International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V74I1P105

Volume 74 Issue 1, 65-84, January 2026

Original Article

A Deep Learning-Based BCI System for Emotion Classification Using EEG Signals

Raju Ramakrishna Gondkar[1], Surekha R Gondkar[2], Ramkumar Sivasakthivel[3], R. Gobinath[4], Manikandan Rajagopal[5]

[1,3,4,5]*Christ (Deemed to be University), Bangalore, Karnataka, India.*
[2]*Department of Electronics and Communication, BMS Information Technology and Management, Bangalore, Karnataka, India.*

[3]*Corresponding Author : ramkumar.s@christuniversity.in*

Received: 28 September 2025 Revised: 02 December 2025 Accepted: 26 December 2025 Published: 14 January 2026

Abstract *- Electroencephalography-based Brain-Computer Interfacing (EEG-BCI) technologies allow for effortless interaction between external hardware and the human brain through monitoring its electric signals. These systems rely on EEG recordings, which provide non-invasive and real-time neural information through electrodes placed on the scalp. To advance emotion-recognizing efficiency and accuracy, this study proposes a deep learning-based method that can extract valuable temporal and spatial information from EEG signals. The proposed model includes the use of a Graph Convolution Network (GCN) for learning spatial relationships between different EEG channels to model the data in graph form and gain features through that modelling. A Convolutional Autoencoder (CAE) is then used to compress data to low dimensions and to reconstruct it so that major features are not ignored. Furthermore, the model uses an Attention-based Bidirectional Gated Recurrent Unit (ABiGRU) for temporal classification, which can emphasize the most important time steps in both backwards and forward directions. Two standard datasets are employed to test the developed approach. The DEAP dataset is used for emotion recognition with a binary response, and SEED is used with multi-class classification. The model attains great results of 98.12% accuracy on DEAP and 97.58% on SEED datasets. The very high performances show the efficacy of the model for decoding emotional states from EEG signals and very strong potential for real-time emotion recognition in affective computing and BCI.*

Keywords *- Graph Convolutional Network, Convolutional Autoencoder, Attention-Based Bidirectional Gated Recurrent Unit, DEAP, SEED.*

1. Introduction

EEG-BCIs are high-tech systems equipped to facilitate a direct, immediate interface between the brain and an external device. EEG technology records brain signals and decodes them into application commands [1]. They have attracted massive attention in the fields of neuroscience, biomedical engineering, and human-computer interaction due to their non-invasive nature and versatility in applications [2]. Electrodes are set on the scalp in a non-invasive neuroimaging method used to detect the brain's electrical activity [3]. BCIs are considered safe and utilize EEG signals that reflect dynamic brain activity generated by the collective functioning of numerous neurons [4].

1.1. Challenges in EEG Signal Processing

However, the signals are nonlinear and non-stationary most of the time, susceptible to interferences from muscle artifacts, environmental noise, and hardware imperfections. In addition, EEG data are high-dimensional and largely differ from person to person; hence, it becomes difficult to gain good and consistent interpretation and for the model to be generalized [5].

1.2. Role of DL in EEG Emotion Recognition

To address these limitations, DL has become more popular because it can separately acquire complex structures from unprocessed EEG signals. Although DL models typically demand more computational resources and larger labeled datasets, they provide superior generalization and reduce the necessity for manually engineering features [9]. DL also contributes to signal enhancement, noise reduction, and temporal feature discovery in EEG analysis. BCIs are generally classified as invasive or non-invasive.

Non-invasive BCIs, such as EEG-based systems, are widely used due to their lower risk, ease of deployment, and user-friendliness. They support real-world applications. To get high-resolution recordings with low interference, invasive BCIs place electrodes directly into the brain [10]. While they share superior signal unity, their clinical use is limited due to

surgical risks and ethical considerations. These systems are mainly used in emotional expression in patients with severe neurological impairments.

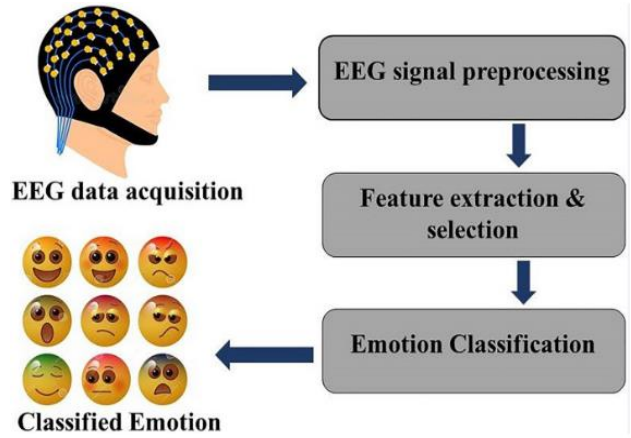

Fig. 1 EEG Signal processing flow for emotion detection

1.3. Advancements in EEG-BCI Applications

Recent advancements in signal processing and wearable hardware are rapidly moving EEG-BCIs from research labs to real-world use. They are making substantial progress in areas like healthcare (e.g., emotion monitoring) [11]. Given the potential of EEG-BCIs, it is essential to improve signal quality. With continued innovation and responsible development, EEG-BCIs can enhance human abilities and improve quality of life across medical, social, and interactive domains [12].

1.4. Limitations of Existing Methods

DL methods have made major and impressive advancements in emotion recognition. At the same time, there are still several common restrictions that have not been overcome. The models based on CNN concentrate largely on the extraction of spatial features. However, their main drawback is the inability to model the temporal dimension. The case is different for LSTMs and GRUs, which, to a great extent, capture temporal dependencies but fail to consider the spatial relationships between EEG channels. Methods based on GCNet are similar in that they incorporate the spatial structure but still use shallow temporal encoders that are not able to capture the long periods and perform the gradual transitions in the emotional state. The combination of the Transformer with CNN or RNN achieves high accuracy, but at the same time, incurs high computational costs and training instability, which is one of the reasons they are not suitable for real-time BCI databases.

1.5. Research Gap

Although using DL techniques to recognize emotions on EEG signals has brought about significant improvements, there are still many limitations in the architectures. The CNN and LSTM models do not have the capacity to discover the spatial organization of the EEG electrodes, and the opposite applies to GCN methods, which do not have strong temporal modeling. However, the transformer-based hybrids give the highest accuracy, but at the same time, they bring about high computational cost and instability during the training process. Moreover, there are various studies that depend on single-stage feature extraction, and this solely limits the richness of the spatiotemporal patterns that are being learned.

Most recent research has pointed out even more limitations. A majority of current CNN/LSTM models assess EEG channels as separate and do not define connectivity according to their own criteria, ultimately resulting in a partial representation of the functional brain relations. However, views based on GCN have a strong understanding of the spatial structure but struggle with limited usage of deep recurrent units in their ability to track dynamic emotional shifts. Moreover, the approach based on the transformer requires massive labeled datasets and suffers from unstable convergence, which is a limitation for subject-dependent EEG data. Also, the prior works predominantly test their models on one data set, which restricts their generalization claims. Not many studies have successfully united feature compression, graph-based spatial learning, attention-based temporal modeling, and computational efficiency into a single architecture. The existence of these gaps also drives the researchers to come up with a more integrated and robust framework.

1.6. Research Problem and Hypothesis

A wide range of DL methodologies have been presented for recognizing emotions from EEG signals; however, the current techniques do not combine feature compression, learning of spatial dependency, and attention-driven temporal modeling within the same framework. It is a major drawback, since it does not allow the models to go deep in detecting the complex spatiotemporal dynamics that are associated with EEG signals. Furthermore, the need for such architectures has grown, which would be capable of providing high accuracy without accompanying huge computational costs, especially in the framework of real-time BCI applications.

By combining CAE to reduce the dimensionality of features, GCN to represent the spatial relationships among EEG channels, and an Attention-Based BiGRU to capture temporal dependencies in both forward and backward directions, more distinct spatiotemporal representations will be produced, leading to a considerable gain in classification performance over current DL models.

1.7. Problem Statement

The technique reduces EEG data size using a CAE while preserving its important temporal and spatial features, and it can also rebuild the original signals. The reduced data is passed to GCN, which considers EEG channels as points in a graph to learn spatial relationships between each other. Then,

an attention-based bidirectional GRU analyzes the sequence data in a backward and forward manner and uses attention to select the most relevant time steps. DL methods such as CAE, GCN, and Attention-BiGRU facilitate the automatic learning of complicated spatial and temporal patterns in EEG signals needed in BCI applications. Data normalization is performed before applying the method to both DEAP and SEED datasets so that all samples are on a consistent scale. The aims of the study are as follows:

- To develop an effective DL-based framework for classifying EEG signals in BCI applications.
- To use CAE for extracting and compressing relevant temporal and spatial EEG features.
- To apply a GCN to learn spatial interactions between EEG channels.
- To use an Attention-Based Bidirectional GRU to capture meaningful physiological temporal dependencies and emphasize crucial time-steps.
- To evaluate the proposed method using standard EEG datasets such as DEAP and SEED, with consistent data normalization.
- To validate the model's accuracy and reliability using several performance indicators.

Section II reviews the related work and previous DL research, referencing some important concepts and advancements. Section III briefly covers data sets used in the study and the implementation of the research methodology. Section IV covers related evaluation criteria for assessing model accuracy. Section V summarizes the findings and concludes the study.

2. Literature Survey

In this section, a comprehensive study and evaluation of previously developed approaches and algorithms for EEG-based identification systems is presented. Emotion identification has recently attracted considerable interest, and Machine Learning (ML) and DL have become essential tools across various fields. Table 1 shows the disadvantages and advantages of existing models.

Although recent DL approaches have markedly improved EEG-based emotion recognition, several recurring limitations remain. Many CNN- or LSTM-centric models excel at capturing either spatial or temporal patterns but rarely both with equal effectiveness, producing suboptimal spatio-temporal representations. Transformer- and capsule-based hybrids boost representational power but typically incur high computational cost, larger memory, and greater training instability on small, subject-dependent EEG datasets. Graph-based approaches capture inter-electrode relationships effectively but are often applied directly on high-dimensional raw features, allowing noise and redundant information to propagate into the graph representation. Furthermore, a number of studies report strong single-dataset performance without rigorous cross-validation, limited ablation analyses, or explicit reporting of inference latency and model size — all of which are critical if the system is intended for real-time BCI deployment. These gaps motivate a design that (i) reduces input noise and dimensionality before graph construction, (ii) separates compact spatial encoding from efficient temporal modeling, and (iii) reports computational efficiency and robustness measures alongside accuracy. The CAE–GCN–ABiGRU pipeline proposed here is designed to meet these needs by combining early-stage compression, topology-aware graph learning, and attention-guided BiGRU temporal modeling to balance accuracy, robustness, and runtime constraints.

The research in [13] presented a state-of-the-art ML model that extensively utilized the temporal and spatial properties of EEG channels to enhance emotion EEG-based classification. To enhance the learning of features and the generalization of models across different datasets, the model integrated attention mechanisms with GRU. On the EEG Brainwave Dataset, the model achieved a high classification accuracy.

An ML-based real-time emotion detection model [14] was developed in the study, which estimated VAD (valence, dominance, and arousal) every 5 seconds. The DEAP and SEED datasets were used. Optimal band powers for the top eight channels were determined by applying Random Forest (RF), Extra-Trees, Principal Component Analysis (PCA), and Power Spectral Density (PSD). Different models were also tested with cross-validation and shift-based data division. Extra-Trees performed better than average after evaluation.

The paper [15] developed the Dual Attention Mechanism Graph Convolutional Neural Network (DAMGCN method). To extract representative spatial information, the brain network was modeled by GCNs. Further, while assigning weights to electrode networks and signal frequency bands, the Transformer model's self-attention mechanism prioritized certain brain locations and band frequencies. The process of attention mechanism effectively demonstrated the weight assignment obtained using DAMGCN. It was implemented and tested on three datasets, DEAP, SEED-IV, and SEED, with the best result obtained on the SEED dataset. The research in [16] was carried out with the aim of exploring another approach to DL-based emotion recognition from EEG input. The approach employed an autoencoder so that LSTM networks are integrated with 2D CNNs. The autoencoder layers best encoded the input signals into a lower dimension; the 2D CNN/LSTM layers then captured the emotion classes in the data efficiently, both temporally and spatially. Experimental results showed higher performance in four-category emotion classification, with an accuracy of 90.04%, conducted using the publicly available DEAP dataset.

Based on multi-domain features, the model in [17] called the Multi-domain Emotion-aware Spatiotemporal Capsule Transformer Network (MES-CTNet) was designed for EEG-based emotion recognition. The model's main components consist of an Squeeze-and-Excitation (SE) block and an Efficient Channel Attention (ECA) block. There was also a temporal coding layer based on Transformers incorporated into a multichannel Capsule neural Network (CapsNet). In the first instance, the multi-domain feature's space-feature-time properties were fused and employed as model inputs. The enhanced CapsNet accommodated these feature maps and performed the extraction of local emotion features. The last emotional state was determined by a time-oriented coding layer based on transformers, which kept regularly catching and recording emotional characteristic data on a global scale. Experimental investigations were carried out on the DEAP and SEED datasets, two extensively used benchmarks with varied emotion labels. MES-CTNet reached amazing accuracy on the DEAP dataset.

In [18], the study minimized individual differences and captured emotion-relevant information using a combination of a Four-Dimensional Convolutional Recurrent Neural Network (4DCRNN) and Random Forest Weights (RFWs). The model was to enhance the accuracy and generalizability of emotion identification through integration. Identification accuracy was then assessed using DEAP and SEED in experimental evaluations. RFW-4DCRNN exhibited excellent emotion recognition performance with respective high accuracies. The study in [19] presented an architecture for emotion identification using EEG signals, combining CNNs and Transformers. To learn well from global patterns, the architecture made the best of both the self-attention mechanism of Transformers and the spatial pattern detection capabilities of CNNs. Performance tests on the architecture employed the DEAP dataset, which contained EEG recordings from 42 subjects. The findings indicated that the design attained an accuracy of 87% for the DEAP dataset.

The approach in [20] used a hybrid model, a Bidirectional Long Short-Term Memory for temporal dependency and a 1D-CNN for extracting features, to improve emotion classification through learning. The method was evaluated using the DEAP dataset. Additionally, a channel selection method was also presented to determine the EEG channels most relevant to emotion recognition, thereby lowering computational complexity and maintaining accuracy. By selecting the best eight-channel model, the method reached an accuracy of 85.16%.

The research work presented in [21] attempted to contribute to the theory of multimodal emotion detection by analyzing the potential of combining EEG signals with generating state-of-the-art models of facial emotion analysis, GRU, LSTM, and Transformer. The GRU model proved efficient on average with 91.8% accuracy. The method described in [22] involved EEG data classification into five different emotional states using a combination of ResNet18 and differential entropy. The various and deep nature of these states of emotion, the method first calculates the differential entropy of the EEG signal. It was followed by the ResNet18 network learning feature representations from the differential entropy values using residual connections that can effectively capture the spatiotemporal dynamic characteristic of complex emotional EEG signals. The method was verified on a dataset SEED-V, through experimentation with a satisfactory level of accuracy.

The work of [23] considered an innovative DL approach called TSF-MDD (Major Depressive Disorder), merging data from the time, frequency, and spatial domains. The first stage of the data reconstruction scheme involved creating four-dimensional EEG signals with reference to time, space, and frequency. The data were then fed to a 3D-CNN and CapsNet-based model and processed for feature extraction across domains. To avoid data leakage, subject-independent data partitioning was employed during training and testing. The method showed an accuracy of 92.1%. A novel method for emotion recognition using exclusive datasets and DL concepts in [24]. The method combined attention layers with LSTM algorithms, and the main feature of the methodology was its use of cost, compact biometric sensors, and complex sensor systems. EEG and its developmental phases have become standard. Even with the inexpensive sensor setup, the classifier attained a remarkable accuracy of 93.75%.

The research in [25] intended to improve accuracy for four- and three-class emotion classification. The model contained N emotion classes, with each classifier functioning as an Adaptive Neuro-Fuzzy Inference System (ANFIS). The features that were best distributed were selected to be the input vectors for the respective ANFIS architectures; they were then trained. Outputs of the trained ANFIS models were pooled further to construct a feature vector for input to adaptive networks, allowing the system to perform emotion recognition. Results showed 73.49% and 95.97% on the DEAP and Feeling Emotions datasets, respectively. In the study [26], Domain Adversarial Neural Network with Multiple Adversarial Tasks (DANN-MAT) was used. An emotion classifier was designed to be adversarially challenged by multiple emotion-unrelated classification tasks, and the results removed irrelevant data while preserving emotion-related characteristics. The results showed that subjects' emotion categorization accuracy improved with fewer tasks used and that the model's generalizability was enhanced with more adversarial challenges. Applying the model method to the SEED-IV and SEED sets provided state-of-the-art results.

The study in [27] used a DL-based model that initially faced difficulties in concurrently recording the spatial topological and spatial activity components of EEG data. To address the Spatial Activity Topological Feature Extractor

Module (SATFEM), an extractor module for topological features and spatial activity in EEG signals was developed. Subsequently, Domain-adaptation Spatial-feature Perception (DSP)-EmotionNet was constructed with SATFEM as its feature extractor, which significantly enhanced the model used with cross-subject emotional EEG identification tasks. With highly accurate cross-subject EEG identification of emotions, the model surpassed state-of-the-art methods on the datasets SEED and SEED-IV. Recent hybrid architectures for EEG emotion recognition commonly combine transformers, capsule networks, GCNs, and CNN/LSTM blocks to capture spatial–temporal patterns (e.g., DAMGCN, MES-CTNet, DANN-MAT, and related GCN/Transformer hybrids). While these works achieve high accuracy, they typically (i) rely on heavy transformer/capsule modules that increase FLOPs and memory footprint, (ii) tightly couple spatial learning and global self-attention, which raises training instability for small subject-dependent EEG sets, and (iii) often omit early-stage compression that reduces noise while preserving temporal structure. For example, DAMGCN emphasizes dual attention on graph nodes and frequency bands but remains computationally heavy for real-time deployment. MES-CTNet fuses capsule and transformer blocks to exploit multi-domain features, yet its capsule-transformer pipeline increases inference latency. Domain adversarial approaches such as DANN-MAT improve cross-subject generalization but do not directly address lightweight temporal–spatial encoding for low-latency BCI. The proposed CAE–GCN–ABiGRU differs in three critical ways. First, insert a CAE compression stage prior to graph construction to denoise and reduce dimensionality while retaining temporal continuity. This reduces downstream GCN/GRU compute and improves robustness to low-SNR EEG segments. Second, graph construction uses a lightweight adjacency design tuned to electrode topology and CAE-latent features (N × 32 node features), allowing effective spatial modeling with low GFLOPs compared to transformer/capsule hybrids. Third, the ABiGRU with attention focuses on salient time steps while avoiding the training instability and data of transformer layers. These combined choices yield a practical tradeoff: accuracy comparable to heavy hybrids but with substantially lower model size, FLOPs, and CPU inference time.

Table 1. Advantages and disadvantages of existing models

Ref	Methods	Dataset	Advantages	Disadvantages
[13]	GRU	EEG Brainwave Dataset	Utilizes spatial and temporal EEG features with high accuracy	The model overfits on small datasets
[14]	PCA, PSD, RF, Extra-Trees	DEAP, SEED	Enables accurate real-time emotion detection with minimal EEG channels and low complexity.	Model performance can degrade with noisy channels.
[15]	DAMGCN	DEAP, SEED, SEED-IV	Utilizes graph and attention mechanisms to prioritize critical brain regions and frequencies.	Graph construction and attention layers significantly increase model complexity and training time.
[16]	2D CNN, LSTM	DEAP	Combines spatial and temporal learning for efficient four-class emotion recognition.	Needs careful tuning of encoding dimensions
[17]	MES-CTNet (CapsNet, Transformer)	DEAP, SEED	Fuses multi-domain features for superior temporal and local emotion feature extraction.	High model complexity
[18]	RFW-4DCRNN	DEAP, SEED	Integrates 4D data modeling with weighted voting to reduce inter-subject variability.	4D data modeling increases the model's complexity and demands large training datasets.
[19]	CNN	DEAP	Improves global emotion pattern interpretation by contextual learning.	Limited interpretability of attention outputs and higher computational overhead.
[20]	1D-CNN, Bi-LSTM	DEAP	Reduces computational complexity while maintaining strong classification performance.	Channel selection reduces performance.
[21]	Multimodal (GRU, LSTM, Transformer)	EEG + Facial Emotion Data	Multimodal fusion enhances emotion classification accuracy.	Requires synchronized multimodal data, which complicates data collection and processing.
[22]	ResNet18	SEED-V	Learns deep spatiotemporal features using entropy-based EEG representations.	Relies heavily on entropy features that do not capture all emotion-relevant variations.
[23]	TSF-MDD (3D-CNN, CapsNet)	Mumtaz2016	Extracts cross-domain features with high generalization using subject-independent data.	Subject-independent partitioning requires large and diverse datasets to avoid overfitting.

[24]	LSTM	Private Dataset	Achieves high accuracy using minimal sensor configurations.	Lower-cost sensors introduce more noise.
[25]	ANFIS	DEAP, Feeling Emotions	Handles nonlinear, uncertain EEG features with interpretable neuro-fuzzy logic.	Scalability is limited due to fuzzy rule explosion in multi-class scenarios.
[26]	DANN-MAT (Multi-adversarial tasks)	SEED, SEED-IV	Enhances the model using adversarial learning on emotion-irrelevant tasks.	Adversarial training is sensitive to hyperparameters and causes instability during learning.
[27]	SATFEM	SEED, SEED-IV	Boosts cross-subject accuracy by extracting spatial activity and topological EEG features.	Cross-subject generalization remains suboptimal, especially in highly variable real-world data.

Compared to existing models, the proposed GRU-GCN framework offers a balanced trade-off between accuracy and efficiency. While prior methods like DAMGCN and MES-CTNet achieve strong performance, they suffer from high model complexity and longer training times. Simpler models, such as 1D-CNN or basic GRU, show efficiency but often lack deeper spatial or temporal insights. By combining GCN for spatial feature extraction and Attention-Based Bidirectional GRU for temporal classification, the proposed model effectively captures inter-channel relationships and critical time steps. It avoids the over-complexity of multimodal or 4D approaches, making it more practical for real-time applications. Moreover, it performs competitively better than many models on the SEED and DEAP datasets in terms of classification metrics. The GRU-GCN architecture for EEG-based emotion recognition offers significant potential with regard to scalability and accuracy.

3. Proposed Methodology

As illustrated in Figure 2, the proposed model provides a framework for EEG-based BCI classification with the DL methods. Using a CAE, the quantity of EEG data can be reduced without losing its real spatial and temporal features.

The CAE will take the data by compressing and reconstructing it. These compressed features are then passed to the GCN, which learns the relationship between EEG channels by considering them to be points connected in a graph. The GCN output is then fed to an attention-based bidirectional GRU that passes over the data backward and forward and uses the attention mechanism to give more importance to critical time steps.

It starts with EEG activity, the recording, while brain activity is measured through electrodes positioned on the head. These raw signals are cleaned with procedures including normalization into smaller epochs to maintain uniform and dependable input. Afterward, the clean EEG data are compressed with the CAE, preserving key spatial and temporal features.

The reduced features are passed to a GCN to learn spatial relationships between EEG channels. Following this, an attention-based bidirectional GRU captures time-based patterns and concentrates on those moments crucial for the accurate result. Finally, standards are used to assess the efficiency of the models in detecting and identifying the emotional states derived from the EEG data.

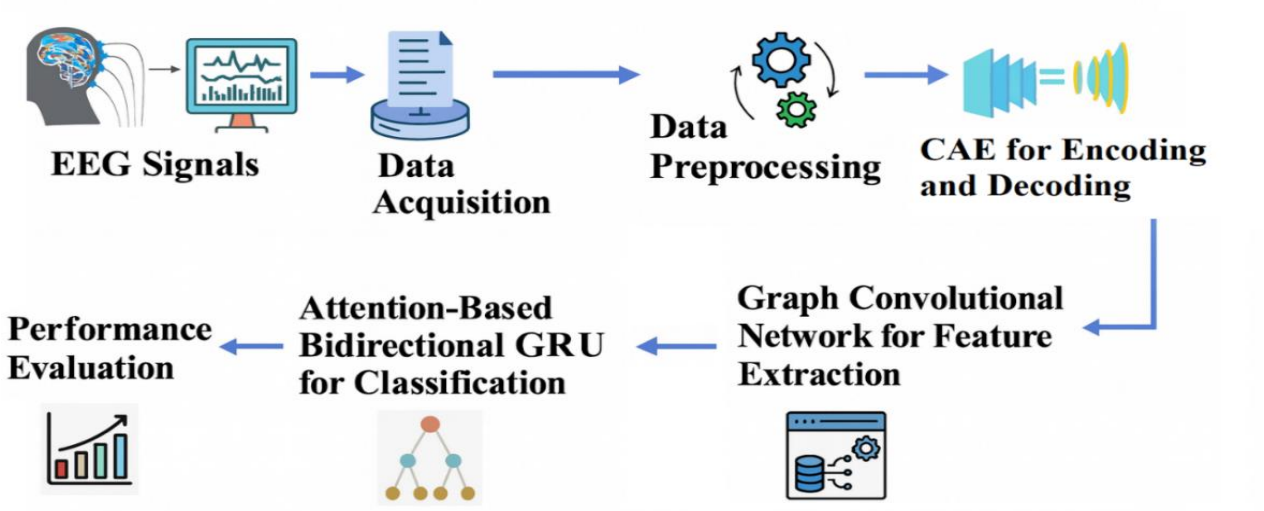

Fig. 2 Proposed model flow diagram

3.1. Datasets

Open-source EEG datasets are used in this segment for emotion recognition. These datasets advance the research space into EEG signal emotion recognition.

3.1.1. DEAP

The dataset for DEAP is a frequently used benchmark in the domain of emotion recognition with EEG and physiological data. It contains records of 32 participants from ages 19 to 37 years, each of whom had watched forty-one-minute music video clips that were selected to induce a broad spectrum of emotional reactions. The group was evenly mixed in terms of gender. Once after viewing a video, the participants rated their emotional state along the arousal and valence dimensions on a scale of 1 to 9. EEG was obtained with 32 electrodes placed according to the worldwide 10–20 method and sampled at a frequency of 512 Hz. They were also obtained along with the EEG. Preprocessing involved down-sampling to 128 Hz, bandpass filtering, and removal of artifacts through Independent Component Analysis (ICA). The raw data is included in the dataset as well. They are also compatible with valence-arousal models. For research purposes, the dataset is a landmark in EEG-based emotion recognition work.

3.1.2. SEED

The dataset is generally used for emotion recognition through EEG. It consists of EEG recordings collected from fifteen individuals while they were shown fifteen handpicked short Chinese films, four minutes each, to evoke positive, neutral, or negative emotions. The EEG signals were gathered by an ESI NeuroScan system comprising 62 channels at a 1000 Hz sample rate. They essentially guaranteed that precise brainwave data would be collected. Each participant was engaged in three sessions, held on three separate days to accommodate variability across time. The data are labeled on the basis of emotional states as recorded by individuals themselves, to offer a reliable ground truth for the supervised learning tasks. To be clear with the signals, some preprocessing steps were taken, such as bandpass filtering and artifact removal. Thus, SEED allows three-emotion classification, while SEED-IV has been extended to allow for the classification of four emotions, including fear. When it comes to the evaluation of ML and DL models for emotion recognition, it has gained wide acceptance. The dataset is available to the general public for academic research use. Because of its consistency, enriched EEG properties, and emotional variance, it serves as a resource of importance in affective computing and EEG-BCI research [28].

3.2. Data Preprocessing

Preprocessing should be done to translate the raw EEG signals into analysis-ready data in EEG-based BCI systems. The next step involves the application of artifact removal techniques, using ICA to mitigate interferences. The signal is afterwards segmented into epochs corresponding to stimuli or events and normalized to get them all on the same scale. These steps will immensely help in accuracy and fast feature extraction for EEG-BCIs and in classification.

3.3. Data Normalization

EEG-BCI data normalization brings about the scaling of the EEG features into a uniform range so that ML models have better stability during performance. The more common methods are min-max normalization, which normalizes values into a uniform distribution. It would eliminate the feature with a bigger magnitude from dominating the model training procedure and would lessen the effect of variability from individuals. For EEG-based emotion detection algorithms to obtain balanced and accurate classification, data normalization must be taken into consideration [29].

3.4. Graph Convolutional Network for Feature Extraction

In the construction of a GCN, the two primary functions are convolution and pooling. Graph convolution and feature transformation are the two primary components of the graph convolution layer's processing of graph signals. Figure 3 illustrates the architectural structure of the GCN model.

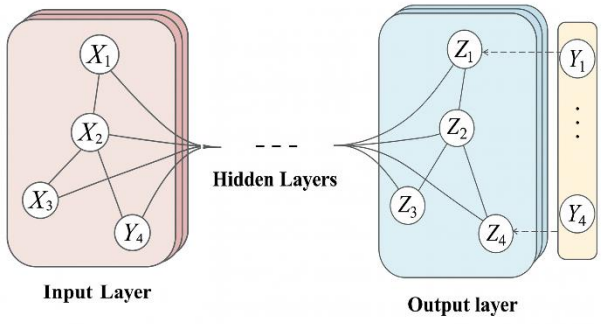

Fig. 3 Architecture of GCN

3.4.1. Graph Convolution

Equation (1), which is based on a generalized convolution method, uses the graph Laplacian, since the traditional convolution operation is not directly applicable to graphs. It is possible to express graph convolution as Equation (2).

$$Y = g_\theta(L_s)X, \tag{1}$$

$$Y = U_s g_\theta(\Lambda_s)U_S^T X \tag{2}$$

A Fourier domain filter with parameters $\theta \in R^n$, where $X \in R^{m\times s}$ is the input matrix, $g_\theta(.) = diag(.)$ and $g_\theta(\Lambda_s)$ is a function of the eigenvalues. Both $Y \in R^{m\times s}$ and $U_s \in R^{m\times m}$ are eigenvector matrices; the latter is the result of graph convolution. In the conventional graph convolution, however, calculating $g_\theta(\Lambda_s)$ is a challenging and computationally costly process. The issue was addressed by introducing a Chebyshev polynomial expansion. Up to K^{th} order, it implies that $g_\theta(\Lambda_s)$ can be accurately given by a formulation using

Chebyshev polynomials $T_k(x)$ as an expansion. As shown in Equation (3).

$$g'_\theta(\Lambda_s) = \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\Lambda}_s) \quad (3)$$

Equation (4) shows that a recursive calculation, $T_K(x)$, is performed when θ_k is a vector of Chebyshev coefficients. Equation (5) can be used to obtain $\widetilde{\Lambda}$, which is a normalized version of Λ.

$$\begin{cases} T_0(x) = 1, T_1(x) = x, \\ T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), k \geq 2, \end{cases} \quad (4)$$

$$\widetilde{\Lambda} = \frac{2\Lambda_s}{\lambda_{max}} - I_n, \quad (5)$$

where $\widetilde{\Lambda}$ components range from -1 to 1 and λ_{max} are the greatest elements of Λ_s. The meaning of a signal x convolution with a filter g'_θ is used to calculate the graph convolution. As shown in Equation (6).

$$Y = \sum_{k=0}^{K-1} \theta_k U_s T_k(\widetilde{\Lambda}) U_S^T X, \quad (6)$$

3.4.2. Feature Transformation

To filter the signal without affecting the feature's dimension, it can use the graph convolution technique. As demonstrated by Equation (7), once an adjustable weight matrix has been applied to the graph signal, the feature's dimension can be changed.

$$X' = Cheb\big(X, W^{(i)}\big) = YW^{(i)} \quad (7)$$

$X' \in R^{m \times j}$ is the final output of the graph convolution of i^{th} layer, the function of $Cheb(\cdot)$ Chebyshev convolution, and $W^{(i)} \in R^{m \times j}$ is a trainable weight matrix of that layer i^{th} [30].

In BCI-EEG, a GCN represents the physical or functional interactions between the electrodes by modeling their spatial relationships as a graph. Because EEG signals are not geometric, GCNs can detect inter-channel relationships that regular CNNs miss. Graph convolutions synthesize data from nearby nodes to process the input features (raw signals, frequency power, etc.) from each electrode. In the brain, the operation aids in the learning of spatial patterns. If two electrodes are close enough in proximity or have similar signals, their effects on one another can be described by the adjacency matrix. To make it more discriminative, these learned properties are fed through a number of layers. The collected features are then interpreted by a classifier for applications such as emotion recognition.

3.5. Convolutional AUTOENCODER

AE is a typical DL technique that can learn effective representations of unlabeled data and perform feature extraction and dimensionality reduction. It works by building an encoder and decoder, then recreates the output after mapping raw data to any hidden areas. In terms of reducing data dimensions, AE is similar to PCA (Principal Component Analysis). AE basically uses dimensionality reduction algorithms of high-dimensional raw data to obtain representative features. The computational complexity and robustness of AE are great, despite the fact that it is usually built on fully connected networks. However, convolutional layers' local connections and weight-sharing properties allow them to require fewer parameters to acquire richer information.

As a result, CAE employs convolutional layers rather than fully connected ones. To help the encoder identify the hidden space model, the input data is down-sampled, and a latent representation with fewer dimensions is produced. The hidden layer in the work uses input features that are typical of the whole dataset, and the feature extraction method is unsupervised. The goal of CAE is to make the decoder's reconstruction more similar to the encoder's input. Rather than determining the optimal replacement of input data, CAE is used in the study to learn representations in latent vectors. Various tasks, including dimension reduction and classification, are performed by the well-trained encoder following CAE training. The decoder is used for information reconstruction, whereas the encoder converts temporal data into potential space data. As illustrated in Equation (8), the encoding process is explained below.

$$h^k = \sigma\,(x * W^k + b^k) \quad (8)$$

Among these, it finds σ, the activation function of the method CAE, b^k the offset of the entire mapping ($*$ indicating a convolution operation) w is the shared weight matrix, x the input feature, and h^k the possible representation of the k^{th} mapping. As shown in Equation (9), the construction process is described below.

$$y = \sigma\big(\sum_{k \in H} h^k * \widehat{W}^k + C\big) \quad (9)$$

H stands for the potential mapping group, y is the reconstruction features used for output, and the flipping of weights is $\widehat{w}$, the bias of each input channel is c. Decrease reconstruction errors by optimizing weight and bias through backpropagation. MSE (Mean Squared Error) is the activation function that has been selected, and it is represented by the following Equation (10).

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad (10)$$

where x_i and y_i are the original features and the reconstructed features after ith iteration, respectively, and the dimensionality of an input vector is N. In the study, a CAE structure for the encoder that primarily uses activation and normalization layers, max pooling, and convolutional layers.

A feature extractor is a convolutional layer. Using convolution kernels, it applies convolutional calculations to the input signal while preserving its primary characteristics. The amount and dimensionality of features are reduced using iterative multiple convolution and pooling methods. The decoder is made up of the same components as the encoder, but it operates in the opposite way. For signal recovery in the decoder, the recovered features are utilized for deconvolution and up-sampling reconstruction on the latent space. The intricate architecture of CAE is seen in Figure 4.

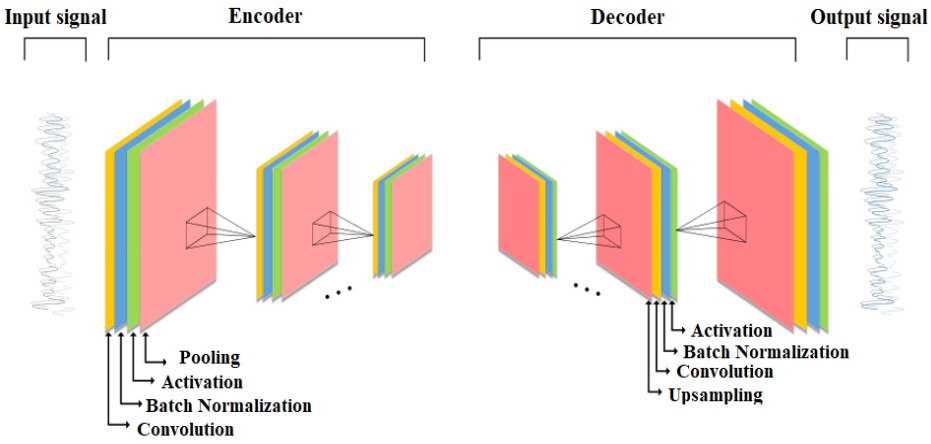

Fig. 4 Structure of CAE

The encoder and decoder are making use of an unsupervised learning method. The training features are used to obtain all the parameters, and no label information from the data is used. The most effective parameters learned by CAE are memorialized throughout its training phase. A new layer, a dense layer, and a softmax classifier have been added in place of the decoder. Once the final classification task is finished, EEG signal features are extracted using the encoder and then applied to the model. The Rectified Linear Unit (ReLU) is the activation function of the paradigm. All the convolutional layers have a 3×3 convolution kernel size [31].

In BCI-EEG, a CAE is employed to generate compact and meaningful representations from EEG data automatically. The encoder part uses convolutional layers to subtract geographical and temporal information from the raw EEG data, hence decreasing its dimensionality. It helps in removing background noise while collecting important patterns associated with brain activity. Decoders employ transposed convolutions to recover the original input from the compressed feature map. To keep important data while reducing the reconstruction error. CAEs are effective in feature learning and EEG data denoising. Then, these acquired characteristics can be applied to classification tasks like emotion identification.

3.6. Bidirectional Gated Recurrent Unit

The BiGRU is used in the study to reproduce the first scenario. The most well-known DL models for processing sequential data are Residual Neural Networks (RNNs). Yet, issues including disappearing gradient and growing gradient have been noted to impact RNNs. As a result, RNNs are unable to detect long-term dependencies. These problems have inspired the development of several specific RNN architectures, such as GRU and LSTM. The latest feature of the LSTM model has the capability to preserve long-term dependency utilizing Input, Forget, and Output (IFO) gates. LSTM requires more inputs than GRU. It can be acquired faster than what LSTM provides. In addition, while LSTM necessitates four gates, GRU requires only two: the update and reset gates". Identifying the best way to combine the current input with the data that has already been saved is the main function of the reset gate. The amount of required historical memory can be controlled via the update gate. Figure 5 illustrates the GRU architecture, with the Equations detailed in (11) to (14).

$$\text{Update gate: } (Z_t) = \sigma(W_z h_{t-1} + U_z X_t) \quad (11)$$

$$\text{Reset gate: } (r_t) = \sigma\,(W_r h_{t-1} + U_r X_t) \quad (12)$$

$$\text{New state: } (h_t) = (Z_t \circ C_t) + ((1 - Z_t) \circ h_{t-1}) \quad (13)$$

$$\text{Cell state: } (C_t) = \tanh(W_c(r_t \circ h_{t-1}) + U_c X_t) \quad (14)$$

The sigmoid function is represented by r, and X_t is the input vector at time t. Both the h_t and h_{t-1} state vectors are thought to be confidential. W_r stands for the reset of the gate, update the gate is W_z, and current cell state W_r in the parameter matrices. The hidden state of the vector h_{t-1} is connected to all these matrices. U_r stands for the reset gate, U_z for the update gate, and U_c for the current cell state in a parameter matrix. The vector of input X_t is connected to each of these matrices. Fundamental matrix multiplication is needed in the framework of the circ, which transforms the state h_t to represent the vector output.

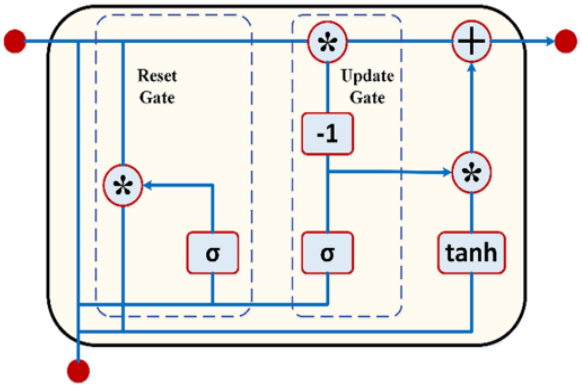

Fig. 5 Gated Recurrent Unit (GRU) architecture

The classic GRU approach of forecasting a time series evaluates the values recorded at previous time points. BiGRU proves its worth in problems like speech recognition, proper text representation, and missing data prediction, primarily when the prior and subsequent values are accessible. The BiGRU generic architecture is illustrated in detail in Figure 6. Two new GRU layers, operating in opposite directions, are integrated in the BiGRU, an innovative design. Each hidden state is stored in its own layer; Separate layers are used to store the forward and reverse states, respectively. For each time step, BiGRU executes forward pass computation on a given input sequence constructed as $\dots, X_{t-n}, \dots, X_{t-2}, X_{t-1}$. X_t for the step time $\dots, t, t-n, \dots, t-2, t-1, t,$ and the input sequence is subsequently subjected to reverse pass processing for each of the subsequent time steps $t, t-1, t-2, \dots, t-3$, ...Consolidation of the hidden states occurs after both backward and forward passes have been completed. The observed sequence $\dots, y_{t-n}, \dots, y_{t-2}, y_{t-1}, y_t$ y_t is generated, when the highly connected layer analyzes its final hidden states.

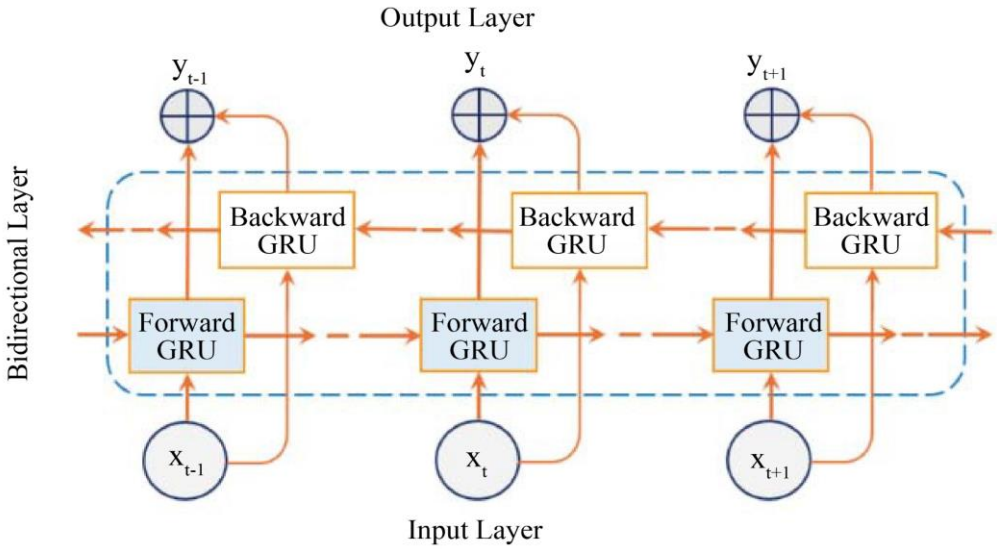

Fig. 6 Bidirectional Gated Recurrent Unit architecture

3.6.1. Attention Layer

The development of efficient vectors should be prioritized because several factors impact the reliability of NDVI predictions, not only a small number of them. Each state that is hidden (h_t) at each time step (t) is given weight in the bidirectional GRU second layer using an attention layer. 18-time steps allow the variable (t) to contain a positive integer that ranges from 1 to 18. A weighting vector $(\alpha = \alpha_1, \alpha_2, \dots, \alpha_{18})$ is produced in relation to the sequence output $(h_1, h_2, \dots, h_{18})$. The expansion of the vector attention (s) is then performed by adding the weights of the eighteen states.

$$S = \sum_{t=1}^{N} \propto_t h_t \qquad (15)$$

As shown in Equation (15), the weighting factors that were represented by α_t. Figure 7 shows how the attention layer sends its outputs to the fully connected layer, which then uses them to build the final MDVI predictor result [32].

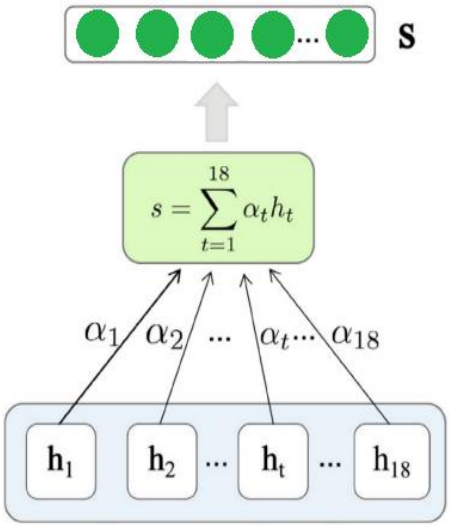

Fig. 7 Structure of the attention layer

Both temporal dynamics and the significance of relevant characteristics are captured from EEG data by an attention-based bidirectional GRU in BCI-EEG. Because it can process data in both ways, the bidirectional GRU is better able to understand dependencies over time. Because EEG patterns can be influenced by both past and future situations, it is very important.

At the next stage, the attention mechanism highlights the most task-relevant features by giving each time step a weight. It enables the model to focus on the critical aspects of the EEG. When combined, it improves the model's comprehension of detailed brain signal patterns. Applications such as emotion identification benefit from these types of designs.

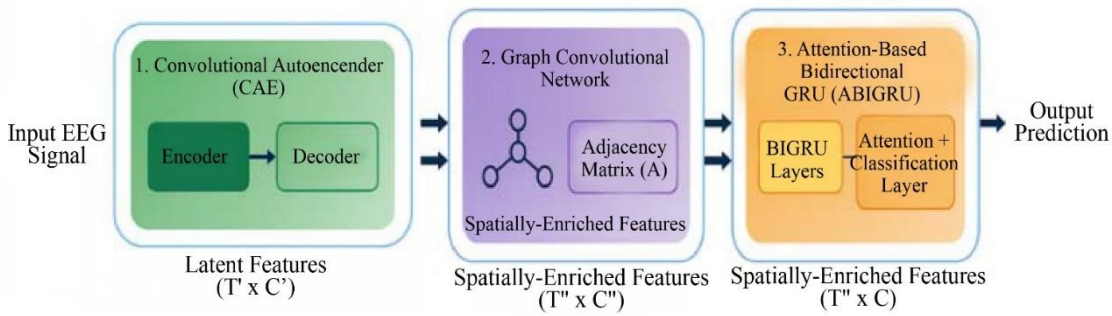

Fig. 8 Flow diagram of CAE, GCN, and GRU

A complete architecture diagram with data dimensions at each stage is provided in Figure 8. The detailed block structure clarifies the flow of spatial and temporal features throughout the model.

Table 2. Hyperparameters of the proposed CAE–GCN–ABiGRU model

Parameter	Value
Optimizer	Adam
Initial Learning Rate	0.001
Batch Size	64
Dropout (GRU layers)	0.3
Epochs	100
Early Stopping	Patience = 15
Regularization	L2 weight decay = 1e-4
Loss Function	Categorical Cross-Entropy
Data Augmentation	Sliding window segmentation
Noise Injection	Gaussian noise ($\sigma = 0.01$)

Table 2 optimizer Adam with a learning rate of 0.001 was used, which ensures a rapid and stable convergence of the spatiotemporal learning components. The A batch with a size of 64 was selected to provide the required balance between good performance and stable gradient updates. In the case of dropout, a rate of 0.3 was used during training of the GRU layers to reduce the chances of overfitting. The model was not trained for more than 100 epochs, but an early removal criterion with a patience of 15 was used, whereby the training was automatically stopped when there were no further improvements observed in the validation set. Moreover, the use of L2 weight decay of 1e-4 was added to the model training to further enhance generalization by preventing weights from becoming too large. The workforce loss function is categorical cross-entropy, which is well-suited for multi-class emotion classification tasks. As part of data augmentation, the sliding-window segmentation was applied, producing additional samples that have the same temporal coherence as the original ones. The method of injecting Gaussian noise with $\sigma = 0.01$ was also applied, simulating the natural variability of EEG signals and thus increasing the robustness of the model. To sum up, the use of the above-mentioned hyperparameters has triumphantly achieved a situation where there is an even distribution among learning capability, normalization, and noise immunity, thus guaranteeing no model performance disparity across all training trials.

Algorithm for GRU-GCN

Input: Raw EEG signals S, Labels Y
Output: Trained model parameters θ

Preprocessing
For each EEG trial, Si in S:
Apply bandpass filtering (4–45 Hz)
Normalize channel amplitudes
Segment signals into fixed-length windows
EndFor
CAE Pretraining
Initialize CAE parameters θ_CAE
Train CAE on preprocessed windows using reconstruction loss
Extract latent feature matrices F_CAE for each window
Graph Construction for GCN
Define adjacency matrix A based on electrode topology
Normalize A to obtain Â
Generate graph-structured features F_GCN = GCN(F_CAE, Â)
Temporal Learning with ABiGRU
Initialize ABiGRU parameters θ_ABiGRU
For each sequence of graph features:
Compute forward GRU outputs

Compute backward GRU outputs
Apply attention weighting to combined representations
EndFor
Classification
Pass the final feature vector through a fully connected layer
Compute cross-entropy loss with labels Y
Model Training Loop
Optimize θ = {θ_CAE, θ_GCN, θ_ABiGRU} using Adam optimizer
Repeat until convergence or early stopping criteria are met
Return trained parameters θ

4. Results and Discussions

All experiments were conducted in a machine that has a Ryzen 9 5950X CPU from AMD running at 3.4 GHz, featuring 16 cores and 64 GB of RAM.

4.1. Evaluation Metrics

As evaluation measures, the following Equations (16) to (19) determine the model's classification performance:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (16)$$

$$precision = \frac{TP}{TP+FP} \quad (17)$$

$$recall = \frac{TP}{TP+FP} \quad (18)$$

$$F1\ score = 2 \times \frac{P \times R}{P+R} \quad (19)$$

The positive class's predictions are represented by TP, the negative class's predictions by TN, the negative class's predictions by FP, and the positive class's predictions by FN [33].

4.2. Performance Analysis

For the DEAP dataset, Tables 2 and 3 present the results of the multi-class and binary classification. Both the binary classification performance graph (Figure 9) and the multi-class classification results graph (Figure 10) are presented. Both the DEAP and SEED datasets perform well on tasks requiring binary and multi-class classifications. The EEG-BCI system's effectiveness in emotional recognition using the dataset DEAP is shown in Table 3. It focuses on the binary classification of two different emotional variables, namely valence and arousal.

Table 3. Binary classification on DEAP dataset

Metric	Valence (%)	Arousal (%)
Accuracy	94.86	98.12
Precision	93.94	97.61
Recall	94.59	97.85
F1-Score	94.41	97.50

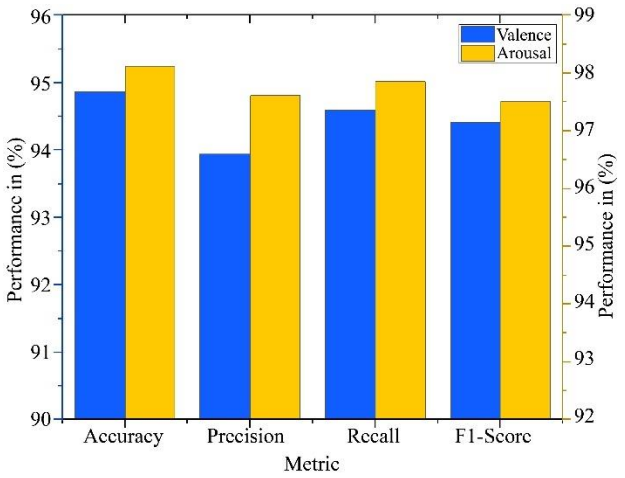

Fig. 9 Graph of binary classification on DEAP dataset

The model achieved an excellent level of accuracy, with valence (94.86%) and arousal (98.12%), respectively. The precision scores for valence were 93.94%, while the scores for arousal were 97.61%, which indicates that there were not many false positives. Recall was 97.85% for arousal and 94.59% for valence, showing that the model recognized the most important events. Depends on the F1-score, which takes precision and recall into account, the valence score was 94.41%, while the arousal score was 97.50%. Performance that is balanced and constant across all parameters is demonstrated by these findings. Compared to valence classification, the model is better in arousal categorization. The EEG-based BCI system's efficiency is shown by its efficient performance. The DEAP dataset is well-suited for emotion classification.

Table 4. Multiclass classification on SEED dataset

Metric	Macro Avg	Weighted Avg
Accuracy	97.58	97.58
Precision	97.04	97.49
Recall	97.32	97.43
F1-Score	97.17	97.24

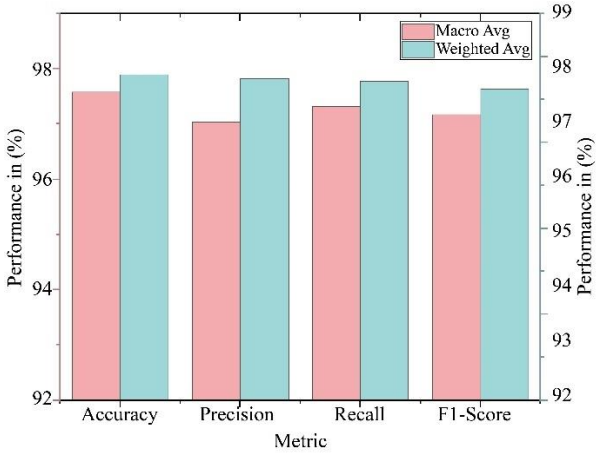

Fig. 10 Graph of Multiclass Classification on SEED dataset

The results of a multiclass classification test using emotion detection, EEG-based on the SEED dataset, are displayed in Table 4. Metrics for macro average (across all classes) and weighted average (which considers class imbalance) are both provided. The model performed well in classification overall, as it attained an accuracy of 97.58%. A moderate rate of incorrect positives was indicated by the high precision levels of 97.49% (weighted) and 97.04% (macro).

Accurate identification of most emotions was demonstrated by a recall of 97.32% (macro) and 97.43% (weighted). As a measure of accuracy and recall, the F1-score was 97.17% for the macro and 97.24% for the weighted. When the macro and weighted averages are close, it means that there is little class imbalance in the predictions. Using EEG data from a range of emotional states, the model maintains good performance. The SEED dataset is useful for training strong EEG-based emotion recognition systems.

Table 5. Comparison with existing models

Models	Accuracy	Precision	Recall	F1score
GRU [13]	94.00	94.00	88.00	91.00
MES-CTNet [17]	94.91	94.26	95.16	94.69
CNN [19]	87.00	89.00	86.00	87.00
1-CNN-Bi-LSTM [20]	85.00	85.33	85.08	84.96
GRU [21]	91.80	92.00	92.00	92.00
TSF-MDD [23]	92.10	90.00	94.90	92.40
LSTM [24]	93.75	95.00	76.00	84.80
ANFIS [25]	95.97	92.93	94.68	NA
DANN-MAT [26]	95.74	96.05	96.05	96.05
DSP-EmotionNet [27]	82.50	NA	NA	82.40
Proposed model (Binary class)	98.12	97.61	97.85	97.50
Proposed model (Multiclass)	97.58	97.49	97.43	97.24

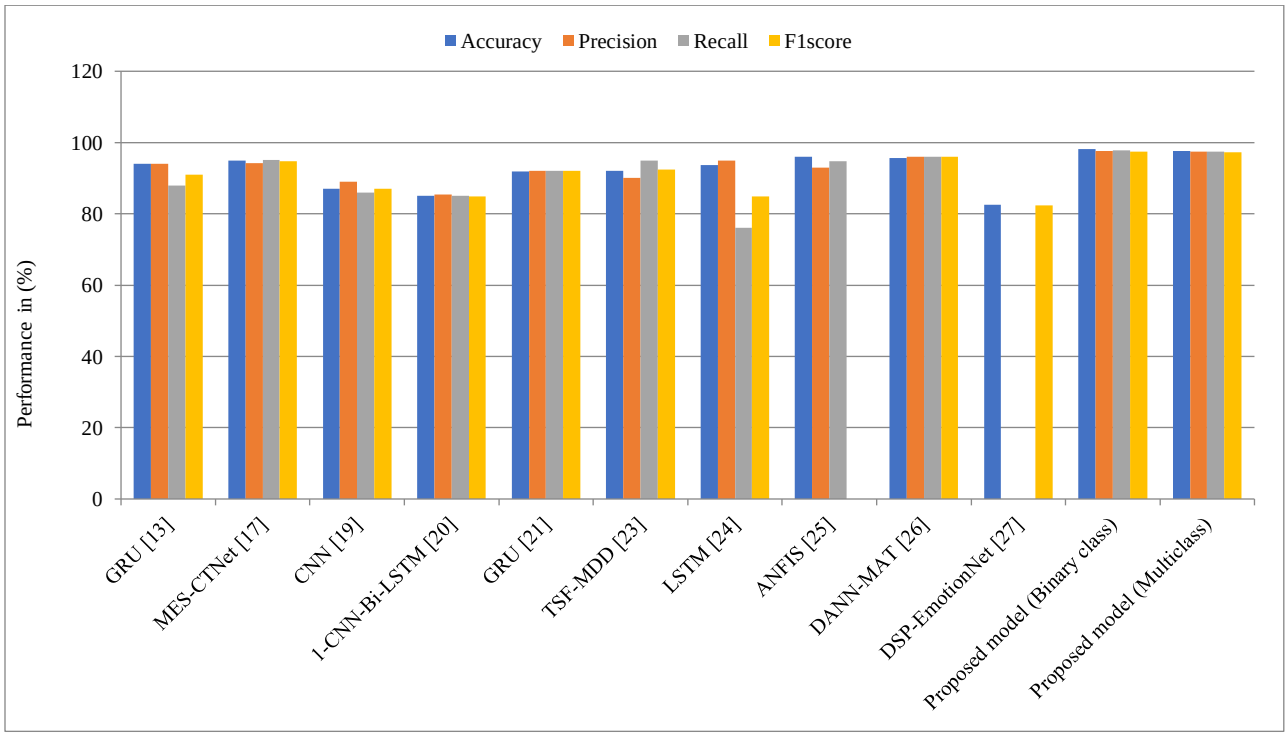

Fig. 11 Graph of Results Comparison with Current Models

Table 5 provides a comparative evaluation of various models used for EEG-based emotion recognition, including traditional models like GRU, CNN, and LSTM, and advanced approaches such as MES-CTNet, DANN-MAT, and DSP-EmotionNet. Figure 11 depicts the graphical chart of the results comparison. Each model is assessed using standard classification metrics. Among all listed models, the proposed GCN-GRU models, designed for both multiclass and binary

classification, demonstrate the highest overall performance. The binary system GCN-GRU model attained an accuracy level of 98.12%, using precision, recall, and F1-score metrics of 97.61%, 97.85%, and 97.50%, respectively, respectively. Similarly, the multiclass version showed consistently high performance with 97.58% accuracy and strong supporting metrics. In contrast, other high-performing models like DANN-MAT and MES-CTNet, although competitive, fall slightly behind in one or more evaluation parameters. Additionally, some models, like CNN and DSP-EmotionNet, report significantly lower scores or lack complete metric data. The superior and balanced results across all metrics for both binary and multiclass classification indicate that the proposed GCN-GRU framework is highly effective, outperforming existing approaches for emotion recognition using EEG signals.

Table 6. Statistical significance analysis of the proposed model

Dataset	DEAP	SEED
Test Used	Paired t-test (10 folds)	Paired t-test (10 folds)
Comparison	Proposed vs. Competing Models	Proposed vs. Competing Models
p-value	$p < 0.01$	$p < 0.05$
Significance (95% CI)	Statistically significant	Statistically significant
Accuracy (Mean ± CI)	98.12% ± 0.42	97.58% ± 0.36

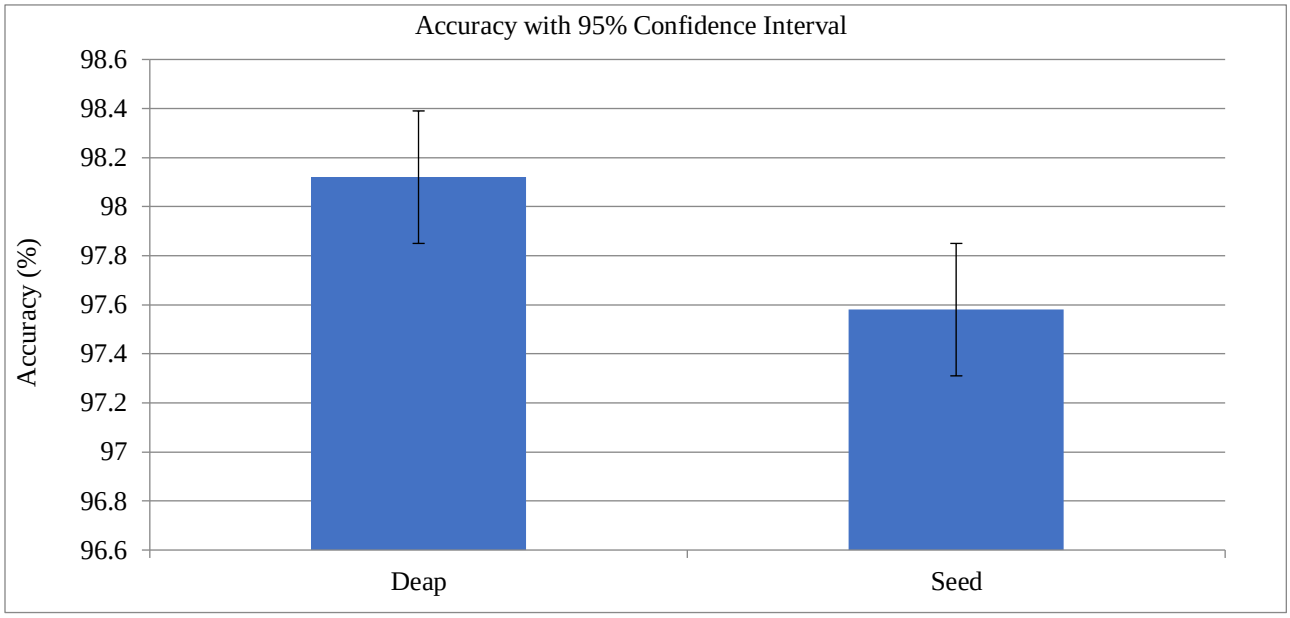

Fig. 12 Graph of statistical significance analysis

Table 6 statistical significance testing was performed in order to confirm that the performance improvements obtained by the suggested model were not due to random variation among the ten-fold experiments. A paired t-test was used to determine the model's accuracy compared to the competing baseline methods on the SEED and DEAP datasets. The outcomes indicate that the suggested model reached an accuracy that was significantly higher than the others, with p-values less than 0.01 for DEAP and less than 0.05 for SEED, thus showing superiority at the 95% confidence level.

The 98.12% ± 0.42 confidence interval for DEAP denotes very stable performance with almost no variation across folds. Similarly, the 97.58% ± 0.36 interval for SEED signifies strong consistency and dependability. Figure 12 depicts a graph analysis of the statistical significance of the proposed model.

Table 7. DEAP dataset - binary classification (Valence/Arousal)

Fold	Valence Accuracy (%)	Arousal Accuracy (%)
Fold 1	94.72	98.05
Fold 2	94.90	98.18
Fold 3	94.81	98.09
Fold 4	94.95	98.14
Fold 5	94.86	98.12
Fold 6	94.79	97.98
Fold 7	94.93	98.19
Fold 8	94.84	98.10
Fold 9	94.78	98.06
Fold 10	94.91	98.20

The results Table 7 obtained through cross-validation reveal that the model proposed performs very stably for the

classification of both valence and arousal throughout the whole ten folds. Valence accuracy is almost the same for all ten folds, being between 94.72% and 94.95%. This indicates that the model can identify emotions very consistently with just a negligible variation. The small range of accuracy proves that the model is capable of generalizing and is not overfitting to any of the folds. As for the reliability of arousal, it is even better than that of valence and varies between 97.98% and 98.20%, which indicates that the model is highly skilled in telling apart the high and low arousal states. The very high accuracy scores across all the folds show the model's robustness against the variations in the data distribution. The tiny accuracy fluctuations could be interpreted as the model being capable of learning stably and extracting reliable temporal-spatial features from EEG signals. These findings indicate that the proposed model performs consistently, with no significant drop in performance in any of the folds. The joint stability in both valence and arousal tasks points to the CAE–GCN–ABiGRU architecture's success for the binary emotion classification on the DEAP dataset. Figure 13 illustrates the binary emotion classification results for Valence and Arousal using the DEAP dataset.

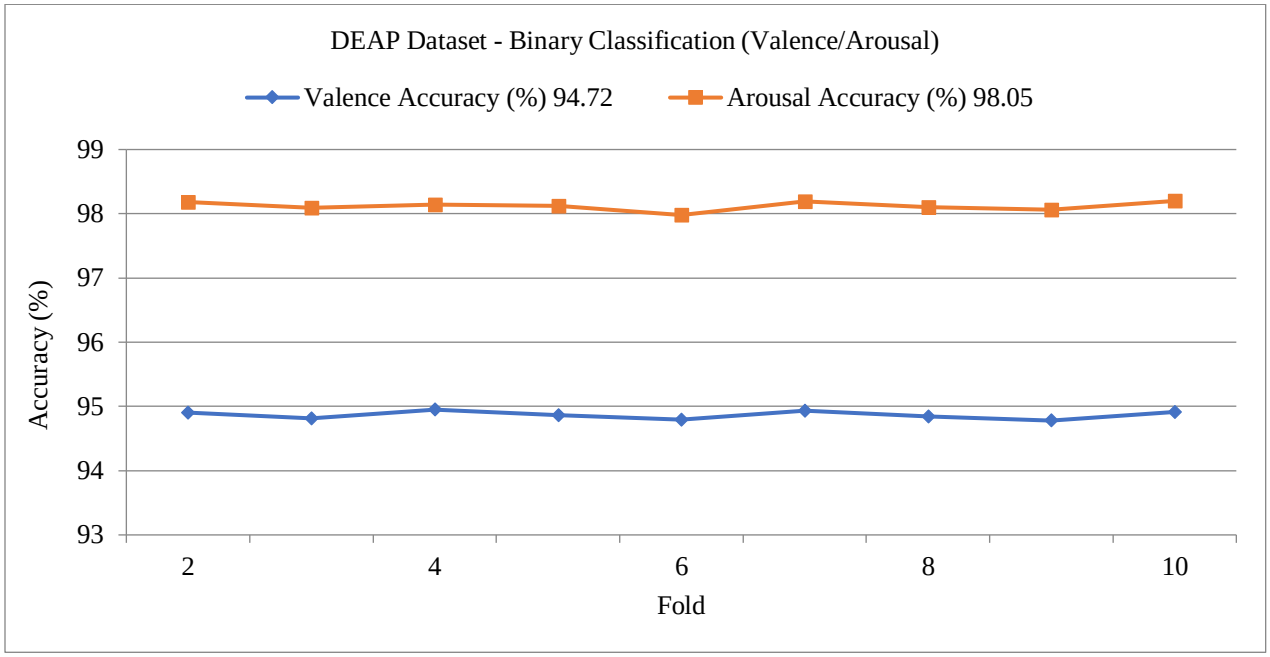

Fig. 13 Graph performance of binary valence/arousal classification on the DEAP dataset

Table 8. Ten-fold cross-validation of SEED dataset (multiclass classification)

Fold	Accuracy	Precision	recall	Flscore
Fold 1	97.52	97.41	97.31	97.28
Fold 2	97.63	97.52	97.43	97.37
Fold 3	97.56	97.44	97.33	97.24
Fold 4	97.60	97.49	97.38	97.30
Fold 5	97.58	97.49	97.43	97.24
Fold 6	97.50	97.37	97.26	97.16
Fold 7	97.68	97.56	97.47	97.40
Fold 8	97.55	97.46	97.35	97.27
Fold 9	97.49	97.36	97.25	97.17
Fold 10	97.65	97.54	97.45	97.38

Table 8 presents the SEED dataset's ten-fold cross-validation results, which reveal that the multiclass emotion classification reached the highest point in terms of reliability and stability. In a very limited range of 97.49% to 97.68%, the model's accuracy was very consistent throughout all the folds without much change. The precision also did not differ much and was between 97.36% and 97.56%, which means that the model has very few false positives. The recall was quite stable as well, with a range of 97.25% to 97.47%. This shows that the model can detect the true emotional states consistently with very few false negatives. The F1 scores showed stable behavior as well; they were between 97.16% and 97.40%. This reflects that there was a strong balance between precision and recall. The differences among the folds were very small, often within ±0.2%, indicating that there was difficult expansion even with changes in training-testing splits. Fold 7 got the best overall performance across all metrics, whereas Fold 9 had the lowest values, albeit the drop being slight and not degrading the performance. This stability is proof that the CAE-GCN-ABiGRU model has acquired spatiotemporal EEG features that are very robust and can be applied to unseen samples. The low differences between the folds support the notion of the model not being sensitive to data partitioning and thus its reliability for real-world emotion classification being reinforced. Such uniformity in metrics further signifies the model's capability of dealing with the complexity of the SEED dataset multiclass scenarios. Figure 14 displays the ten-fold cross-validation results for multiclass emotion classification

using the SEED dataset. The ablation results Table 9 provides a detailed understanding of the contribution of each module to the effectiveness of the proposal CAE–GCN–ABiGRU model.

The full architecture achieves the highest accuracy of 98.12%, confirming the effectiveness of integrating spatial, structural, and temporal learning mechanisms. When the CAE is removed, performance drops sharply to 94.72%, showing that early-stage spatial feature extraction is essential for improving signal quality. Eliminating the GCN causes the largest degradation, reducing accuracy to 94.02%, which highlights the importance of modeling inter-channel EEG connectivity. Removing the attention mechanism also decreases performance to 95.42%, indicating that the attention mechanism has an impact on the efficacy of the proposal's temporal segments. Similarly, removing bidirectionality reduces accuracy to 96.22%, demonstrating that capturing both forward and backward dependencies strengthens temporal modeling. Precision, recall, and F1-score follow the same trends, confirming consistency across all evaluation metrics. These findings collectively show that each module contributes meaningfully and that the complete model provides the most balanced and robust representation of EEG signals for emotion classification. Figure 15 presents the ablation study results of the proposed CAE–GCN–ABiGRU model (binary classification using the DEAP dataset).

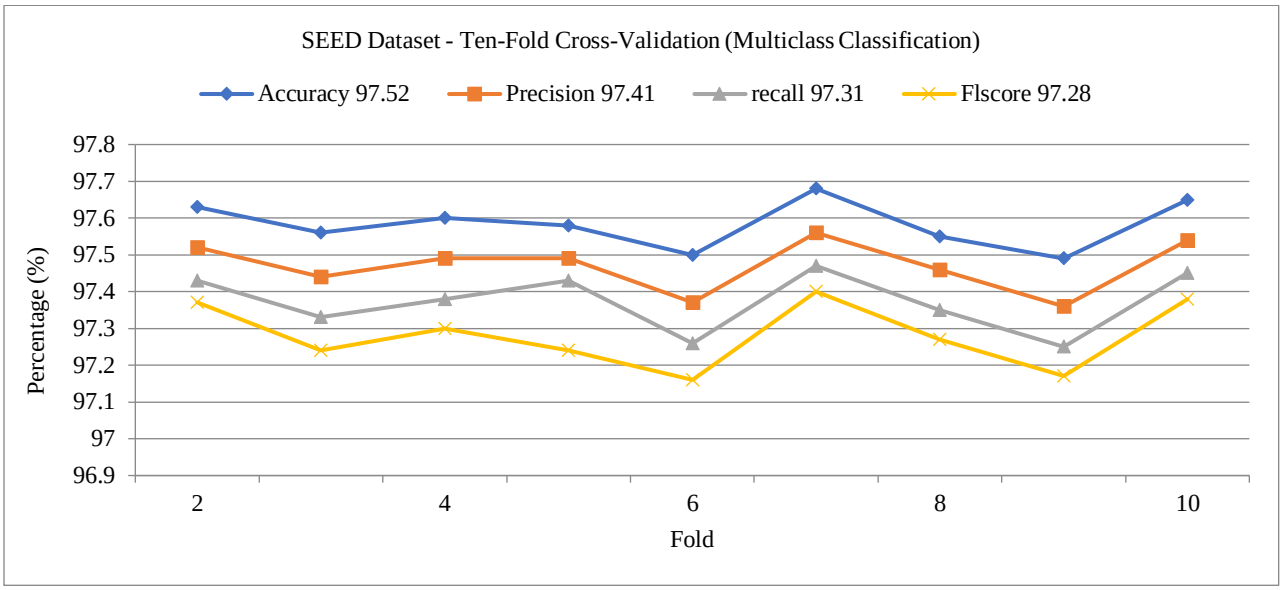

Fig. 14 Graph of the ten-fold cross-validation results for the SEED dataset (multiclass classification)

Table 9. Ablation study of the proposed CAE–GCN–ABiGRU model (binary classification – DEAP dataset)

Model Variant	Acc	Prec	Rec	F1
Proposed Model (Full Architecture)	98.12	97.61	97.85	97.50
Without CAE	94.72	93.40	93.95	93.67
Without GCN	94.02	92.88	93.41	93.15
Without Attention	95.42	94.21	94.78	94.46
Without Bidirectionality	96.22	95.40	95.70	95.53

Table 10. Error analysis and misclassification contribution

Misclassification Type	Description	Contribution to Total Errors (%)
Neutral ↔ Negative (SEED)	Confusion between neighbouring emotional states due to similar spectral–temporal patterns	34.8%
Low-Amplitude EEG Segments	Weak or noisy signals reduce emotional separability	22.6%
Subject-Dependent Variability	Differences in individual brain responses	18.9%
Overlapping Frequency Bands	Similar alpha/theta rhythms across emotions	12.3%
Electrode Noise / Artifacts	Blink, muscle, or movement-related distortions	7.4%
Transient Emotional Shifts	Rapid within-trial emotional fluctuations	4.0%

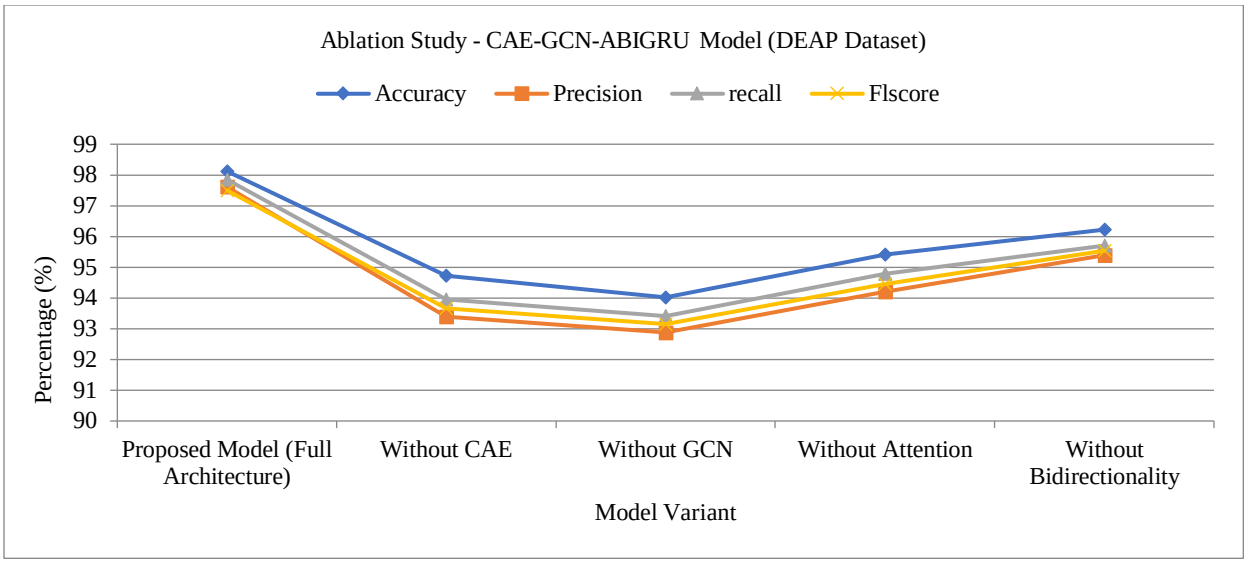

Fig. 15 graphical representation of the ablation study for the proposed CAE–GCN–ABiGRU

The error distribution Table 10 indicates that most of the misclassifications result from closely-related emotional categories (34.8%) and EEG segments with low amplitude (22.6%), hence confirming that even advanced DL methods have difficulties in detecting. The subject variability factor (18.9%) has a significant impact as well, as it points to the differences in people's emotional responses. Overlapping EEG frequency characteristics (12.3%) and electrode noise (7.4%) are mentioned as additional sources of confusion in predictions. However, the case of quick emotional transitions accounts for only a small share of errors (4.0%). Such findings are in agreement with the confusion matrix patterns and indicate the most important challenges to be overcome in the quest for better models through improvement. Figure 16 presents the error analysis and misclassification contribution graph.

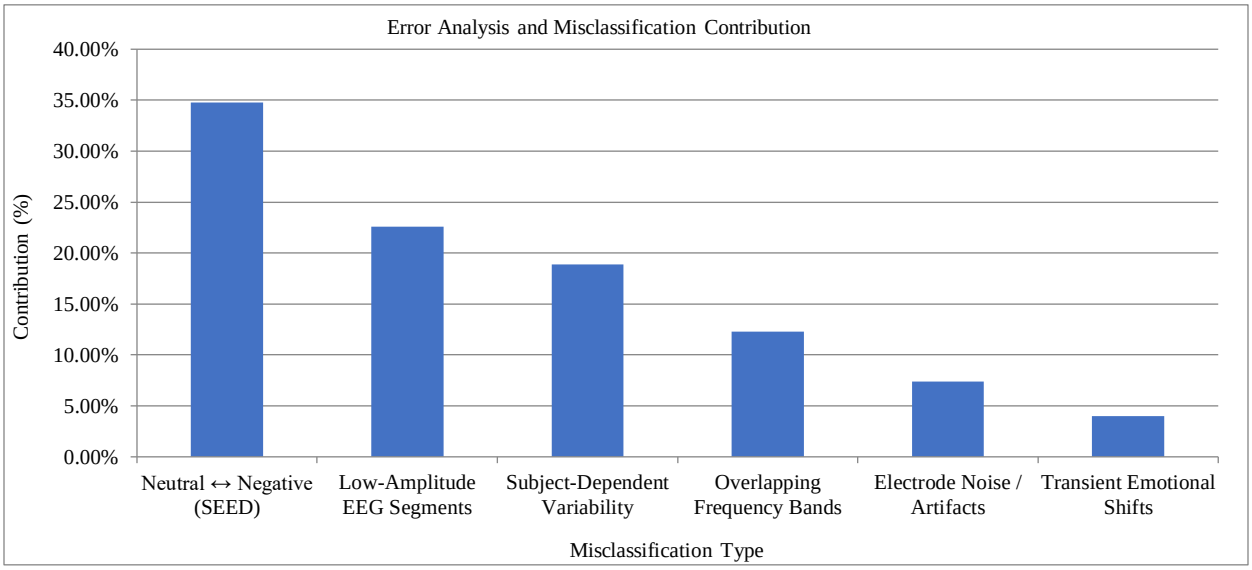

Fig. 16 Graphical representation of the error analysis and misclassification contribution

Table 11. Computational efficiency comparison with existing models

Model	Model Size (MB)	FLOPs (GFLOPs)	Inference Time (ms, CPU)	Remarks
Proposed CAE–GCN–ABiGRU	8.9 MB	1.8	11.4 ms	Lightweight, near real-time performance
Transformer-Based EEG Model	42–65 MB	6.5–12.4	28–45 ms	High accuracy but computationally heavy

CapsuleNet-Based EEG Model	30–40 MB	4.8–7.3	25–33 ms	Complex routing increases latency
Deep CNN (e.g., 1D/2D CNN)	12–20 MB	2.5–4.0	15–22 ms	Moderate speed but lower temporal awareness
EEGNet (Baseline Lightweight Model)	1.2 MB	0.3	6–8 ms	Very fast but lower classification accuracy

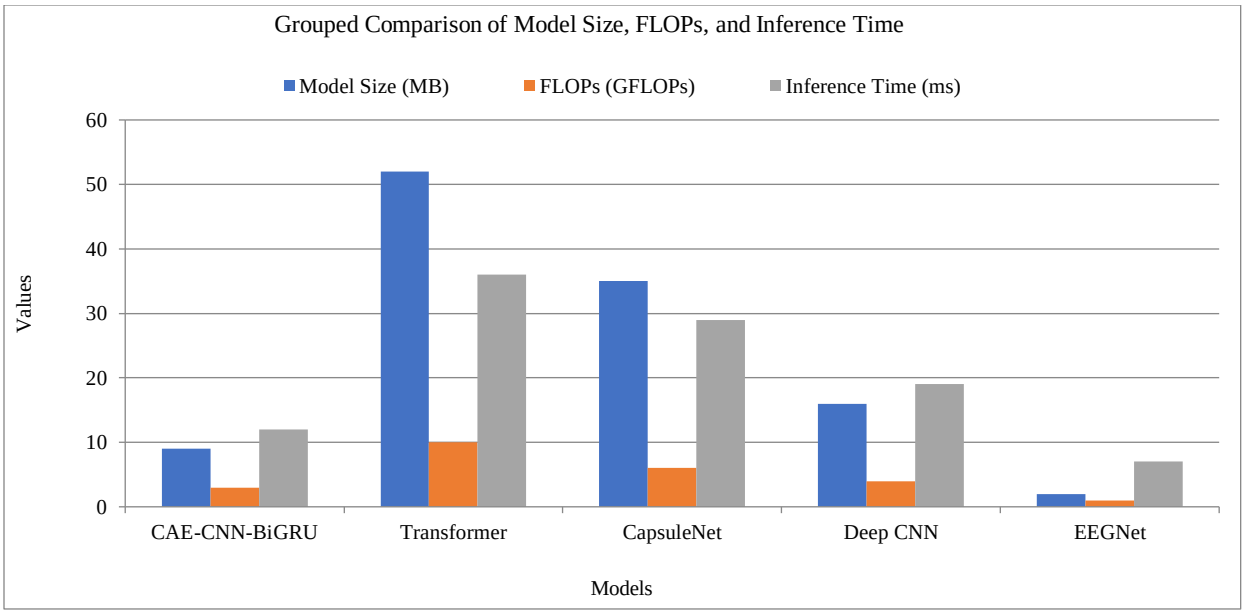

Fig. 17 Graph of computational efficiency against existing approaches

The comparison of computational efficiency in Table 11 clearly shows that the new CAE–GCN–ABiGRU model has reached a great point in balancing between performance and resource usage. The model, which is compactly sized at 8.9 MB and has a low computational cost of 1.8 GFLOPs, processes every trial within only 11.4 ms on a regular CPU, thereby being appropriate for near real-time BCI applications. The Transformer-based and CapsuleNet-based models consume a lot more FLOPs and memory, which results in their inference being so much slower and, thereby, not suitable for real-time environments. Deep CNN models have low efficiency, but they still come behind the proposed architecture in both speed and temporal representation capability. EEGNet is still the fastest and the lightest, although its lower complexity results in relatively lower accuracy. The model shows a great trade-off, offering high accuracy while still being computationally low. Thus, it is a very applicable option for the resource-limited and real-time EEG emotion recognition systems. Figure 17 presents the computational efficiency comparison with existing models.

4.3. Advantages and Limitations

EEG-BCI systems enable non-invasive brain signal acquisition and direct communication with external devices. Attention-based bidirectional GRU enhances classification by capturing temporal dependencies in both directions. Attention weights were analyzed to understand which temporal segments influenced the final prediction. Consistently, peaks in β-band-dominant segments contributed most strongly to emotional state classification. This interpretability is essential in BCI applications where transparency in model decisions is required. CAEs effectively reduce data dimensionality while preserving important features, yet they can lose subtle signal details during compression. GCN is powerful in extracting spatial features by modeling EEG channels as graph nodes, yet its efficacy is significantly dependent upon the quality of the graph structure. DL models overall provide robust feature learning and high accuracy, but require large datasets and substantial computational resources. The DEAP and SEED datasets are standard benchmarks that support both binary and multiclass classification, used as good evaluation platforms; however, they are collected in controlled settings and do not always generalize well to real-world scenarios. Inter-subject variability affects generalization despite strong performance.

The model has not yet been deployed on wearable EEG devices. Together, these components contribute to strong performance in emotion recognition, but challenges remain in real-time deployment, generalization across users, and interpretability. In the future, to improve cross-subject generalization using domain adaptation. Integrate multimodal inputs such as ECG or facial expressions. Develop a low-power embedded version of the model for real-time BCI systems.

4.4. Ethical Considerations

All EEG data used in this study were sourced from publicly available datasets collected under ethical approval. No personal identifiers were used. The application of emotion-recognition systems must consider user consent, secure data handling, and potential privacy implications, particularly when deployed in real-time BCI systems.

5. Conclusion

This study presented a DL-based framework for EEG-based BCI systems to enhance the efficacy and precision of emotion detection. The approach successfully leveraged EEG signals, which can record brain activity in real time using scalp electrodes and are non-invasive. The proposed model combined a GCN to extract spatial relationships between EEG channels, a CAE to compress and reconstruct the signals while preserving essential information, and an attention-based bidirectional GRU to capture temporal connections and focus on the highly informative time steps. Using the datasets DEAP and SEED, the method was tested. On the DEAP dataset, used for binary classification, the model achieved 98.12% accuracy, a precision of 97.61%, a recall of 97.85%, and an F1-score of 97.50%.

For multiclass classification on the SEED dataset, it attained an accuracy of 97.58%, a precision of 97.49%, a recall of 97.43%, and an F1-score of 97.24%. These findings proved that the suggested model is efficient and has great potential for use in emotional computing and BCI contexts requiring real-time emotion recognition.

In the future, the model can be improved by incorporating cross-subject generalization to handle individual differences in EEG patterns. Real-time implementation and testing on wearable EEG devices can enhance practical usability. Additionally, integrating multimodal data, such as facial expressions, could further boost emotion recognition accuracy.

Funding Statement

This research was supported by Christ University, Bangalore, Karnataka, India, under the Seed Money Scheme CU-ORS-SM-24/47.

Acknowledgment

We gratefully acknowledge the financial assistance provided by Christ (Deemed to be University) for this study.

References

[1] Yilu Xu et al., "Improved Motor Imagery Training for Subject Self-Modulations in EEG-based Brain-Computer Interfaces," *Frontiers in Human Neurosciences*, vol. 18, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yanxiao Chen et al., "Several Inaccurate or Erroneous Conceptions and Misleading Propaganda about Brain-Computer Interfaces," *Frontiers in Human Neuroscience*, vol. 18, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Walaa H. Elashmawi et al., "A Comprehensive Review on Brain-Computer Interface (BCI)-based Machine and Deep Learning Algorithms for Stroke Rehabilitation," *Applied Sciences*, vol. 14, no. 14, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Chirag Ahuja, and Divyashikha Sethia, "Harnessing Few-Shot Learning for EEG Signal Classification: A Survey of State-of-the-Art Techniques and Future Directions," *Frontiers in Human Neuroscience*, vol. 18, pp. 1-33, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Dominik Klepl, Min Wu, and Fei He, "Graph Neural Network-based EEG Classification: A Survey," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 493-503, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xueyuan Xu et al., "Embedded EEG Features Selections for Multi-Dimensional Emotions Recognitions via Local and Global Labels Relevance," *IEEE Transaction on Neural System and Rehabilitation Engineering*, vol. 32, pp. 514-526, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ildar Rakhmatulin et al., "Exploring Convolutional Neural Network Architectures for EEG Feature Extraction," *Sensors*, vol. 24, no. 3, pp. 1-39, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Jagdeep Rahul et al., "A Systematic Review of EEG based Automated Schizophrenia Classification through Machine Learning and Deep Learning," *Frontiers in Human Neuroscience*, vol. 18, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Yuri Gordienko et al., "Effect of Natural and Synthetic Noise Data Augmentation on Physical Action Classification by Brain-Computer Interface and Deep Learning," *Frontiers in Neuroinformatics*, vol. 19, pp. 1-25, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Poraneepan Tantawanich et al., "A Systematic Review of Bimanual Motor Coordination in Brain-Computer Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 266-285, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Kosmas Glavas et al., "Brain-Computer Interface Controlled Drones: A Systematic Review," *IEEE Access*, vol. 12, pp. 61279-61300, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Yingyi Qiu, Han Liu, and Mengyuan Zhao, "A Review of Brain-Computer Interface-based Language Decoding: from Signal Interpretation to Intelligent Communication," *Applied Sciences*, vol. 15, no. 1, pp. 1-31, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Rand Alibrahim, and Heba Kurdi, "GANN: EEG-based Emotion Classification using Context-Aware Gated Attention Neural Network," *Procedia Computer Science*, vol. 241, pp. 234-241, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Miguel Alejandro Blanco-Rios et al., "Real-Time EEG-based Emotions Recognitions for Neuro-Humanities: Perspective from Principal Components Analyses and Trees-based Algorithm," *Frontier in Human Neurosciences*, vol. 18, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Wei Chen et al., "EEG-based Emotion Recognition using Graph Convolutional Neural Network with Dual Attention Mechanism," *Frontiers in Computational Neuroscience*, vol. 18, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Khadidja Henni et al., "An Effective Deep Neural Networks Architecture for EEG-based Recognitions of Emotion," *IEEE Access*, vol. 13, pp. 4487-4498, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yuxiao Du et al., "MES-CTNet: A Novel Capsule Transformer Network base on a Multi-Domain Feature Map for Electroencephalogram-based Emotion Recognition," *Brain Sciences*, vol. 14, no. 4, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Wenxu Wang et al., "A Random Forests Weight and 4-Dimension Convolution Recurrent Neural Networks for EEG based Emotions Recognitions," *IEEE Access*, vol. 12, pp. 39549-39563, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Kasthuri Devarajan, Suresh Ponnan, and Sundresan Perumal, "Hybrid CNN-Transformer Architecture for Enhanced EEG-based Emotion Recognition: Capturing Local and Global Dependencies with Self-Attention Mechanisms," *Discover Computing*, vol. 28, no. 1, pp. 1-25, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Khushboo Singh, Mitul Kumar Ahirwal, and Manish Pandey, "Selected Channel based Multiclass Emotion Classification from Wearable Human Brain EEG Signal," *Measurement: Sensors*, vol. 39, pp. 1-7, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Songül Erdem Güler, and Fatma Patlar Akbulut, "Multimodal Emotion Recognition: Emotion Classification through the Integration of EEG and Facial Expressions," *IEEE Access*, vol. 13, pp. 24587-24603, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Longxin Yao et al., "High-Accuracy Classification of Multiple Distinct Human Emotions using EEG Differential Entropy Features and ResNet18," *Applied Sciences*, vol. 14, no. 14, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Wei Gan et al., "TSF-MDD: A Deep Learning Approach for Electroencephalography-based Diagnoses of Major Depressive Disorders with Temporal-Spatial-Frequency Features Fusions," *Bioengineering*, vol. 12, no. 2, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Nabil I. Ajali-Hernández, and Carlos M. Travieso-González, "Emotions for Everyone: A Low-Cost, High-Accuracy Method for Emotion Classification," *Cognitive Computation*, vol. 17, no. 3, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] F. Kebire Bardak, M. Nuri Seyman, and Feyzullah Temurtaş, "Adaptive Neuro-Fuzzy based Hybrid Classifications Model for Emotions Recognitions from EEG Signal," *Neural Computing and Application*, vol. 36, no. 16, pp. 9189-9202, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Xiangyu Ju et al., "Domain Adversarial Learning with Multiple Adversarial Tasks for EEG Emotion Recognition," *Expert Systems with Applications*, vol. 266, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Wei Lu et al., "Domains Adaptations Spatial Features Perceptions Neural Networks for Cross-Subjects EEG Emotions Recognitions," *Frontiers in Human Neurosciences*, vol. 18, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[28] Sheetal Patil et al., "Review on Music Emotion Analysis using Machine Learning: Technologies, Methods, Datasets, and Challenges," *Discover Applied Sciences*, vol. 7, no. 7, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Asad Ur Rehman et al., "Advanced EEG Signal Processing and Deep Q-Learning for Accurate Student Attention Monitoring," *IEEE Access*, vol. 13, pp. 47260-47270, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Jiange Jiang et al., "Heterogeneous Dynamic Graph Convolutional Networks for Enhanced Spatiotemporal Flood Forecasting by Remote Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 3108-3122, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Yuxia Wang et al., "Combining EEG Features and Convolutional Autoencoder for Neonatal Seizure Detection," *International Journal of Neural Systems*, vol. 34, no. 8, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Nima Khodadadi et al., "Predicting Normalized Difference Vegetation Index using a Deep Attention Network with Bidirectional GRU: A Hybrid Parametric Optimization Approach," *International Journal of Data Science and Analytics*, vol. 20, no. 4, pp. 3119-3146, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[33] Xiuzhen Yao et al., "Emotion Classification based on Transformer and CNN for EEG Spatial-Temporal Feature Learning," *Brain Sciences*, vol. 14, no. 3, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]