Dataset: Feeling Emotions" from Kaggle, recorded with a 14-channel Emotiv EPOC headset under the international 10–20 electrode layout. Participants viewed emotionally evocative stimuli while EEG was sampled at 128 Hz; each labelled instance provides pre-extracted frequency-domain features rather than a raw voltage trace, so no additional filtering or artefact-removal step was required. The dataset holds 2,132 samples split almost evenly across three classes (Table 1), limiting the risk of a majority-class bias.

Table 1. Class Distribution of the Dataset

Label	Sample Count
Neutral	716
Negative	708
Positive	708
Total	2,132

Labels were one-hot encoded and features standardised with a StandardScaler; each sample was reshaped to a (128, 1) sequence to match the Conv1D/LSTM/GRU input format. Data were split 80:20 into train and test sets (stratified), with the training portion further split 70:30 into training and validation subsets.

2.2 Model Architectures

Figure 2 shows the shared pipeline, from stimulus and EEG acquisition to the final Negative/Neutral/Positive prediction; the three models differ only in the network placed between the standardised input and the softmax output.

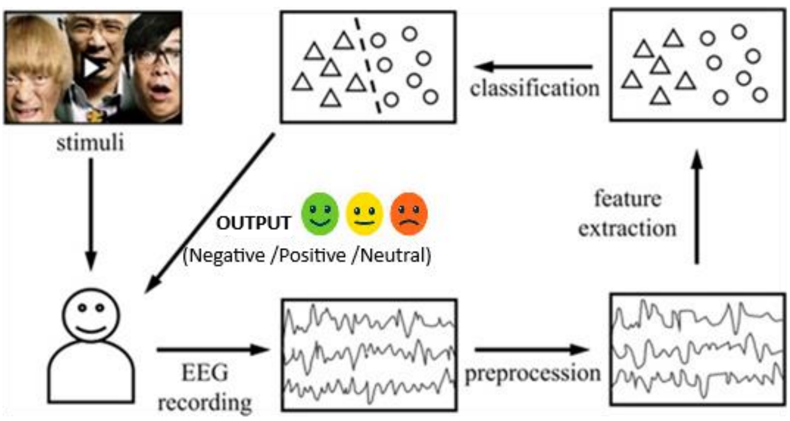

Fig. 2. Proposed emotion-classification workflow using EEG-derived features.

The CNN branch stacks two Conv1D blocks (64 and 128 filters, kernel size 3, ReLU) with batch normalisation and max-pooling, then two dense layers (128 and 64 units) with 0.3 dropout before a three-unit softmax layer — effective at extracting local patterns but with no explicit mechanism for tracking how those patterns evolve over time.

The LSTM branch replaces the convolutional blocks with two stacked LSTM layers (128 then 64 units, dropout 0.3) followed by a 64-unit dense layer, carrying a cell state forward across time steps to capture longer-range temporal dependence.

The GRU branch mirrors the LSTM branch layer-for-layer but uses a GRU cell, which merges the input and forget gates into a single update gate — reducing parameter count while retaining the ability to model sequential dependence.

2.3 Training Configuration and Evaluation

All models were built with the Keras Sequential API and trained in Google Colab on an NVIDIA Tesla T4 GPU under an identical configuration (Table 2): categorical cross-entropy loss, Adam optimiser (learning rate 0.001), batch size 32, up to 20 epochs, with early stopping (patience 10), checkpointing, and a learning-rate scheduler to limit overfitting.

Table 2. Training Configuration and Hyperparameters

Parameter	Value
Loss function	Categorical cross-entropy
Optimiser	Adam
Learning rate	0.001
Batch size	32
Maximum epochs	20
Dropout rate	0.3
Callbacks	Early stopping, model checkpoint, LR scheduler
Train / Val / Test split	56% / 24% / 20% (80:20, then 70:30)
Hardware	NVIDIA Tesla T4 GPU (12 GB)

Performance was measured with accuracy, precision, recall, and F1-score (per-class, macro-, and weighted-averaged), together with normalised confusion matrices to visualise class-wise misclassification.

3. Results

3.1 Training Behaviour

All three models converged smoothly within the 20-epoch budget. CNN training accuracy reached close to 99% against roughly 96% validation accuracy, with both loss curves stabilising quickly; LSTM triggered early stopping at epoch 14, with validation accuracy tracking training accuracy closely throughout; and GRU showed the most favourable profile of the three (Fig. 3), with validation accuracy climbing to around 98.5% and only a small, stable gap between training and validation loss — evidence of genuine generalisation rather than memorisation.

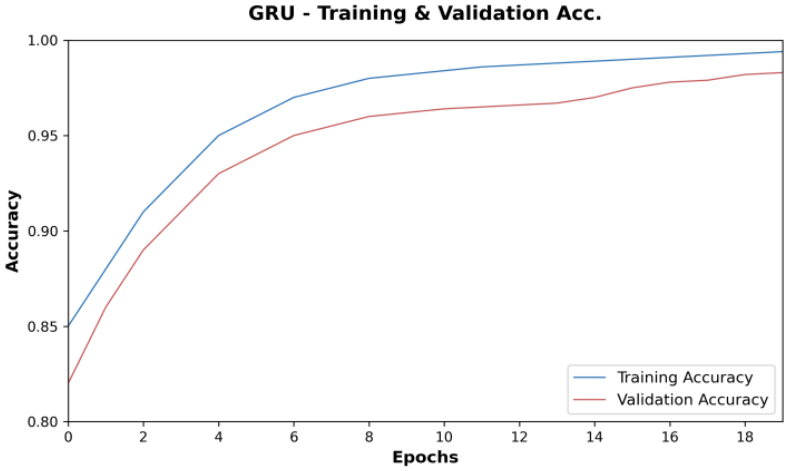

Fig. 3. GRU training and validation accuracy across epochs (representative of the best-performing model).

3.2 Classification Performance

Table 3 reports per-class precision, recall, and F1-score on the 640-sample test set, and Table 4 summarises overall accuracy and macro-averaged metrics.

Table 3. Per-Class Classification Report for CNN, LSTM, and GRU Models

Model	Class	Precision	Recall	F1-Score
CNN	Negative	0.93	0.96	0.94
CNN	Neutral	0.97	0.95	0.96
CNN	Positive	0.94	0.93	0.93
LSTM	Negative	0.94	0.98	0.96
LSTM	Neutral	1.00	0.97	0.98
LSTM	Positive	0.95	0.94	0.94
GRU	Negative	0.97	0.99	0.98
GRU	Neutral	0.99	0.98	0.98
GRU	Positive	0.98	0.97	0.97

Table 4. Comparative Overall Performance of CNN, LSTM, and GRU Models

Model	Accuracy (%)	Macro Precision	Macro Recall	Macro F1-Score
CNN	95.1	0.947	0.947	0.943
LSTM	96.2	0.963	0.963	0.960
GRU	98.1	0.980	0.980	0.977

GRU obtained the highest overall accuracy (98.1%) and the highest macro-averaged scores of the three architectures, ahead of LSTM (96.2%) and CNN (95.1%). All three models were weakest on the Positive class, most often confused with Neutral — CNN, LSTM, and GRU classified 95%, 94%, and 98% of Positive samples correctly, respectively (Fig. 4) — while Negative samples were the easiest to separate throughout, consistent with more distinctive EEG patterns for that class in this dataset.

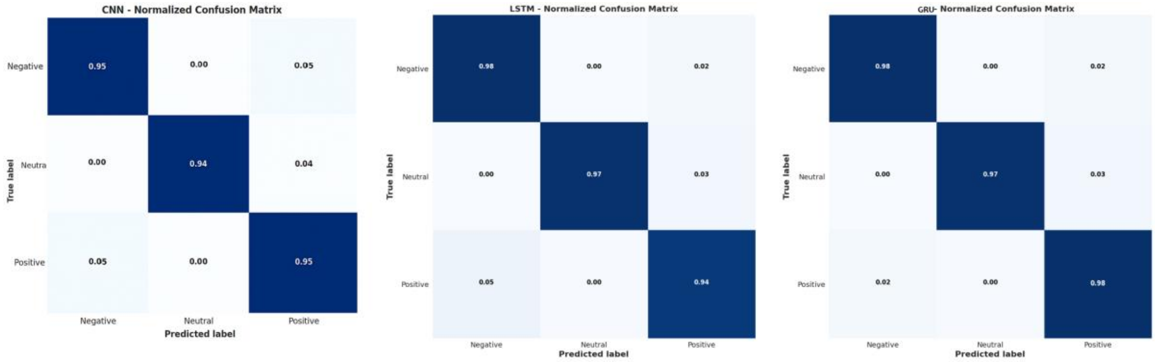

Fig. 4. Normalised confusion matrices for the CNN, LSTM, and GRU models on the test set.

4. Discussion

The consistent ranking — GRU ahead of LSTM ahead of CNN — matches a broader pattern in this literature: architectures that explicitly model how a signal evolves over time tend to outperform ones that treat each input as a static vector, and GRU's simpler gating appears to offer a genuine edge over LSTM's more elaborate gating here, not merely a computational saving. This is consistent with Pamungkas et al.'s [10] finding that a well-tuned GRU can match or exceed LSTM accuracy while training faster, and with Sailaja and Kumar Rao's [7] comparison, where lightweight recurrent models reached accuracy close to considerably more elaborate alternatives.

Set against the wider literature, the 98.1% GRU accuracy obtained here compares favourably with several more complex designs on related tasks. Nguyen's Echo-GRU [9], evaluated on a dataset of identical scale and class structure, reported 95.58% accuracy — over two points below the plain GRU used here. Graph-based methods such as Song et al.'s DGCNN [11] and Setiawan et al.'s GCN/GAT comparison [12] reach comparable, or on the harder five-class SEED-V benchmark marginally higher, accuracy at substantially higher structural cost, and Transformer-based designs show a similar pattern: Devarajan et al.'s edge-deployed CNN-Transformer [13] reached only about 87% accuracy, while Cheng et al.'s MSDCGTNet [14] reports very high accuracy at considerably higher implementation complexity. Classical machine-learning baselines fare worse still [15], [16], and earlier CNN-LSTM hybrids [3], [4] report accuracy broadly consistent with the CNN and LSTM figures obtained here, without matching the GRU branch. Taken together, this suggests the accuracy gap observed here (95.1–98.1%) is not an artefact of an unusually favourable dataset, and supports treating a well-tuned GRU as a strong, resource-efficient default rather than assuming architectural complexity guarantees better performance.

These results should, however, be read against a single, comparatively small, laboratory-collected, three-class dataset evaluated offline on a fixed split; cross-subject and real-time performance remain untested, and none of the three architectures offers insight into which features drove a given prediction.

5. Conclusion

This study compared CNN, LSTM, and GRU architectures for three-class EEG emotion recognition under a single controlled training configuration. The GRU model achieved the highest test accuracy (98.1%), ahead of LSTM (96.2%) and CNN (95.1%), with all three models converging stably and generalising well to the held-out test set — evidence that gated recurrent architectures capture the temporal structure of EEG signals more effectively than a purely convolutional model, while remaining considerably cheaper to train than graph- or Transformer-based alternatives.

For practitioners building resource-constrained or real-time emotion-aware systems, these results support a well-tuned GRU as a strong default choice. Future work should test cross-subject generalisation on larger, more diverse datasets, explore multimodal fusion with other physiological signals, and incorporate explainable-AI techniques so that predictions can be interpreted alongside their accuracy.

REFERENCES:

[1] X. Li, D. Song, P. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition using a combined CNN and LSTM network," 2016.

[2] J. Yang, Q. M. J. Wu, W. Zheng, and B. Lu, "Parallel convolutional recurrent neural network for EEG-based emotion recognition," 2018.

[3] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby Shalaby, "3D-convolutional neural network for EEG-based emotion recognition using the DEAP dataset," International Journal of Advanced Computer Science and Applications.

[4] M. S. K. Hosseini, S. M. Firoozabadi, K. Badie, and P. Azadfallah, "Personality-Based Emotion Recognition Using EEG Signals with a CNN-LSTM Network," Brain Sciences, vol. 13, no. 6, p. 947, 2023.

[5] H. Aldawsari, S. Al-Ahmadi, and F. Muhammad, "Lightweight one-dimensional convolutional neural network for EEG-based emotion classification," 2023.

[6] M. K. Chowdary, J. Anitha, and D. J. Hemanth, "Emotion Recognition from EEG Signals Using Recurrent Neural Networks," Electronics, vol. 11, no. 15, p. 2387, Jul. 2022, doi: 10.3390/electronics11152387.

[7] G. Sailaja and N. B. Kumar Rao, "Emotion Detection from EEG Using Advanced Deep Neural Architectures: LSTM, GRU, and DNN Approaches," Traitement du Signal (IIETA), vol. 42, no. 6, pp. 3615–3625, Dec. 2025, doi: 10.18280/ts.420643.

[8] S. Davarzani, S. Masihi, M. Panahi, A. Olalekan Yusuf, and M. Atashbar, "A Comparative Study on Machine Learning Methods for EEG-Based Human Emotion Recognition," Electronics, vol. 14, no. 14, p. 2744, Jul. 2025, doi: 10.3390/electronics14142744.

[9] Q.-T. Nguyen, "Echo-GRU: Emotion Recognition Using Wearable EEG Supporting Early Alzheimer's Disease Detection," in Intelligent Systems Conference (IntelliSys), pp. 570–586, Springer, 2025, doi: 10.1007/978-3-031-87657-8_1.

[10] Y. Pamungkas, S. Pratasik, M. Krisnanda, and P. N. Crisnapati, "Optimizing Gated Recurrent Unit Architecture for Enhanced EEG-Based Emotion Classification," J. Robot. Control (JRC), vol. 6, no. 3, pp. 1450–1461, May 2025.

[11] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," IEEE Transactions on Affective Computing, vol. 11, no. 3, pp. 532–541, 2020, doi: 10.1109/TAFFC.2018.2817622.

[12] A. Setiawan, Y. Devianto, C. Dewi, and I. Sembiring, "Integration of artificial intelligence and graph neural networks (GNN) for modeling human emotions based on electroencephalogram (EEG) signals," Array, vol. 30, 2026, art. 100950, doi: 10.1016/j.array.2026.100950.

[13] K. Devarajan, S. Ponnan, and S. Perumal, "Hybrid CNN-transformer architecture for enhanced EEG-based emotion recognition: capturing local and global dependencies with self-attention mechanisms," Discover Computing, vol. 28, art. 87, 2025, doi: 10.1007/s10791-025-09596-0.

[14] S. Cheng et al., “EEG-based emotion recognition using multi-scale dynamic CNN and gated transformer (MSDCGTNet),” Scientific Reports, vol. 14, art. 31207, 2024, doi: 10.1038/s41598-024-82705-z.
[15] Anagha Prakash and Alwin Poulose, “Electroencephalogram-based emotion recognition: a comparative analysis of supervised machine learning algorithms,” Data Science and Management, vol. 8, no. 3, pp. 342–360, 2025, doi: 10.1016/j.dsm.2024.12.004.
[16] Shashi Kumar G S, Niranjana Sampathila, and Tanishq Tanmay, “Wavelet based machine learning models for classification of human emotions using EEG signal,” Measurement: Sensors, vol. 24, 2022, art. 100554, doi: 10.1016/j.measen.2022.100554.