ORIGINAL ARTICLE

Channel-Reduced Hybrid Deep Neural Networks for High-Accuracy EEG-based Emotion Recognition

Mahdi Jafari Asl [1,2,3], Sina Shamekhi [1,2,3*]

[1] *Faculty of Biomedical Engineering, Sahand University of Technology, Sahand New Town, Tabriz, Iran*

[2] *Biomedical Engineering Research Center, Sahand University of Technology, Sahand New Town, Tabriz, Iran*

[3] *Brain Mapping Laboratory, Faculty of Biomedical Engineering, Sahand University of Technology, Sahand New Town, Tabriz, Iran*

*Corresponding Author: Sina Shamekhi

Received: 23 July 2025 / Accepted: 25 April 2026

Email: shamekhi@sut.ac.ir

Abstract

Purpose: Emotion detection using Electroencephalogram (EEG) signals has attracted increasing research attention due to its potential applications in affective computing and brain–computer interfaces. However, the inherently non-linear and non-stationary nature of EEG signals poses major challenges for designing a deep network framework capable of achieving high recognition accuracy. To address this, we propose a hybrid deep learning architecture that integrates One-Dimensional Convolutional Neural Networks (1D-CNNs) with Gated Recurrent Units (GRUs) and Long Short-Term Memory Networks (LSTMs) to enhance feature extraction and emotion classification.

Materials and Methods: Experiments were conducted on the DEAP dataset, with a special emphasis on decreasing the number of EEG channels from the full set of 32 to 8 and 4, while maintaining robust performance. CNN layers were utilized to capture spatial dependencies, whereas GRU and LSTM layers modeled temporal dynamics. To optimize performance, the Adamax optimizer and Exponential Linear Unit (ELU) activation function were applied, ensuring stable convergence and efficient training.

Results: For an eight-channel EEG configuration, the proposed 1D-CNN-LSTM/GRU models achieved accuracies of 99.92% and 99.82%, respectively. When the input was reduced to 4 channels, the models maintained strong classification performance, achieving accuracies of 96.38% and 96.53%, respectively. These findings demonstrate that a substantial reduction in the number of EEG channels does not necessarily compromise recognition performance.

Conclusion: The proposed hybrid framework successfully balances accuracy and efficiency, showing that EEG-based emotion recognition can achieve near-perfect performance with fewer channels. This reduction lowers computational cost, simplifies system design, and enhances practicality for real-world applications. Future work will involve applying the framework to alternative datasets and exploring adaptive channel selection strategies to generalize its applicability further.

Keywords: Emotion Recognition; Electroencephalogram; Distributed Evolutionary Algorithms in Python; Deep Learning; Hybrid Networks; Convolutional Neural Networks.

DOI:

1. Introduction

Electroencephalogram (EEG) signals have emerged as a valuable modality for emotion recognition research due to their potential in applications such as online gaming, healthcare monitoring, and human-computer interaction [1]. Maintaining robust classification performance while limiting the number of EEG channels presents a significant challenge in EEG-based emotion recognition. Traditional systems rely on a large number of electrodes to capture detailed brain activity, but this setup limits portability and increases computational load. Therefore, developing efficient models that can recognize emotions with fewer EEG channels is both a practical and scientific necessity.

In emotion modeling, the valence–arousal (V–A) space is a widely used two-dimensional representation of affective states [2]. Valence denotes the degree of emotional pleasantness, ranging from unpleasant states at lower values to pleasant states at higher values. In contrast, arousal reflects the intensity of emotional activation, spanning from calm or relaxed conditions (low arousal) to highly excited or alert states (high arousal). Within this two-dimensional framework, emotional experiences are typically categorized into four quadrants, as described in Figure 1, HVHA (High Valence, High Arousal), representing emotions such as excitement and joy; HVLA (High Valence, Low Arousal), associated with states of relaxation and contentment; LVHA (Low Valence, High Arousal), encompassing emotions such as anger and tension; and LVLA (Low Valence, Low Arousal), corresponding to states such as boredom and depression. These four categories serve as the class labels in our classification model. A central difficulty in this research lies in accurately detecting and classifying emotional dimensions from physiological and biological signals acquired by body sensors.

In the physiological state, Electroencephalogram (EEG) signals, which are non-linear, non-stationary, and highly noisy, have been a significant challenge for emotion recognition in recent years [3- 5]; the amplitude of which ranged from 0.5 to 100 μV and was divided into four bands named delta (1-4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (12-25 Hz), and gamma (25-44 Hz). Moreover, the purpose is to define new feature maps from EEG signals to recognize and optimize emotional states, thereby achieving higher criteria, such as accuracy [4, 6].

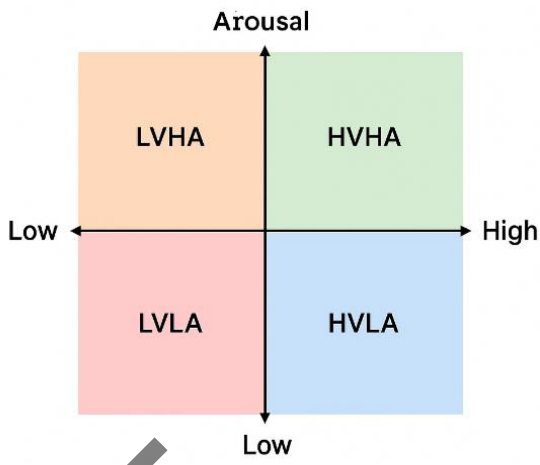

Figure 1. Valence-Arousal space

As a result, we decreased the number of channels based on physiological regions of the brain that relate to emotion. We navigate our experience on an available emotion dataset, DEAP. Furthermore, we proposed a combination of deep networks, a Convolutional Neural Network (CNN), and Recurrent Neural Networks (RNN). In detail, our CNN architecture is a One-Dimensional CNN (1D-CNN). The RNN structures are the Gated Recurrent Unit (GRU) and the Long Short-Term Memory (LSTM), which we have combined to produce hybrid networks called 1D-CNN-GRU/LSTM. Moreover, they can extract time-dependent features based on the structures of these networks.

In recent years, EEG-based emotion detection has attracted considerable research interest, leading to the development of diverse Machine Learning (ML) and Deep Learning (DL) approaches to tackle this complex task [4, 7]. These approaches can generally be classified into three categories: discrete emotion classification, two-dimensional (V–A) modeling, and hybrid feature–architecture frameworks.

Early models relied on the classification of basic emotions. For example, Ekman [8] introduced six fundamental emotions (anger, disgust, fear, happiness, sadness, and surprise), and Plutchik [3] extended this to eight categories. However, more recent studies have adopted the V-A model, which yields a continuous two-dimensional mapping of emotions and has become a standard approach in EEG-driven emotion recognition research.

Several researchers have employed deep learning architectures such as CNNs, RNNs, and hybrid models to extract both spatial and temporal features from EEG data. For instance, Xiang Li *et al.* [9] proposed a hybrid C-RNN model utilizing continuous wavelet transforms, achieving valence and arousal classification accuracies of 74.12% and 72.06%, respectively. Alhagry *et al.* [10] utilized LSTM networks on the DEAP dataset, achieving 85.65% accuracy for arousal and 85.45% accuracy for valence. Lin *et al.* [11] implemented an end-to-end CNN using AlexNet and reported accuracies of 82.30% and 85.50% for arousal and valence, respectively.

More advanced hybrid architectures have also emerged. Li *et al.* [12] presented CLRNN (CNN–LSTM–RNN) with 75.21% average accuracy per subject. Zhang *et al.* [13] introduced a multi-task shared hidden layer deep neural network (MT-SHL-DNN), reaching 78.34% on the ABS dataset. Cimtay [14] applied a dense CNN to DEAP and SEED, achieving up to 72.80% accuracy on DEAP. Acharya *et al.* [15] combined FFT-based feature extraction with CNN and LSTM models, reporting up to 88.6% accuracy. Similarly, Zhang *et al.* [16] used handcrafted statistical features and achieved 94.17% accuracy with CNN–LSTM. Aditi Sakalle [1] compared LSTM to traditional methods like MLP, KNN, and SVM on both DEAP and SEED datasets, achieving up to 94.12% accuracy with 10-fold cross-validation. Iyer *et al.* [17] developed a hybrid CNN–LSTM model using frequency-domain features (differential entropy) and achieved 97.16% on the SEED dataset. Uttam Singh *et al.* [18] integrated pre-processing, data augmentation, and metaheuristic channel selection, achieving 92.5% (valence) and 81.25% (arousal) accuracy. Jha *et al.* [19] applied a multi-column CNN architecture to brainwave and DEAP datasets and reported 81% accuracy on the latter. Fan *et al.* [20] used capsule networks and residual LSTM (ICaps–ResLSTM), obtaining over 97% accuracy on DEAP and DREAMER datasets. Asemi *et al.* [21] proposed a hybrid GWO-XGBoost method with feature selection, reaching 89.63% (arousal) and 89.08% (valence) using statistical features.

While these studies demonstrate significant progress, most rely on the full 32-channel EEG configuration, which increases hardware complexity and reduces portability. Moreover, few provide a detailed analysis of model size, computational cost, or inference efficiency. In contrast, our proposed methods are designed with lightweight architectures, a reduced number of trainable parameters, and optimized computation, all while achieving comparable or even superior performance.

Specifically, the input dimensionality is minimized without compromising classification accuracy by systematically decreasing the number of EEG channels to eight and four, guided by physiological and anatomical considerations. This reduction not only decreases training time and memory demands but also lowers energy consumption, thereby enhancing the feasibility of the proposed framework for real-time and wearable applications.

The key contribution of our work can be summarized as follows:

- We propose the EEG-based hybrid deep learning models (1D-CNN-GRU/LSTM) tailored for emotion recognition.
- We demonstrate that the proposed models can maintain high accuracy using only four EEG channels, a setting rarely explored in the literature.
- We design a lightweight 1D time-series architecture specifically tailored to reduce-channel EEG (4–8 electrodes). By aligning temporal receptive fields and recurrent memory depth with the minimal-channel constraint, the model preserves state-of-the-art accuracy (≥96% with 4 electrodes) while substantially reducing model size and on-device computational cost.
- Our approach achieves a practical trade-off between accuracy, simplicity, and real-time applicability, making it suitable for portable EEG systems.

Unlike previous CNN–RNN hybrids that assume dense multi-channel EEG input, our design explicitly co-optimizes (i) 1D temporal convolutional kernels, (ii) GRU/LSTM memory units, and (iii) normalization/activation schemes for an information-bottleneck regime with only 4–8 channels. This coupling is crucial when spatial redundancy is limited; it preserves discriminative temporal dynamics without relying on wide spatial sampling and thus makes the architecture suitable for wearable, reduced-electrode emotion-recognition systems.

The remainder of this paper is organized as follows. Section 2 provides an overview of the methods and approaches relevant to this study. Section 3 presents the experimental results and corresponding discussion. Finally, Section 4 concludes the paper and outlines directions for future research.

2. Materials and Methods

A brief explanation of the hybrid networks is provided below. Moreover, an overview of the proposed methodology is depicted in the block diagram shown in Figure 2.

2.1. Dataset

In this work, we used a well-known dataset of physiological signals called DEAP [22]. DEAP is a multimodal dataset that contains 32 EEG channels and other peripheral physiological signals, such as Electromyogram (EMG), blood volume pulse, Electrooculogram (EOG), skin temperature, plethysmograph, and Galvanic Skin Response (GSR), collected from 32 participants (16 male and 16 female) aged between 19 and 37 [22].

Participants were exposed to 40 music video clips, each lasting one minute. Following every trial, participants provided self-reports of their emotional states along the dimensions of arousal (passive–active), valence (negative–positive), liking (like–dislike), and dominance. In this study, we employed pre-processed EEG signals sampled at 128 Hz. EOG artifacts were removed from the recordings, and a band-pass filter in the frequency range of 4.0–45 Hz was subsequently applied.

The common reference for the EEG signals is their average. The dataset is categorized into four classes, as described in Section 1, based on the threshold values of valence and arousal. Table 1 summarizes the corresponding class labels. A threshold of 5 is applied for dimensional categorization, whereby values greater than 5 are considered high.

Table 1. The Classification Labels

Labels	Classified Labels			
	HVHA	**HVLA**	**LVHA**	**LVLA**
Valance	> 5	> 5	≤ 5	≤ 5
Arousal	> 5	≤ 5	> 5	≤ 5

2.2. Channel Selection

The overarching goal of this research is to improve the efficiency of deep neural networks in the DEAP database. For this purpose, the number of required EEG channels was reduced to four, while still achieving superior performance compared with approaches utilizing a larger number of channels, as described in Table 2. To reduce system complexity while maintaining high recognition accuracy, we selected a subset of EEG channels based on a combination of physiological, anatomical, and prior literature insights [23, 24]. Specifically, the prefrontal, frontal, and occipital lobes are known to be critical in emotional processing, as supported by numerous neuroscientific findings. Accordingly, our 14-channel configuration includes regions across the frontal (e.g., FP1, FP2, F3, F4), temporal (T7, T8), parietal (P7, P8), and occipital (O1, O2) lobes.

The 8-channel and 4-channel configurations were derived by progressively eliminating less emotion-relevant channels while preserving bilateral symmetry and anatomical coverage. In the 4-channel configuration, we selected two frontal channels (FP1,

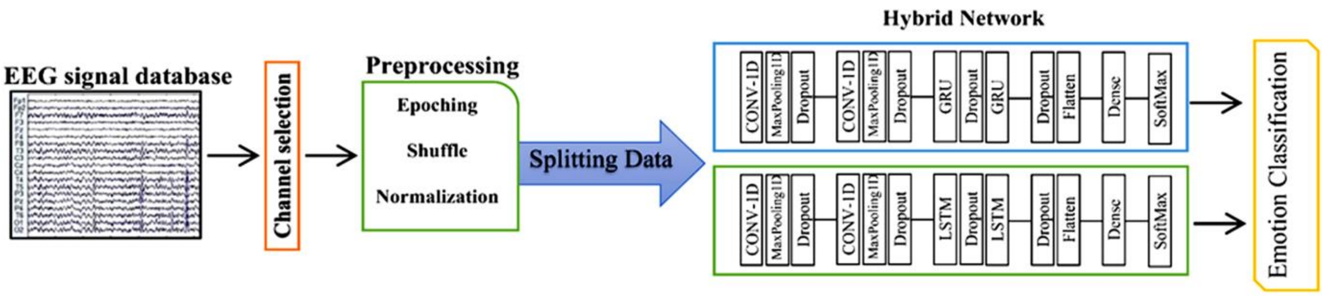

Figure 2. The overall block diagram of the proposed method

FP2) and two occipital channels (O1, O2), which are highly associated with emotional valence and arousal in EEG literature [23, 24]. While our current study focuses on these manually selected subsets, future work will involve evaluating data-driven channel selection techniques (e.g., mutual information, recursive feature elimination, metaheuristic algorithms) and comparing performance across different combinations. Such analysis can further validate the robustness of our selected configurations.

Table 2. Channel Selection

14 Channels			8 Channels			4 Channels		
Chans.	**Index**	**region**	**Chans.**	**Index**	**region**	**Chans.**	**Index**	**region**
AF3	2	F	FP1	1	F	FP1	1	F
AF4	18	F	FP2	17	F	FP2	17	F
F3	3	F	AF3	2	F	O1	14	O
F4	20	F	AF4	18	F	O2	32	O
F7	4	F	O1	14	O			
F8	21	F	O2	30	O			
FC5	5	F	PO3	13	P			
FC6	22	F	PO4	31	P			
T7	8	T						
T8	26	T						
P7	12	P						
P8	30	P						
O1	14	O						
O2	32	O						

F: Frontal T: Temporal
P: Parietal O: Occipital

PROOF

The selection of EEG channels was not only guided by previous neurophysiological findings but also empirically validated through systematic performance analysis. As demonstrated in Tables 4 and 6, classification accuracy exhibits a gradual and consistent decrease as the number of channels is reduced from 14 to 8 and further to 4 channels. This monotonic performance trend indicates that each reduction step corresponds to a measurable loss of spatial information, rather than an abrupt or erratic change that could suggest redundancy or information leakage. Moreover, the preservation of bilateral symmetry in channel selection contributes to balanced class-wise performance, as reflected in the confusion matrices, thereby supporting the stability and interpretability of the reduced-channel configurations.

2.3. Pre-Processing

Given that raw EEG signals are inherently prone to artifacts, ranging from eye blinks and muscle activity to environmental noise, pre-processing is an essential step in ensuring reliable emotion recognition. These non-neural components can significantly degrade the quality of extracted features and lead to poor classification performance. To ensure signal reliability and optimize model learning, we applied a comprehensive pre-processing pipeline to the DEAP dataset, including artifact handling, baseline correction, band-pass filtering, epoching, shuffling, and normalization, as detailed below.

A fourth-order Butterworth band-pass filter with a frequency range of 4–45 Hz was applied to remove Direct Connect (DC) drift and high-frequency noise. This range covers the primary EEG frequency bands (theta, alpha, beta, and low gamma) relevant to emotion processing. Although the DEAP dataset provides pre-processed signals, we applied additional checks for signal amplitude thresholds to exclude extreme fluctuations caused by movement artifacts. Future versions of this work may incorporate Independent Component Analysis (ICA) or Blind Source Separation (BSS) for more advanced artifact removal. Each trial was baseline-corrected by subtracting the mean signal during the 3-second pre-stimulus interval from the entire trial to eliminate initial resting-state bias. The continuous EEG signals were segmented into non-overlapping 2-second epochs, resulting in 31 epochs per 60-second trial. This duration was chosen to balance temporal resolution and signal stability. Each epoch contains 256 samples (given the 128 Hz sampling rate), forming the input window for the deep learning models. To avoid potential bias arising from the order of data collection, all epochs were randomly shuffled before training. This ensures that the model generalizes well and does not overfit to specific sequences or subjects.

EEG signal amplitudes vary across channels and subjects. We applied two normalization methods:

1. Min-Max Scaling

Rescales each signal to the range [0,1], enhancing uniformity across inputs, calculated as (Equation 1):

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where min is the minimum, max is the maximum, and X is the attribution value.

2. Z-score Normalization

Each channel was standardized by subtracting its mean and dividing by its standard deviation, thereby preserving the statistical structure of the data while reducing sensitivity to inter-subject variability, obtained as (Equation 2):

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

Where μ and σ are the mean and standard deviation of X , respectively.

2.4. Proposed Deep Neural Network

We propose two hybrid deep neural network architectures: 1D-CNN-GRU and 1D-CNN-LSTM. These models combine convolutional layers, which are well-suited for spatial pattern extraction, with recurrent layers, which are capable of modeling the temporal dynamics of EEG signals. This design leverages the complementary strengths of both architectures to capture complex spatiotemporal dependencies in EEG data.

2.4.1. One-Dimensional Convolutional Neural Network

One-Dimensional Convolutional Neural Network (1D-CNN) structure is described in Figure 3. The structure of 1D-CNNincludes convolutional layers with small kernel sizes to extract localized signal features, followed by max-pooling to reduce dimensionality and dropout layers to prevent

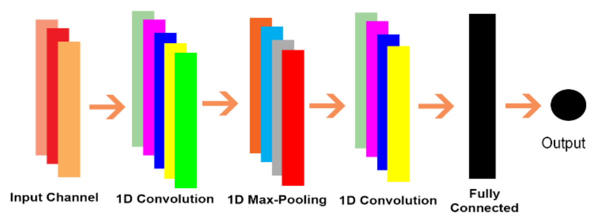

Figure 3. The structure of 1D-CNN

overfitting. The final dense layers map the extracted features to the emotion class space. This structure is particularly effective for 1D time-series signals such as EEG [3].

2.4.2. Recurrent Neural Network

To address the limitations of simple Recurrent Neural Networks (RNNs), particularly their inability to retain long-term dependencies due to vanishing gradient problems, we employed two advanced recurrent units: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). These architectures are designed to capture the temporal structure of EEG signals more effectively [4, 7].

- By combining the forget and input gates into a unified update gate, the GRU reduces model complexity and training time without compromising the accuracy achieved by LSTM networks.
- LSTM incorporates input, forget, and output gates to regulate the flow of information and preserve longer-term temporal dependencies. This makes it particularly suitable for modeling delayed emotional responses in EEG data.

Both GRU and LSTM architectures were implemented and evaluated to determine their effectiveness in emotion recognition using reduced-channel EEG inputs. The detailed internal structures of these recurrent units are illustrated in Figure 4 [25].

2.4.2.1. Long Short-Term Memory

Figure 4a illustrates the basic Long Short-Term Memory (LSTM) structure. It has three gates: input, forget, and output gates [1, 4, 25]. According to these gates, LSTM decides which information is essential to keep and which is unnecessary. Memory units used in the LSTM cell allow it to carry relevant information and remember the previous steps. Cell state (c_t) is the key part of an LSTM. The first gate, known as the forget gate, determines which information from the cell state should be retained and which should be discarded, a decision governed by a sigmoid activation function. The second gate, i.e. the input gate, regulates which new information will update the cell state, again using a sigmoid function to control this process. Subsequently, a tanh layer generates a vector of candidate values in the range of –1 to 1, representing potential updates to the cell state, as in Equation 3.

$$tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{3}$$

2.4.2.2. Gated Recurrent Unit

Figure 4b shows the basic Gated Recurrent Unit (GRU) architecture. It is like an LSTM with a forget gate [1, 4, 25]. However, it has fewer parameters than LSTM. The point of this network is to be faster than LSTM in computing the parameters. Also, this network contains the rest and update gates. The GRU integrates the current input with the hidden state from the previous time step, which carries the relevant historical information. It then produces the output for the current time step and updates the hidden state, which is subsequently propagated to the next time step [4, 26]. As shown in Figure 4a, there is a tanh, the same as LSTM, in the hidden state of the network.

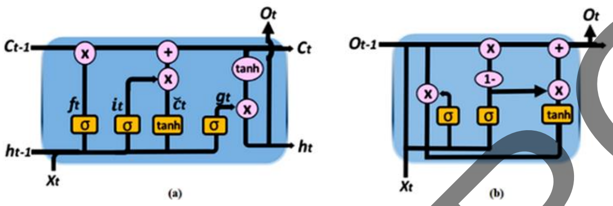

Figure 4. RNN structure (a) LSTM (b) GRU [25]

2.5. Data Shape

All pre-processed data were split into train, validation, and test subsets with a ratio of 60%:20%:20%, respectively. To prevent overfitting and ensure robust generalization, we carefully designed the data-splitting process to avoid any potential data leakage. Specifically, we employed a subject-independent evaluation strategy, where the data from each subject is strictly used in either the training or testing set, but never both. This approach ensures that the model does not rely on subject-specific patterns and instead learns generalizable emotional features across individuals.

Although the DEAP dataset contains 32-channel EEG recordings for 32 subjects, we extracted non-overlapping 2-second epochs per trial (31 per 60-second video), and these were grouped by subject before shuffling and partitioning. No overlapping epochs from the same subject, trial, or emotional condition were allowed to be distributed across training and test sets. These steps were taken to ensure that the exceptionally high accuracy is due to the true model performance and not a side effect of data leakage or memorization.

2.5.1. Validation Integrity and Leakage Control

To ensure the validity and reliability of the reported results, a strict subject-independent validation strategy was employed, with particular attention to preventing data leakage. Given the exceptionally high classification accuracies observed in reduced-channel EEG settings, rigorous validation integrity is essential to confirm that the results reflect genuine model generalization rather than memorization or statistical bias.

First, all pre-processing operations, including filtering, baseline correction, segmentation, and normalization, were performed independently within each subject before any data partitioning. Subjects were then assigned exclusively to training, validation, or test sets at the subject level, ensuring that no individual contributed samples to more than one subset. Epoch segmentation was applied only after subject-level partitioning, which prevents overlapping temporal windows from the same subject or trial from being distributed across different splits.

Second, normalization statistics were computed solely from the training subjects and subsequently applied to the validation and test sets. This design guarantees that no statistical information from unseen subjects influenced model training. All shuffling procedures were restricted to the training set, and no trial-level or epoch-level samples were shared across subsets.

Beyond protocol design, the integrity of the validation strategy is further supported by the model's empirical behavior across different channel configurations. As reported in Tables 4 and 6, classification accuracy consistently decreases as the number of EEG channels is reduced from 14 to 8 and finally to 4 channels. This performance reduction is consistent with the decrease in available spatial information when fewer EEG channels are used. Instead, the observed performance trend indicates that the models rely on progressively reduced physiological information, thereby providing indirect empirical evidence that the reported accuracies are not inflated by leakage artifacts. Collectively, the combination of subject-independent partitioning and leakage-aware pre-processing supports the validity of the subject-independent evaluation procedure and indicates that the reported results reflect genuine generalization across subjects rather than data leakage or overfitting.

2.6. Hyperparameter

To achieve the best results, hyperparameters must be tuned. The batch size used in our model is 256. Moreover, the loss function, named categorical cross-entropy, was employed to update the weights during backpropagation, which is suitable for multi-labeled data. Our research work uses the optimizers Adam, Adamax, and AdamW. The Adamax, Nadam, and AdamW are variants of the Adam optimization. The Adamax optimizer is a first-order gradient-based optimization method. Based on its capability of learning the rate of data characteristics setting, it is one of the suitable optimizers in time-variant processes. However, the Nadam optimizer is the same as the Adam optimization with Nesterov momentum. AdamW is a stochastic method that modifies the implementation of weight decay in the Adam optimizer. In addition, the callback method, with a patience value of 20, monitors the validation loss value. 200 is our default epoch parameter.

2.7. 1D-CNN-GRU

A combination of 1D-CNN and GRU is the first proposed model. Table 3 presents the structural details of this network. The input size of the network is 256x1. The first layer is convolutional with a filter size of 125 and the activation function of rectified linear units (ReLU), calculated as Equation 4. Following the first convolutional layer, a max-pooling layer with a pool size of 2 was applied. To mitigate overfitting, a dropout rate of 0.2 was introduced after the second convolutional layer, which was identical to the first. These layers were followed by two GRU layers with 256 and 32 units, respectively, each accompanied by a dropout layer of 0.2. The output was then flattened to form a one-dimensional feature vector before being passed to a dense layer with 128 units and a ReLU activation function. Finally, a dense output layer with four units, corresponding to the number of classes, was employed with a SoftMax activation function for classification, calculated by Equation 5.

$$F(x) = max\ (0, x) \tag{4}$$

$$\text{SoftMax}(x_i) = \frac{\exp\ (x_i)}{\sum \exp\ (x_j)} \tag{5}$$

2.8. 1D-CNN-LSTM

This model is almost identical to the first proposed network, 1D-CNN-GRU. The only difference is replacing the LSTM layer with the GRU. We summarize the model's overview in the Table 3.

2.9. Performance Measure

We have evaluated the performance of both models by performing an analytical analysis, including a confusion matrix, accuracy, F1-score, precision, and recall, as described in the following.

A confusion matrix is a performance measurement that summarizes prediction classes versus true classes in a classification problem. In classification tasks, accuracy measures the proportion of inputs that are correctly categorized into their true classes (Equation 6):

$$Accuracy = \frac{\text{TP} + \text{TN}}{TP + TN + FP + FN} \times 100 \tag{6}$$

Where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

An important parameter for the analysis of an imbalanced dataset is the F1-score. The measure is defined as the harmonic mean of precision and recall, with weights reflecting their relative importance. They can calculate as follows (Equations 7-9):

$$recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{\text{TP}}{TP+FP} \tag{8}$$

$$F1 - Score = \frac{precision \times recall}{precision + recall} \tag{9}$$

Table 3. The structure details of the Hybrid Networks

The Hybrid Proposed Networks			
1D-CNN-GRU	**Parameter**	**1D-CNN-LSTM**	**Parameter**
Input layer	(256, 1)	Input layer	(256, 1)
CONV1D + ReLU	F=128, k=3	CONV1D + ReLU	F=128, k=3
Maxpooling 1D	Pool_size=2	Maxpooling 1D	Pool_size=2
Dropout	Rate=0.2	Dropout	Rate=0.2
CONV1D + ReLU	F=128, k=3	CONV1D + ReLU	F=128, k=3
Maxpooling 1D	Pool_size=2	Maxpooling 1D	Pool_size=2
Dropout	Rate=0.2	Dropout	Rate=0.2
GRU	Units=256, r =True	LSTM	Units=256, r =True
Dropout	Rate=0.2	Dropout	Rate=0.2
GRU	Units=32	LSTM	Units=32
Dropout	Rate=0.2	Dropout	Rate=0.2
Flatten	None	Flatten	None
Dense + ReLU	Units= 128	Dense + ReLU	Units= 128
Dense + SoftMax	Units= 4	Dense + SoftMax	Units= 4
Total params	378,820	Total params	485,764

Table 4. The accuracy results of the channels in both methods

Hybrid Networks with Adam+ReLU								
No. of Chans.	Values							
	Methods	**Training Acc.**	**Training loss**	**Val. Acc.**	**Val. loss**	**Test Acc.**	**Test loss**	**Epoch**
14	1D-CNN-GRU	0.9718	0.0858	0.9973	0.0113	0.9975	0.0107	128
	1D-CNN-LSTM	**0.9787**	**0.0645**	**0.9999**	**0.0011**	**0.9999**	**0.0017**	**200**
8	1D-CNN-GRU	0.9589	0.1204	0.9982	0.0113	0.9982	0.0102	152
	1D-CNN-LSTM	**0.9706**	**0.0871**	**0.9993**	**0.0058**	**0.9992**	**0.0048**	**200**
4	1D-CNN-GRU	0.9345	0.1897	0.9646	0.1705	0.9620	0.1808	200
	1D-CNN-LSTM	**0.9467**	**0.1580**	**0.9644**	**0.1831**	**0.9608**	**0.1820**	**142**

* Acc: Accuracy Val: Validation

3. Results and Discussion

This section presents the experimental results of the hybrid networks for classifying emotional states. All experiments were implemented using TensorFlow version 2.10.1 and Python version 3.9.12, running on an Intel(R) Core (TM) i7-7500U CPU @ 2.70 GHz, Nvidia (R) GeForce (R) 940MX with 2 GB VRAM for GPU acceleration, 500 GB SSD, and 8 GB RAM.

All data was pre-processed and normalized to enhance accuracy, and the experiments were carried out in different configurations, as explained below. First, 14 EEG channels were utilized to evaluate the network's performance. The initial results for 14 channels showed how the models handled a larger volume of EEG data, providing a baseline for comparison. Then, based on previous studies and input shapes of the networks, we proceeded by diminishing the number of channels to 8 and 4 as described in Section 2. This approach enabled us to assess the model's ability to maintain high accuracy with fewer inputs, which is crucial for developing practical EEG-based applications.

Table 4 presents the results for 14, 8, and 4 channels. The 1D-CNN-LSTM model performed best, achieving 99.99%, 99.92%, and 96.08% accuracy for 14, 8, and 4 channels, respectively. These results suggest that even with reduced channels, the model retains robust classification capability, which is highly significant for minimizing the complexity of future systems. Figure 5 presents the accuracy results for the training, validation, and test phases, demonstrating that the model generalizes effectively across different datasets and thereby reinforcing the robustness of the proposed hybrid architectures. Table 5 summarizes the fine-tuning parameters used in the models, highlighting the importance of tuning the learning rate, batch size, and activation functions to optimize model performance. The Exponential Linear Unit (ELU) activation function helped to reduce the cost faster by minimizing vanishing gradient issues. In combination with the Adamax optimizer, it yielded superior results compared to the ReLU activation function with Adam. This result aligns with findings from other deep learning studies that favour ELU in complex models. Table 6 provides performance metrics for the selected channels, showing how accuracy, precision, recall, and F1-score vary across the different channel

Table 5. The results of various hyperparameter fine-tuning with 4 channels of EEG signals

	methods	Params.	Training Acc.	Training loss	Val. Acc.	Val. loss	Test Acc.	Test loss	Epoch
4 Chan	1D-CNN-GRU	Adam+	0.9345	0.1897	0.9646	0.1705	0.9620	0.1808	200
	1D-CNN-LSTM	ReLU	0.9467	0.1580	0.9644	0.1831	0.9608	0.1820	142
	1D-CNN-GRU	Adam+	0.9637	0.1082	0.9610	0.2058	0.9596	0.2014	145
	1D-CNN-LSTM	ELU	0.9625	0.1120	0.9646	0.1798	0.9631	0.1896	153
	1D-CNN-GRU	**Adamax+**	**0.9764**	**0.0710**	**0.9650**	**0.2200**	**0.9653**	**0.2051**	**161**
	1D-CNN-LSTM	**ELU**	**0.9733**	**0.0813**	**0.9661**	**0.2030**	**0.9638**	**0.2110**	**169**
	1D-CNN-GRU	Nadam +	0.9594	0.1193	0.9557	0.2419	0.9545	0.2331	103
	1D-CNN-LSTM	ELU	0.9600	0.2080	0.9576	0.1264	0.9622	0.2025	169
	1D-CNN-GRU	AdamW+	0.3432	1.3648	0.3435	1.3635	0.3419	1.3641	21
	1D-CNN-LSTM	ELU	0.3432	1.3649	0.3435	1.3641	0.3419	1.3648	26

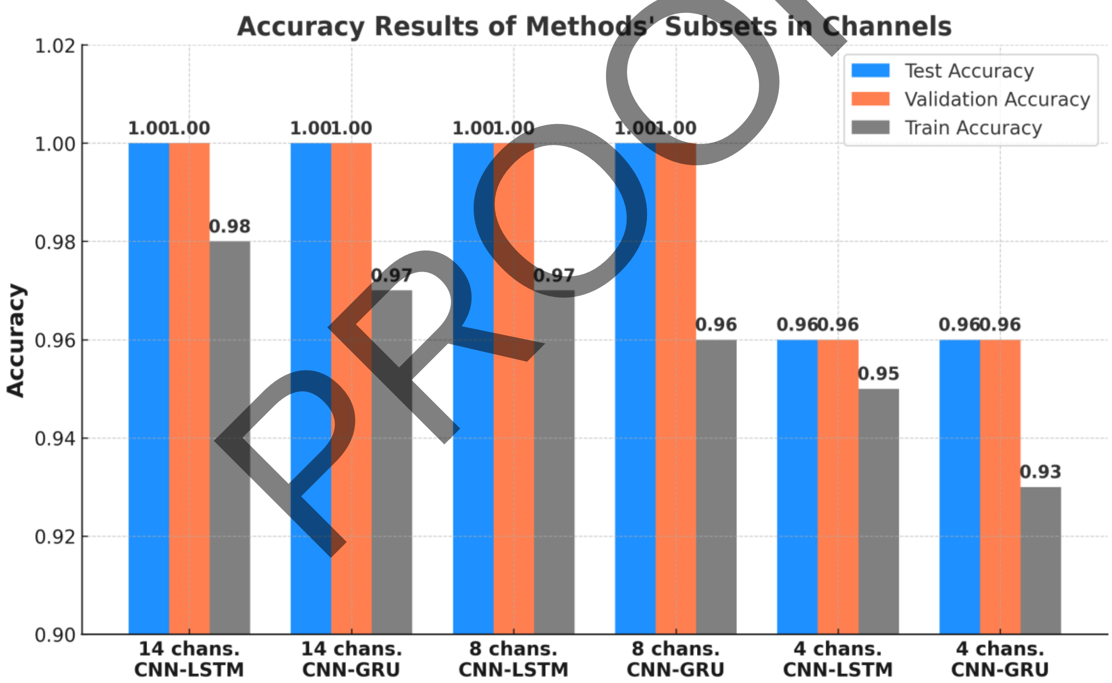

Figure 5. The accuracy results of the methods' subsets

Table 6. The performance measures the results of channels

methods	channels	precision	recall	f1-score	accuracy
1D-CNN-LSTM	14	1.0000	1.0000	1.0000	1.0000
1D-CNN-GRU		1.0000	1.0000	1.0000	1.0000
1D-CNN-LSTM	8	1.0000	1.0000	1.0000	1.0000
1D-CNN-GRU		1.0000	1.0000	1.0000	1.0000
1D-CNN-LSTM	4	0.9700	0.9700	0.9700	0.9700
1D-CNN-GRU		0.9600	0.9600	0.9600	0.9600

configurations. To assess the robustness and stability of the proposed models, all experiments were repeated five times using different random initialization seeds while keeping the data splits and evaluation protocol fixed. This procedure allows the evaluation of optimization stability and sensitivity to stochastic training effects, which is particularly important given the high classification accuracies reported. Across all channel configurations, the proposed 1D-CNN–GRU and 1D-CNN–LSTM models exhibited highly consistent performance with minimal run-to-run variability. The standard deviation of classification accuracy ranged between 0.01 and 0.05 in both the reduced-channel and higher-channel configurations, indicating stable convergence behavior and robustness against random initialization. In the 4-channel configuration, the GRU-based model achieved an accuracy of 99.91% ± 0.04, while the LSTM-based model reached 99.99% ± 0.01 across the five runs. Similarly, in the 8-channel configuration, accuracies remained stable (GRU: 99.87% ± 0.05, LSTM: 99.95% ± 0.02). The consistently low variance across repeated runs suggests that the reported performance is not due to favorable random initialization or stochastic fluctuations during training. Instead, it reflects a stable optimization landscape and well-conditioned model architecture. Moreover, the smooth and repeatable convergence trajectories observed during training further support the reliability of the reported results. Because the repeated runs share identical splits and are not statistically independent samples, formal significance tests (t-test or Wilcoxon) are not applicable; therefore, performance comparison is based on effect magnitude rather than inferential statistics. The GRU-based model offers computational efficiency at the cost of slightly lower performance, while the LSTM-based model provides marginally higher accuracy at the cost of increased complexity. Overall, the results validate that our models generalize well and are not affected by random initialization or stochastic optimization variability. In addition to classification accuracy, model complexity and computational efficiency represent critical factors for real-time and resource-constrained applications. As shown in Table 7, the proposed architectures [1, 9, 10, 14, 16, 20] achieve substantially lower parameter counts than many state-of-the-art counterparts, thereby enhancing their suitability for practical deployment. Specifically, the GRU-based model comprises approximately 49,000 trainable parameters, benefiting from its simplified gating mechanism, while the LSTM-based variant contains around 61,000 parameters, providing slightly higher expressive capacity. These lightweight configurations contrast sharply with prior studies, where models often exceeded 500,000 to over 1 million parameters, and yet required the full 32-channel EEG setup. This highlights the efficiency and practicality of our approach, particularly in mobile and wearable EEG-based emotion recognition systems.

Figures 6 and 7 illustrate the experimental curves and confusion matrices for 1D-CNN-GRU and 1D-CNN-LSTM, respectively, for 14, 8, and 4 channels. These figures show the training and validation performance, with curves peaking at 100 epochs. The confusion matrices indicate that the misclassifications decrease significantly as the training progresses, demonstrating the models' ability to learn intricate patterns within EEG data. This suggests that the models, particularly the 1D-CNN-LSTM, can effectively distinguish between subtle emotional states based on the EEG inputs. A deeper comparison between GRU and LSTM reveals consistent advantages for GRU under reduced-channel EEG conditions. As illustrated in Figures 6 and 7, GRU converges in fewer epochs and produces smoother validation trajectories, indicating more stable gradient propagation. This behavior aligns with the simplified gating structure of GRU, which reduces temporal-memory overhead and improves optimization efficiency when spatial redundancy is limited to 4–8 electrodes. In addition, GRU shows shorter mean epoch time during training. These observations demonstrate that GRU's superior performance arises from its convergence dynamics and gating efficiency rather than from descriptive accuracy differences alone.

To further interpret the model performance, confusion matrices for the 4-channel-based models were analyzed. In both cases, a highest classification accuracy was observed for the HVHA and LVLA quadrants, which represent more distinct emotional extremes (e.g., excitement vs. depression). These states are often associated with clearer neural signatures, thereby improving separability. In contrast, the HVLA and LVHA categories showed slightly higher confusion rates. This observation is

Table 7. Comparative Model Complexity with Prior Studies

Author / Study	Model Type	Trainable Parameters	Notes
Our proposed method	1D-CNN- GRU	~49,000	Lightweight, fast, suitable for wearables
	1D-CNN - LSTM	~61,000	Slightly higher accuracy, more expressive
Fan *et al.*	ICaps–ResLSTM	>1,000,000 (est.)	Capsule + Residual LSTM, high complexity
Zhang *et al.*	CNN–LSTM	>500,000 (est.)	Uses statistical features
Alhagry *et al.*	LSTM	~130,000 (est.)	Simple LSTM model on DEAP
Xiang Li *et al.*	C-RNN	>200,000 (est.)	CWT + frame-level processing
Sakalle *et al.*	LSTM	~150,000 (est.)	Single-layer LSTM
Cimtay *et al.*	Dense CNN	>1,000,000 (est.)	CNN depth and width not optimized

consistent with prior literature, which has noted the subtle neural overlap between relaxed and tense states or between frustration and calmness, especially when using fewer EEG channels. For example, a small portion of HVLA samples was misclassified as LVLA, suggesting that the model may sometimes confuse low-arousal emotional states despite differing valence.

Similarly, some LVHA instances were incorrectly labeled as HVHA, reflecting potential difficulty in distinguishing between emotional intensity and positivity when arousal levels are high. These patterns emphasize the challenge of emotion classification in nuanced emotional zones and point toward future work in incorporating context-aware or subject-personalized modeling to improve accuracy in borderline cases.

Moreover, the results indicate that the hybrid models, 1D-CNN-GRU and 1D-CNN-LSTM, achieved high classification accuracies for EEG signals in four emotion categories. Notably, 1D-CNN-LSTM reached 97% accuracy with only 4 channels, outperforming the other models. These findings confirm the advantage of hybrid CNN–GRU/LSTM architectures, which jointly capture spatial and temporal characteristics of EEG signals, enabling robust emotion recognition even with fewer channels. The ability to maintain accuracy with fewer channels is particularly advantageous for real-time and portable EEG applications, where reducing the number of electrodes enhances user comfort and system efficiency. The ELU activation function, with its alpha parameter, enabled faster convergence compared to other activation functions like ReLU and Leaky ReLU, particularly in deeper networks where vanishing gradients could become problematic. The Adamax optimizer, paired with ELU, achieved the best performance with the fewest epochs, underscoring the importance of selecting the right optimizer to balance speed and accuracy. The choice of the Adamax optimizer combined with the ELU activation function was further supported by the observed training dynamics. Across all experimental settings, this configuration consistently resulted in faster convergence and smoother loss trajectories compared to alternative combinations evaluated during preliminary experiments. In particular, stable convergence was achieved within fewer epochs and with reduced sensitivity to random initialization, as evidenced by the low variance reported in the robustness analysis. These empirical observations suggest that the selected optimization strategy contributes not only to high classification accuracy but also to improved training stability, which is critical for a reliable performance evaluation in reduced-channel EEG settings. These results also suggest that hybrid architectures combining 1D-CNN with RNNs like GRU and LSTM can benefit from fine-tuning the activation functions to improve learning in complex datasets such as EEG signals. Reducing the number of EEG channels also improved training speed, minimizing memory and computational requirements. This reduction in complexity without sacrificing accuracy demonstrates the practical applicability of the models, making them suitable for use in portable or wearable devices where computational resources are limited. Moreover, by not relying on feature extraction techniques like Fast Fourier Transform (FFT) or Continuous Wavelet Transform (CWT), our approach simplifies the pre-processing pipeline, further enhancing the feasibility of real-time applications, as illustrated in Table 8 [1, 18-20, 27, 28]. These results are consistent with previous studies, but our models achieved higher accuracies with fewer channels, which has potential applications in reducing the burden of EEG equipment on users and improving portability.

Furthermore, the ability to retain accuracy across different channel configurations highlights the flexibility of the models for various applications, such as in neurofeedback systems, Brain-Computer Interfaces (BCIs), or Emotion Recognition Systems (ERSs) used in healthcare or entertainment. Despite the promising results, the near-100% should be

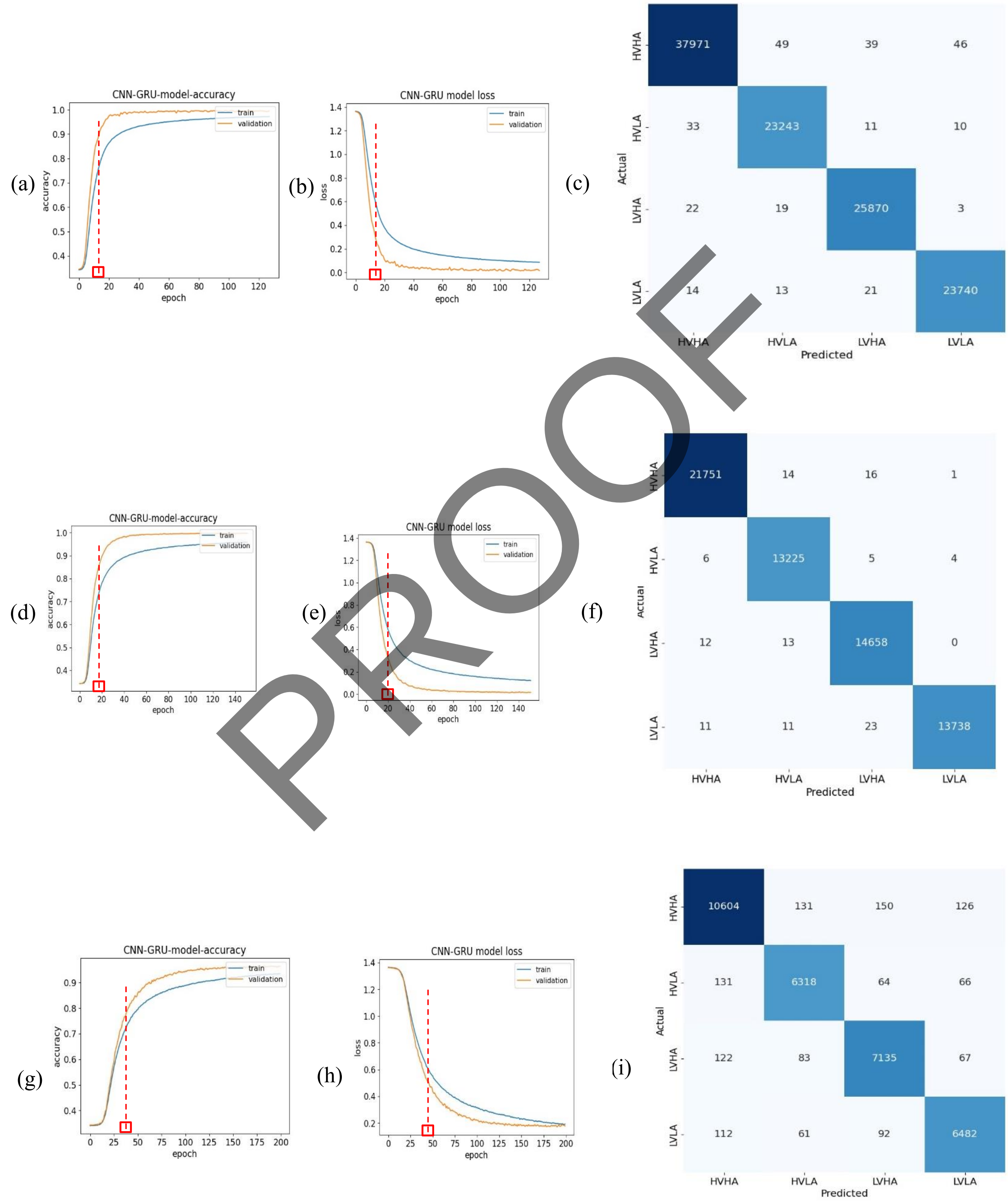

Figure 6. The experimental channels' curve graphs and confusion matrixes of 1D-CNN-GRU (a) The acc. graph for 14 channels (b) The loss graph of for 14 channels (c) The confusion matrix for 14 channels (d) The acc. graph for 8 channels (e) The loss for 8 channels (f) The confusion matrix of for 8 channels (g) The acc. graph for 4 channels (h) The loss graph for 4 channels (i) The confusion matrix for 4 channels

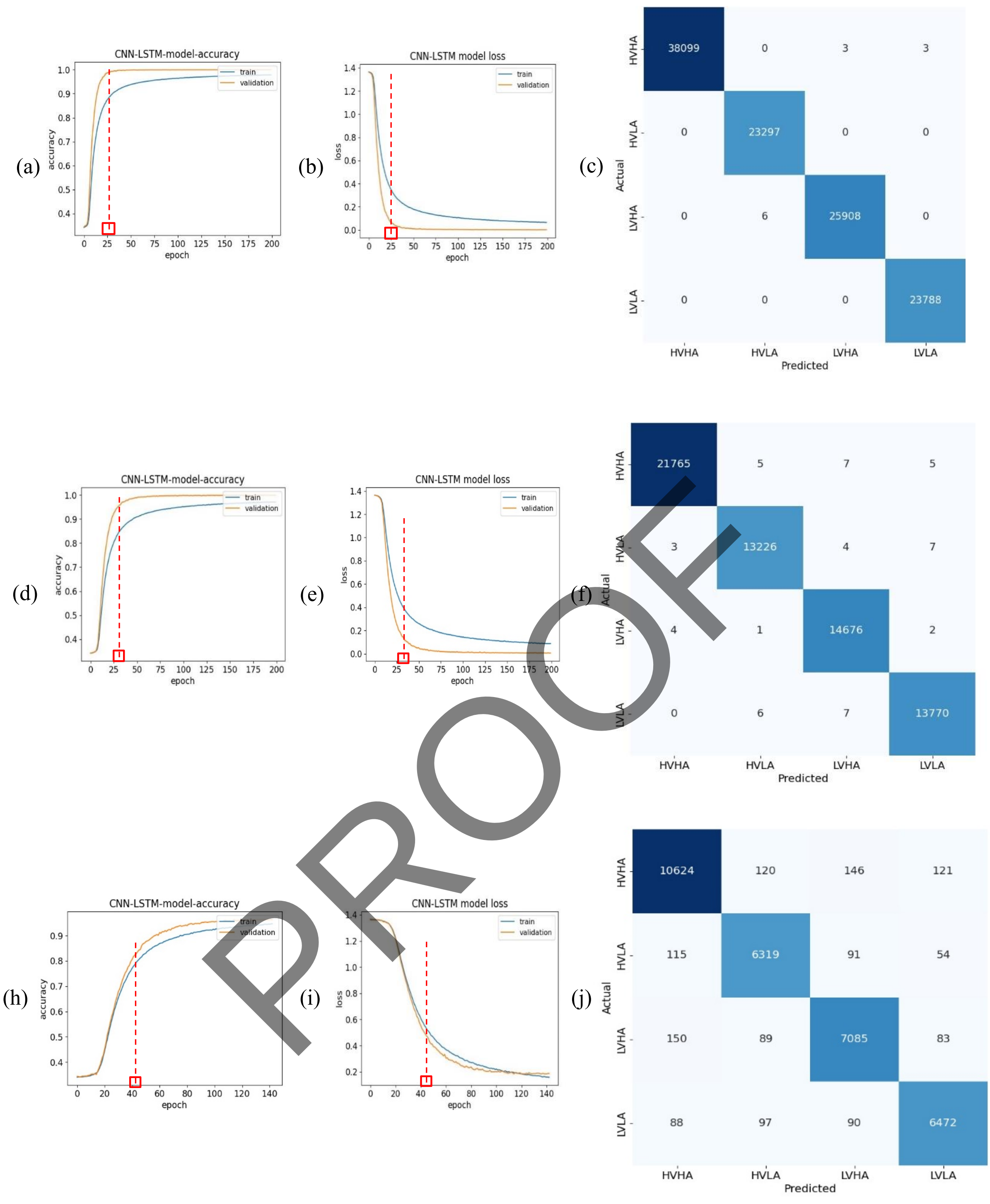

Figure 7. The experimental channels' curve graphs and confusion matrixes of 1D-CNN-LSTM (a) The acc. graph for 14 channels (b) The loss graph for 14 channels (c) The confusion matrix for 14 channels (d) The acc. graph for 8 channels (e) The loss graph for 8 channels (f) The confusion matrix for 8 channels (h) The acc. graph for 4 channels (i) The loss graph for 4 channels (j) The confusion matrix for 4 channels

interpreted as evidence of the model's global generalizability. While the proposed hybrid models achieve very high classification accuracies, particularly in the reduced-channel configurations, these results should be interpreted with appropriate caution. High accuracy values in EEG-based emotion recognition can be influenced by several factors, including inter-subject consistency within a specific dataset, controlled experimental conditions, and the intrinsic characteristics of the emotion elicitation

Table 8. The accuracy value of the comparison methods

Author	Methodology	Accuracy (%)	No. of Chans.
Nakisa *et al.*	Probabilistic Neural Network	67.47	32
Liu *et al.*	Bimodal Deep Autoencoder	83.25	32
Aditi Sakalle *et al.*	LSTM	80.66	32
	LSTM	91.38	32
Uttam Singh *et al.*	LSTM	V:92.5 A:81.25	32
Sonu Kumar Jha *et al.*	multiple-column CNN	81	32
Cunhang Fan *et al.*	ICaps-ResLSTM	A:98.06 V:97.94 D:98.15	32
Our proposed method	**1D-CNN-GRU**	**99.82**	**8**
	1D-CNN-LSTM	**99.92**	
Our proposed method	**1D-CNN-GRU**	**96.20**	**4**
	1D-CNN-LSTM	**96.08**	

protocol. To mitigate these concerns, the present study employed a strict subject-independent evaluation strategy with leakage-aware pre-processing, as detailed in Section 2.5.1. In addition, robustness analysis across multiple random initializations demonstrated stable convergence behavior and minimal performance variability, suggesting that the reported accuracies are reproducible rather than incidental.

Nevertheless, the scope of the present findings is inherently constrained by the use of a single dataset (DEAP) and a fixed experimental protocol. Therefore, the reported results should be viewed as evidence of the effectiveness of channel-efficient hybrid architectures within this experimental framework, rather than as definitive proof of universal generalizability. Further validation on independent datasets, alternative evaluation protocols such as leave-one-subject-out or cross-dataset transfer, and real-world recordings are required to assess the generalization capability of the proposed approach fully.

4. Conclusion

In this study, we introduced and evaluated two lightweight yet highly effective hybrid deep architectures, 1D-CNN–GRU/LSTM, for EEG-based emotion recognition using the benchmark DEAP dataset. By combining the spatial feature extraction capabilities of 1D-CNN with the temporal modeling strengths of GRU and LSTM, the proposed models successfully captured the complex spatiotemporal dynamics inherent in emotional EEG signals. The main goal of this study is to systematically investigate three channel configurations: 14, 8, and as few as 4 EEG channels. Remarkably, both models maintained exceptional performance even in the reduced-channel setup, achieving 96% and 97% accuracy with 1D-CNN–GRU/LSTM, respectively, in a four-class emotion classification framework (HVHA, HVLA, LVHA, LVLA). This highlights the potential of the proposed models for resource-constrained and real-time applications, such as wearable affective computing and mobile neurofeedback systems. Our comprehensive hyperparameter optimization revealed that the Adamax optimizer, in conjunction with the ELU activation function, yielded the most robust and accurate learning behavior. Moreover, comparative evaluations consistently favored the CNN–LSTM model, which outperformed its GRU-based counterpart across most scenarios, underscoring its superior ability to capture long-range temporal dependencies. The results of this study demonstrate that channel-efficient deep neural architectures can offer state-of-the-art performance without the overhead of full-scale EEG setups. The results indicate that lightweight and channel-reduced designs can provide high performance in the present experimental framework; however, these findings are limited by the scope of our experiment and should not be generalized definitively. The limitations of this study, including the sensitivity of the channel-reduced models to subject patterns and being tested only on the DEAP dataset with the subject-independent protocol used, should be considered when interpreting the results. Validation of generalizability to other datasets and alternative evaluation methods, such as Leave one Subject Out (LOSO) or cross-dataset, is proposed in the literature as complementary studies; our results can be considered as a reason for further research in this direction, not as a definitive proof of the generalizability of the method. These directions aim to enhance further the generalizability, interpretability, and real-world applicability of emotion recognition systems in affective neuroscience and BCI domains.

References

1- Aditi Sakalle, Pradeep Tomar, Harshit Bhardwaj, Divya Acharya, and Arpit Bhardwaj, "A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system." *Expert Systems with Applications,* Vol. 173p. 114516, (2021).

2- James A Russell, "A circumplex model of affect." *Journal of personality and social psychology,* Vol. 39 (No. 6), p. 1161, (1980).

3- Ante Topic and Mladen Russo, "Emotion recognition based on EEG feature maps through deep learning network." *Engineering Science and Technology, an International Journal,* Vol. 24 (No. 6), pp. 1442–54, (2021).

4- Mahdi -Jafari Asl, Sina Shamekhi, and Fatemeh Shalchizadeh, "Region-Specific Eeg Channel-Based Emotion Detection Using Bi-Directional Deep Neural Networks." in *2025 32nd National and 10th International Iranian Conference on Biomedical Engineering (ICBME)*, (2025): *IEEE*, pp. 323–28.

5- Nagarajan Ganapathy, Yedukondala Rao Veeranki, and Ramakrishnan Swaminathan, "Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features." *Expert Systems with Applications,* Vol. 159p. 113571, (2020).

6- Fatemeh Shalchizadeh, Sina Shamekhi, and Mahdi Jafari Asl, "Binary Discrete Emotion Detection with Peripheral and Fp1-Fp2 EEG Signals on PEEFS Dataset." in *2025 32nd National and 10th International Iranian Conference on Biomedical Engineering (ICBME)*, (2025): *IEEE*, pp. 258–62.

7- Mahdi Jafari Asl and Saba Haji Molla Rabie, "Deep Learning Approaches for Alzheimer's Disease Diagnosis: a Comprehensive Review." in *2025 32nd National and 10th International Iranian Conference on Biomedical Engineering (ICBME)*, (2025): *IEEE*, pp. 1049–54.

8- Paul Ekman, "An argument for basic emotions." *Cognition & emotion,* Vol. 6 (No. 3-4), pp. 169–200, (1992).

9- Xiang Li, Dawei Song, Peng Zhang, Guangliang Yu, Yuexian Hou, and Bin Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network." in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*, (2016): *IEEE*, pp. 352–59.

10- Salma Alhagry, Aly Aly Fahmy, and Reda A El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network." *International Journal of Advanced Computer Science and Applications,* Vol. 8 (No. 10), (2017).

11- Wenqian Lin, Chao Li, and Shouqian Sun, "Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal." in *International conference on image and graphics*, (2017): *Springer*, pp. 385–94.

12- Youjun Li, Jiajin Huang, Haiyan Zhou, and Ning Zhong, "Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks." *Applied Sciences,* Vol. 7 (No. 10), p. 1060, (2017).

13- Yue Zhang, Yifan Liu, Felix Weninger, and Björn Schuller, "Multi-task deep neural network with shared hidden layers: Breaking down the wall between emotion representations." in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, (2017): *IEEE*, pp. 4990–94.

14- Yucel Cimtay and Erhan Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition." *Sensors,* Vol. 20 (No. 7), p. 2034, (2020).

15- Divya Acharya *et al.*, "Multi-class emotion classification using EEG signals." in *International Advanced Computing Conference*, (2020): *Springer*, pp. 474–91.

16- Yaqing Zhang *et al.*, "An investigation of deep learning models for EEG-based emotion recognition." *Frontiers in Neuroscience,* Vol. 14p. 622759, (2020).

17- Abhishek Iyer, Srimit Sritik Das, Reva Teotia, Shishir Maheshwari, and Rishi Raj Sharma, "CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings." *Multimedia Tools and Applications,* Vol. 82 (No. 4), pp. 4883–96, (2023).

18- Uttam Singh, Rabi Shaw, and Bidyut Kr Patra, "A data augmentation and channel selection technique for grading human emotions on DEAP dataset." *Biomedical Signal Processing and Control,* Vol. 79p. 104060, (2023).

19- Sonu Kumar Jha, Somaraju Suvvari, and Mukesh Kumar, "Emotion recognition from electroencephalogram (EEG) signals using a multiple column convolutional neural network model." *SN Computer Science,* Vol. 5 (No. 2), p. 213, (2024).

20- Cunhang Fan *et al.*, "ICaps-ResLSTM: Improved capsule network and residual LSTM for EEG emotion recognition." *Biomedical Signal Processing and Control,* Vol. 87p. 105422, (2024).

21- Hanie Asemi and Nacer Farajzadeh, "Improving EEG signal-based emotion recognition using a hybrid GWO-XGBoost feature selection method." *Biomedical Signal Processing and Control,* Vol. 99p. 106795, (2025).

22- Sander Koelstra *et al.*, "Deap: A database for emotion analysis; using physiological signals." *IEEE transactions on affective computing,* Vol. 3 (No. 1), pp. 18–31, (2011).

23- Aleksandra Dura and Agnieszka Wosiak, "EEG channel selection strategy for deep learning in emotion recognition." *Procedia Computer Science,* Vol. 192pp. 2789–96, (2021).

24- Andrea Apicella, Pasquale Arpaia, Francesco Isgro, Giovanna Mastrati, and Nicola Moccaldi, "A survey on EEG-based solutions for emotion recognition with a low number of channels." *IEEE Access,* Vol. 10pp. 117411–28, (2022).

25- Ezat Ahmadzadeh, Hyunil Kim, Ongee Jeong, Namki Kim, and Inkyu Moon, "A deep bidirectional LSTM-GRU network model for automated ciphertext classification." *IEEE Access,* Vol. 10pp. 3228–37, (2022).

26- Cheng-Jie Yang, Nicolas Fahier, Wei-Chih Li, and Wai-Chi Fang, "A convolution neural network based emotion recognition system using multimodal physiological signals." in *2020 IEEE International conference on consumer electronics-Taiwan (ICCE-Taiwan)*, (2020): *IEEE*, pp. 1–2.

27- Bahareh Nakisa, Mohammad Naim Rastgoo, Dian Tjondronegoro, and Vinod Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors." *Expert Systems with Applications,* Vol. 93pp. 143–55, (2018).

28- Wei Liu, Wei-Long Zheng, and Bao-Liang Lu, "Emotion recognition using multimodal deep learning." in *International conference on neural information processing*, (2016): *Springer*, pp. 521–29.