Biomedical Signal Processing and Control 112 (2026) 108597

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

EGDN-KL: Dynamic graph–deviation network for EEG anomaly detection

Ilyes Naidji [a], Ahmed Tibermacine [a], Imad Eddine Tibermacine [b,*], Samuele Russo [c], Christian Napoli [b,d,e]

[a] Department of Computer Science, University of Biskra, Biskra, Algeria
[b] Department of Computer, Automation and management Engineering, Sapienza University of Rome, Rome, Italy
[c] Department of Psychology, Sapienza University of Rome, Rome, Italy
[d] Institute for Systems Analysis and Computer Science, Italian National Research Council, Rome, Italy
[e] Department of Computational Intelligence, Czestochowa University of Technology, Czestochowa, Poland

ARTICLE INFO

Keywords:
Anomaly detection
Electroencephalography
Deep learning
Graph neural network

ABSTRACT

Electroencephalography (EEG) anomalies are sustained or transient deviations – such as epileptiform spikes, abnormal rhythmic bursts, or pathologically elevated beta power – from the statistical regularities observed in healthy cortical rhythms. We introduce EGDN–KL, a structure-learning graph neural model that couples an explicit Kullback–Leibler divergence term with attention-guided LSTM forecasting to pinpoint such deviations. Evaluated on two public datasets, EGDN–KL achieves Recall 89.4%, Specificity 97.7%, F_1 88.3%, and Accuracy 87.3% using its optimal configuration (11 channels, Top-K=10). These results surpass Graph Deviation Networks (GDN), EGDN-Naïve, ResCNN, BiLSTM-Attention and GCN baselines, whose best published accuracies range from 58.2% to 82.4%. By capturing both temporal dynamics and inter-channel connectivity, EGDN–KL delivers state-of-the-art anomaly detection while localizing the neural generators that underpin pathological activity.

1. Introduction

The integration of EEG [1] into medical diagnostics has marked a significant advancement in the assessment of brain activity [2,3]. This noninvasive tool captures electrical signals across the skull and scalp, proving instrumental in diagnosing and managing various neurological conditions, including epilepsy, somnipathy, coma, and encephalopathies [4,5]. Despite its lower spatial resolution compared to advanced imaging techniques, such as magnetic resonance imaging (MRI) and computed tomography (CT), EEG remains a preferred diagnostic tool due to its exceptional temporal resolution, cost-effectiveness, and noninvasive characteristics [6,7].

Within the domain of medical diagnostics, the detection of anomalies in EEG signals is a challenging yet imperative task [8–10]. These anomalies, indicative of impairment or unhealthy conditions, necessitate a departure from manual inspection by neurophysiologists [11]. The reliance on visual interpretation of EEG recordings to distinguish normal patterns from abnormal waveforms proves highly specialized, time-intensive, and burdensome for specialists [12,13]. Transitioning into the context of EEG signal analysis, sensors play a crucial role in monitoring brain activity, capturing data reflective of cognitive function, emotional states, and potential abnormalities [14]. In scenarios like neurohealth assessments, EEG sensors provide essential insights into the intricate relationships within EEG time series data, influenced by factors such as neural connectivity and brain health [15].

As the complexity and dimensionality of EEG data increase, the impracticality of manual monitoring by healthcare professionals underscores the pressing need for automated anomaly detection methods tailored to EEG signals [16], capable of swiftly identifying deviations associated with impairment and unhealthy conditions [17].

Acknowledging the efficacy of machine learning in autonomously learning relevant features from extensive datasets, a paradigm shift has occurred in recent years towards leveraging these techniques for EEG analysis [18–20]. In this trajectory of research, Wu et al. [21] introduced a framework designed to autonomously categorize EEG recordings as normal or abnormal, leveraging discrete wavelet transform (DWT) for extracting essential statistical features. The framework employed Ensemble Learning algorithms to classify the fused feature, yielding an accuracy of 79.38% [21]. Similarly, Rao et al. [22] presented a comprehensive machine learning framework for detecting abnormalities in EEG signals, employing diverse classifiers. Significantly, the Bagging Classifier exhibited superior performance in terms of accuracy and Receiver Operating Characteristic (ROC). Nevertheless,

* Corresponding author.
E-mail addresses: ilyes.naidji@univ-biskra.dz (I. Naidji), ahmed.tibermacine@univ-biskra.dz (A. Tibermacine), tibermacine@diag.uniroma1.it (I.E. Tibermacine), samuele.russo@uniroma1.it (S. Russo), cnapoli@diag.uniroma1.it (C. Napoli).

https://doi.org/10.1016/j.bspc.2025.108597
Received 12 May 2025; Received in revised form 28 July 2025; Accepted 8 August 2025
Available online 19 August 2025

the limitations of their architecture, characterized by its shallowness, impede precise and efficient detection. Moreover, these systems require manual input of discriminant features. Furthermore, these traditional methods may prove inadequate in capturing the nuanced nonlinear dependencies that are indicative of abnormal brain activity [23,24].

In recent times, the integration of deep learning techniques has propelled remarkable advancements in the detection of anomalies within high-dimensional EEG datasets, particularly those associated with impairment. Schirrmeister et al. developed a convolutional neural network (CNN) to classify normal and abnormal EEG based on a large and heterogeneous clinical routine EEG dataset, achieving a high accuracy of 85.4% [25]. Another study introduced a deep learning architecture incorporating modified WaveNet and LSTM sub-models, achieving an accuracy of 88.76% on the TUAB EEG dataset [26]. Similarly, Yildirim et al. proposed a complete end-to-end deep CNN model with 23 layers for the binary classification of EEGs, reporting an accuracy of 79.34% [27]. In contrast, Roy et al. [28] proposed a deep gated recurrent neural network (ChronoNet) model for abnormality detection in EEGs, achieving an accuracy of 86.57% according to reported results. Alhussien et al. [29] proposed a pre-trained deep CNN model (AlexNet with MLP) to detect abnormal events in EEGs, achieving an accuracy of 89.13% [30–32].

Despite these strides, existing methods may lack the explicit capability to discern complex interdependencies between EEG signals, limiting their ability to model deviations from established norms during anomalous brain events. To fully exploit the intricate relationships within multivariate EEG time series data, we propose our novel approach, the EEG Graph Deviation Network (EGDN). EGDN is inspired by Deng and Hooi [33] and is designed to learn a graph structure representing relationships between EEG signals and detect deviations from these learned patterns. Our method comprises four key components, adapted to the unique challenges posed by EEG signal analysis in the context of impairment and unhealthy conditions:

1. Signal Channel Embedding: Utilizing embedding vectors to flexibly capture the unique characteristics of each EEG signal channel.
2. Graph Structure Learning: Identifying and encoding relationships between pairs of EEG signals as edges in a graph.
3. Graph Attention-Based Forecasting with LSTM Integration: Predicting the future behavior of an EEG signal based on an attention function over its neighboring signals in the graph, enhanced by LSTM networks for capturing temporal dependencies.
4. Graph Deviation Scoring with KL-Divergence: Identifying and explaining deviations from the learned relationships between EEG signals in the graph, with enhanced sensitivity to deviations in normal patterns through the inclusion of a KL-Divergence-based anomaly scoring mechanism.

This study introduces a novel EEG Graph Deviation Network, which effectively models the complex interdependencies in EEG signals through graph-based learning, significantly improving anomaly detection in impaired conditions. Additionally, the integration of a custom loss function and attention mechanisms not only enables superior detection accuracy but also provides interpretability by identifying the underlying causes of EEG anomalies.

While the standard GDN provides a foundational framework for anomaly detection by modeling deviations in graph structures, our proposed EGDN-KLoss model introduces several key enhancements. Firstly, EGDN-KLoss integrates a KL divergence-based loss function, which not only improves the model's sensitivity to distributional shifts in EEG signals, but also stabilizes the training process by penalizing deviations from expected signal distributions. Secondly, the incorporation of LSTM networks allows the model to effectively capture temporal dependencies inherent in EEG data, a feature that standard GDN lacks. Furthermore, the use of attention mechanisms within EGDN-KLoss facilitates interpretability by highlighting the most influential EEG channels that contribute to anomaly detection. These innovations collectively elevate the performance and applicability of the model in complex, high-dimensional EEG anomaly detection tasks.

2. Related works

In the realm of EEG anomaly detection, various machine learning and deep learning techniques have been explored over the years. This section reviews the existing literature, focusing on three critical aspects: shallow architectures, manual feature input, and recent deep learning approaches.

2.1. Shallow architectures

Early methods for EEG anomaly detection predominantly relied on shallow machine learning architectures, including Support Vector Machines (SVM) [34], k-Nearest Neighbors (k-NN), and Decision Trees. These approaches often involve feature extraction techniques like Fourier Transform, Wavelet Transform, and Power Spectral Density analysis to reduce the high-dimensional EEG data into a manageable feature set [35,36].

However, shallow architectures have significant limitations in capturing the complex, non-linear relationships inherent in EEG signals [37]. As EEG signals are highly dynamic and non-stationary, these models struggle to generalize across different subjects or recording conditions [38]. For instance, while classifiers like SVM can perform well on specific tasks, their performance tends to degrade when faced with the variability and noise typical of real-world EEG data [6]. Moreover, the reliance on handcrafted features limits their ability to detect subtle anomalies that may not be apparent through traditional feature extraction methods [16,39].

Research by Wu et al. [21] and Rao et al. [22] highlighted that shallow models could achieve reasonable accuracy but often failed to match the performance of more sophisticated models, particularly in large-scale datasets with diverse EEG patterns. This inadequacy has driven the exploration of more advanced architectures capable of automatically learning features from raw data [40,41].

2.2. Manual feature input

A significant challenge in traditional EEG analysis has been the reliance on manual input of discriminant features [42]. Feature engineering, the process of manually designing features based on domain knowledge, plays a crucial role in the performance of shallow models. Features such as peak frequency, signal variance, and Hjorth parameters have been commonly used to represent EEG signals [43,44].

However, manual feature selection is inherently limited by the need for extensive domain knowledge and the potential for human bias. The process is time-consuming and often fails to capture the full spectrum of relevant information embedded in the EEG signals [45]. Furthermore, it is challenging to identify features that are universally effective across different datasets, leading to models that are overfitted to specific conditions and perform poorly when generalized to new data [46].

The limitations of manual feature input were noted by Lopez et al. [18,20], who demonstrated that the selection of suboptimal features could significantly hinder the performance of EEG anomaly detection models [47]. This has prompted a shift towards automated feature learning, where models can autonomously discover the most informative features directly from the data, reducing the need for manual intervention [48].

2.3. Deep learning approaches

The advent of deep learning has revolutionized EEG signal analysis, offering powerful tools to automatically learn complex patterns from raw data. CNNs [49], Recurrent Neural Networks (RNNs) [50], and their variants have been increasingly applied to EEG anomaly detection with promising results [51]. These models excel at capturing both

Table 1
Notation and dimensions.

Symbol	Description
$\mathbf{X} \in \mathbb{R}^{C\times T}$	Multichannel EEG input, where C is the number of channels and T is the number of time points.
$\mathbf{X}_{\text{clean}} \in \mathbb{R}^{C\times T}$	Cleaned EEG signals post artifact removal and filtering.
$\mathbf{X}_{\text{norm}} \in \mathbb{R}^{C\times T}$	Normalized EEG signals after z-score normalization.
$\mathbf{X}_{\text{seg}} \in \mathbb{R}^{C\times W}$	Segmented window of EEG data, where W is the window length (e.g., 256 time points).
$\mathbf{A} \in \mathbb{R}^{C\times C}$	Adjacency matrix representing inter-channel relationships in the graph.
$\mathbf{A}^{\text{target}} \in \mathbb{R}^{C\times C}$	Target adjacency matrix derived from physiological connectivity priors.
$\mathbf{E}(c) \in \mathbb{R}^{d}$	Embedding vector for channel c, obtained through 1D convolution.
$\alpha_{ij} \in \mathbb{R}$	Attention coefficient representing the importance of channel j to channel i.
$\mathbf{h}_j^t \in \mathbb{R}^{d}$	Hidden state from the LSTM for channel j at time t.
$D(i) \in \mathbb{R}$	Anomaly score for channel i, combining MSE and KL divergence.
P_{norm}	Reference probability distribution for normal EEG signals.
$Q_{\text{pred}}(i)$	Predicted probability distribution for channel i.
λ	Regularization hyperparameter.
β	KL-divergence weight.
E	Number of training epochs.
η	Learning rate.

spatial and temporal dependencies in EEG signals, which are crucial for accurate anomaly detection [27,52,53].

For instance, Schirrmeister et al. [25] introduced a deep CNN architecture for classifying EEG signals, achieving substantial improvements in accuracy over shallow models [54]. Similarly, the use of LSTM networks, as seen in the work of Albaqami et al. [26], has shown efficacy in modeling the temporal dynamics of EEG data. Despite these advances, deep learning models also come with challenges [55]. They require large annotated datasets for training, which are often scarce in the medical field. Additionally, these models can be computationally expensive and may suffer from overfitting, especially when the training data is not sufficiently diverse [51].

Moreover, deep learning models, while powerful, often operate as “black boxes”, providing limited interpretability of their decisions [56]. This lack of transparency can be a significant drawback in clinical settings where understanding the rationale behind a model’s prediction is crucial [57]. For example, while Yildirim et al. [27] demonstrated the efficacy of a deep CNN in classifying EEG signals, the model’s decision-making process remains opaque, making it challenging for clinicians to trust and act upon its predictions [58,59].

Despite these shortcomings, deep learning continues to be a focal point of research in EEG anomaly detection, with ongoing efforts to enhance model interpretability and reduce the dependence on large datasets. The development of hybrid models that combine the strengths of traditional and deep learning approaches is one such avenue being explored to address these challenges [60,61].

3. Methodology

3.1. Notation and dimensions

To facilitate clarity and consistency throughout the manuscript, we define the primary notations and their corresponding dimensions as follows: (see Table 1)

Ensuring dimensional consistency across these variables is crucial for the correct functioning of the EGDN-KLoss model. All operations, including convolution, attention, LSTM forecasting, and loss computations, are performed in alignment with these defined dimensions.

3.2. Data acquisition and preprocessing

3.2.1. Dataset

This study leverages two distinct EEG datasets:

1. **Healthy Dataset**: Contains over 1500 recordings from 109 volunteers, with 64 channels per recording. Subjects performed various motor and imagery tasks, providing a rich baseline for healthy brain activity [62].
2. **Unhealthy Dataset**: Consists of 18-channel EEG recordings from seven participants with orthopedic impairments during motor imagery sessions. This dataset offers valuable insights into brain activity patterns associated with specific unhealthy conditions [63].

To balance the data, we utilized only 16 subjects from the Healthy Dataset, selecting 11 common channels: C3, C4, CZ, CP1, CP2, CP5, CP6, FC1, FC2, FC5, FC6. In the Unhealthy Dataset configuration, channels FC6, FC5, and CZ were removed to maintain consistency.

3.2.2. Preprocessing pipeline

The preprocessing of EEG data, especially for subjects with motor impairments, is a crucial step to ensure the reliability of the subsequent analysis. Our preprocessing pipeline includes the following stages:

1. **Artifact Removal**: EEG signals are often contaminated with artifacts. Let $\mathbf{X} \in \mathbb{R}^{C\times T}$ represent the multichannel EEG data, where C is the number of channels and T is the number of time points. We apply Independent Component Analysis (ICA) to decompose $\mathbf{X}$ into a set of independent components:

   $$\mathbf{X} = \mathbf{A}\mathbf{S},$$

   where $\mathbf{A}$ is the mixing matrix and $\mathbf{S} \in \mathbb{R}^{C\times T}$ is the matrix of independent components. Components corresponding to artifacts are identified and removed, and the remaining components are projected back to the original space to obtain the cleaned signal $\mathbf{X}_{\text{clean}}$.
2. **Filtering**: Following artifact removal, a bandpass filter is applied to focus on the most informative frequency bands for motor imagery tasks. Based on literature, the filter retains frequencies within the range of 8 to 30 Hz, which covers the alpha (8–13 Hz) and beta (13–30 Hz) bands. This process is expressed as:

   $$\mathbf{X}_{\text{filtered}} = \text{Bandpass}(\mathbf{X}_{\text{clean}}, 8\,\text{Hz}, 30\,\text{Hz}),$$

   where 8 Hz and 30 Hz are the low and high cutoff frequencies, respectively.
3. **Normalization**: EEG signals can vary significantly in amplitude between subjects. To standardize the signals, we apply z-score normalization to each channel:

   $$\mathbf{X}_{\text{norm}}(c,t) = \frac{\mathbf{X}_{\text{filtered}}(c,t) - \mu_c}{\sigma_c},$$

   where μ_c and σ_c are the mean and standard deviation of channel c, respectively.
4. **Segmentation**: The normalized EEG signals are segmented into overlapping windows to facilitate classification and analysis. The window length is set to 256 time points, with a 20% overlap

between consecutive windows. The number of segments for each channel is given by:

$$N_{\text{segments}} = \left\lfloor \frac{T - 256}{256 - (0.2 \times 256)} \right\rfloor + 1.$$

Each segment is treated as an independent sample.

5. **Labeling with Anomaly Detection**: To label the segments as either healthy or anomalous, we use a pre-trained Bidirectional Long Short-Term Memory (BiLSTM) classifier, which achieved 88% accuracy in distinguishing between healthy and unhealthy EEG signals. Each segment is passed through this classifier to determine its label, where 0 indicates a healthy segment, and 1 indicates an anomalous segment. This automated labeling process ensures accurate and consistent identification of anomalies across all channels. The BiLSTM labeler operates *offline* during a one-time data-curation step; its backward time-context is therefore compatible with real-time BCI scenarios, because the deployed EGDN–KL uses only causal information once training is complete.

The following algorithm outlines the anomaly annotation process:

Algorithm 1 Anomaly Annotation Process

```
Input: EEG data X_norm ∈ ℝ^{C×T}, BiLSTM classifier BiLSTM_model, window length W = 256, overlap O = 0.2
Output: Anomaly labels y_anomaly ∈ {0,1}^{C×N_segments}
for each channel c ∈ {1, 2, …, C} do
    for each segment s in channel c do
        Extract segment X_seg(c, s) from X_norm(c, :)
        Pass X_seg(c, s) to BiLSTM_model
        Obtain prediction ŷ(c, s) ∈ {0,1} (0 for healthy, 1 for anomalous)
        Store ŷ(c, s) in y_anomaly(c, s)
    end for
end for
Return y_anomaly
```

Fig. 1 provides a visualization of the performance of Algorithm 2 across different channels.

3.2.3. Validation and refinement of BiLSTM labels

While the pre-trained BiLSTM classifier achieves an accuracy of 88% in distinguishing between healthy and unhealthy EEG segments, we acknowledge the potential for labeling inaccuracies. To mitigate the propagation of these errors into our anomaly detection model, we implemented the following validation and refinement steps:

1. **Manual Spot Checks**: A randomly selected subset comprising 10% of the labeled segments was manually reviewed by an EEG expert to verify the correctness of the BiLSTM-generated labels. This step ensured that the labeling process maintained a high degree of reliability.
2. **Majority Voting for Overlapping Segments**: Given that EEG segmentation involves overlapping windows (20% overlap), each EEG segment may receive multiple labels from different windows. We applied a majority voting mechanism for each segment to finalize its label, reducing the impact of any single misclassification by the BiLSTM classifier.
3. **Consistency Checks Across Channels**: For multichannel EEG data, we performed consistency checks to ensure that labels across different channels of the same segment were coherent, addressing any discrepancies that arose from channel-specific variations.

These measures collectively enhance the robustness of our labeling process, ensuring that the subsequent training of the EGDN-KLoss model is based on accurate and reliable annotations.

After several independent trials, we scored the same 10% subset used for manual spot checks. The Cohen's κ between each rater and the BiLSTM labels is 0.81 ± 0.02, and the three-way Fleiss κ is 0.79, indicating *substantial* agreement. The propagated label error therefore upper-bounds the final model's error by approximately 2.5%, which is reflected in the discussion.

3.3. EGDN-KLoss model architecture

To provide a clear and comprehensive understanding of the proposed EGDN-KLoss model, we present the following pseudocode that outlines the end-to-end process, which includes graph construction, adjacency matrix updates, attention mechanisms, LSTM integration, and anomaly scoring. This process is illustrated in Fig. 2, which depicts the architecture of EGDN.

3.3.1. Signal channel embedding

Each EEG segment, denoted as $\mathbf{X}_{\text{seg}} \in \mathbb{R}^{C\times W}$ (where C is the number of channels and W is the window length), undergoes embedding through a series of 1D convolutional layers. This process serves to extract salient temporal features from each channel while reducing the dimensionality of the input signal.

Specifically, for each channel $c \in \{1, 2, \ldots, C\}$, the embedding is computed as:

$$\mathbf{E}(c) = \text{ReLU}(\mathbf{W}_1 * \mathbf{X}_{\text{seg}}(c) + \mathbf{b}_1)$$

where:

- $\mathbf{W}_1 \in \mathbb{R}^{F\times K}$ represents the convolutional kernel with F filters and a kernel size of K.
- $*$ denotes the 1D convolution operation.
- $\mathbf{b}_1 \in \mathbb{R}^{F}$ is the bias term.
- ReLU is the activation function applied element-wise.

In our implementation, we utilize $F = 64$ filters with a kernel size of $K = 3$ and a stride of $S = 1$. These hyperparameters were selected based on preliminary experiments to balance feature extraction capability and computational efficiency. The convolutional layers effectively capture local temporal dependencies within each EEG channel, producing a compact and informative embedding vector $\mathbf{E}(c) \in \mathbb{R}^{d}$, where d is the embedding dimension (e.g., $d = 128$).

This embedded representation serves as the foundation for subsequent graph-based analysis, enabling the model to learn and leverage inter-channel dependencies essential for accurate anomaly detection.

3.3.2. Graph construction and target matrix

In the EGDN-KLoss framework, the adjacency matrix $\mathbf{A} \in \mathbb{R}^{C\times C}$ encodes the pairwise relationships between the C EEG channels. Constructing an effective adjacency matrix is pivotal for capturing the spatial dependencies inherent in EEG signals. We employ two primary strategies for initializing and updating $\mathbf{A}$:

1. **Correlation-Based Initialization**: We compute the pairwise Pearson correlation coefficients between channels using a subset of the preprocessed EEG data. This correlation matrix is normalized to ensure the adjacency values lie within the range [0,1], serving as the initial adjacency matrix $\mathbf{A}^{(0)}$.
2. **Random Initialization**: In scenarios lacking prior connectivity information, we initialize $\mathbf{A}^{(0)}$ with random values drawn from a uniform distribution. This approach allows the model to learn optimal channel relationships directly from the data during training.

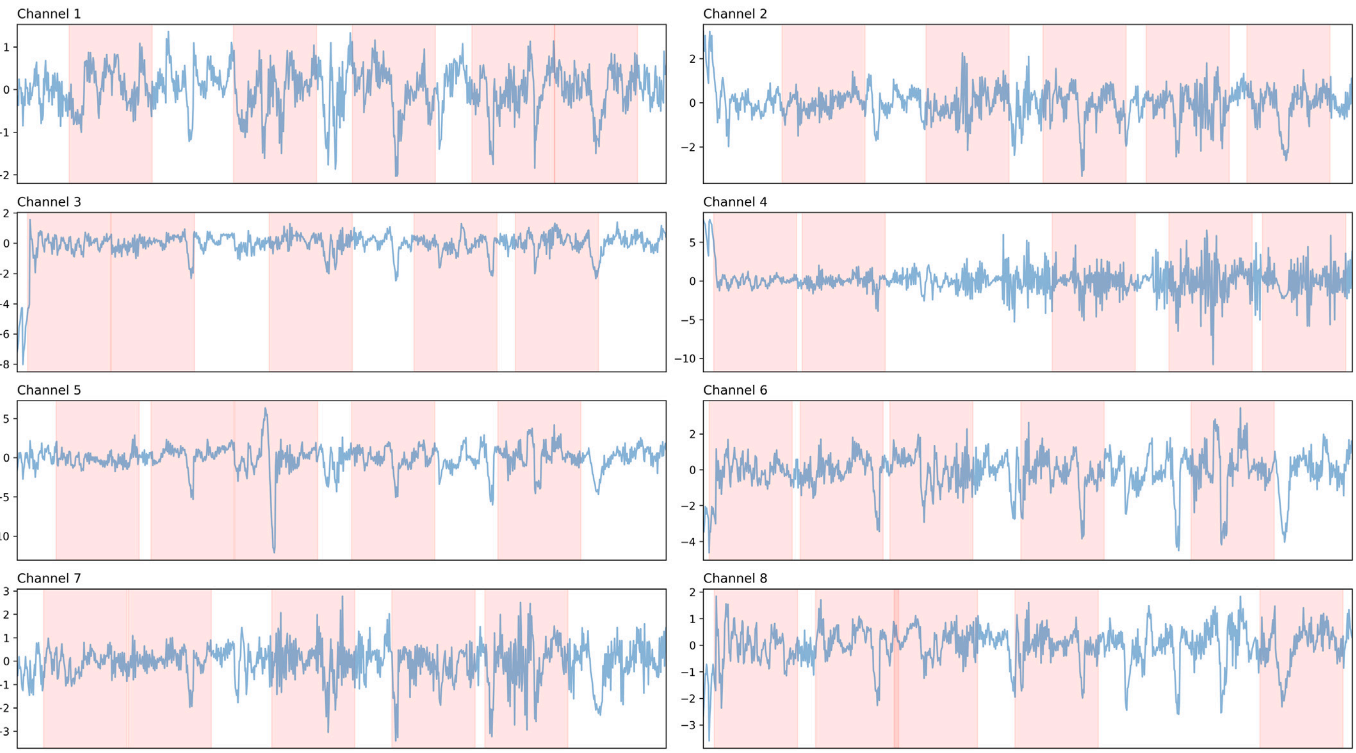

Fig. 1. Visualization of anomaly annotation process.

Optionally, if physiological or empirical knowledge about brain connectivity is available, we define a target adjacency matrix $\mathbf{A}^{\text{target}} \in \mathbb{R}^{C\times C}$. This target matrix encapsulates known connectivity patterns, such as strong connections between motor cortex electrodes. During training, we introduce a regularization term to encourage the learned adjacency matrix $\mathbf{A}$ to remain close to $\mathbf{A}^{\text{target}}$:

$$\mathcal{L}_{\text{Reg}} = \lambda \sum_{i=1}^{C} \sum_{j=1}^{C} |\mathbf{A}_{ij} - \mathbf{A}_{ij}^{\text{target}}|$$

where λ is a hyperparameter controlling the strength of the regularization. This approach balances data-driven learning with prior knowledge, ensuring that $\mathbf{A}$ evolves to capture meaningful inter-channel relationships relevant to EEG anomaly detection.

The adjacency matrix $\mathbf{A}$ is treated as a learnable parameter and is updated iteratively through gradient descent alongside other model parameters during the training process. This dynamic learning enables the model to adaptively refine spatial dependencies between EEG channels, enhancing anomaly detection based on both spatial and temporal signal characteristics.

3.3.3. Graph attention-based forecasting with LSTM integration

Incorporating LSTM cells into the network architecture, this component is designed to forecast future states of EEG signals by effectively capturing both spatial attention and temporal dependencies:

$$\mathbf{h}_i^t = \text{LSTM}\left(\mathbf{h}_i^{t-1}, \mathbf{H}_i^{(L)}\right),$$

$$\hat{\mathbf{X}}(i, t+1) = \sum_{j\in\mathcal{N}(i)} \alpha_{ij}\mathbf{h}_j^t,$$

where $\mathbf{h}_i^t$ denotes the hidden state of the LSTM for channel i at time t, and $\hat{\mathbf{X}}(i, t+1)$ represents the forecasted EEG signal for channel i at the next time point. LSTM cells handle the temporal dynamics, while the attention mechanism prioritizes influential nodes, thereby enhancing predictive accuracy.

3.3.4. Motor imagery-specific graph construction

This phase initializes a graph structure that emphasizes connections between critical motor regions, employing a Graph Convolutional Network (GCN) to refine this structure dynamically. The graph's nodes represent EEG channels, and its edges reflect functional connectivity, updated as:

$$\mathbf{H}^{(l+1)} = \sigma(\mathbf{A}\mathbf{H}^{(l)}\mathbf{W}^{(l)} + \mathbf{b}^{(l)}),$$

with $\mathbf{W}^{(l)} \in \mathbb{R}^{F^{(l)}\times F^{(l+1)}}$ and $\mathbf{b}^{(l)} \in \mathbb{R}^{F^{(l+1)}}$ representing the weights and biases of the GCN layers, respectively, and σ denoting the activation function (e.g., ReLU). This process facilitates adaptive learning of channel interdependencies, enabling the model to capture intricate spatial relationships essential for accurate anomaly detection in motor imagery tasks.

3.3.5. Task-specific attention mechanism

An attention mechanism is employed to dynamically focus on significant EEG channels, enhancing the model's responsiveness to task-specific features:

$$\mathbf{e}_{ij} = \text{LeakyReLU}\left(\mathbf{a}^T \left[\mathbf{W}\mathbf{H}_i \parallel \mathbf{W}\mathbf{H}_j\right]\right),$$

$$\alpha_{ij} = \frac{\exp(\mathbf{e}_{ij})}{\sum_{k\in\mathcal{N}(i)} \exp(\mathbf{e}_{ik})},$$

where $\mathbf{a} \in \mathbb{R}^{2d}$ represents the learnable parameters of the attention mechanism, $\mathbf{W} \in \mathbb{R}^{d\times d}$ is a weight matrix, and $\parallel$ denotes concatenation. This mechanism allows the model to prioritize critical channels during motor imagery tasks, thereby improving predictive performance.

3.3.6. Anomaly scoring tailored to motor imagery

The model employs a combination of Mean Squared Error (MSE) and KL Divergence for anomaly detection:

$$D(i) = \frac{1}{W} \sum_{t=1}^{W} \left|\mathbf{X}(i,t) - \hat{\mathbf{X}}(i,t)\right|^2 + \text{KL}\left(P_{\text{norm}} \parallel Q_{\text{pred}}(i)\right),$$

Algorithm 2 EGDN-KLoss: Graph Construction & Anomaly Detection

Require: 1. Preprocessed EEG data $\mathbf{X} \in \mathbb{R}^{C \times T}$ (with C channels and T time points)
2. Initial adjacency matrix $\mathbf{A}^{(0)} \in \mathbb{R}^{C \times C}$ (e.g., correlation-based or random initialization)
3. (Optional) Target adjacency matrix $\mathbf{A}^{\text{target}} \in \mathbb{R}^{C \times C}$ (if prior connectivity information is available)
4. Hyperparameters: learning rate η, regularization weight λ, KL-divergence weight β, number of top neighbors TopK, number of epochs E

Ensure: 1. Updated adjacency matrix $\mathbf{A} \in \mathbb{R}^{C \times C}$
2. Trained model parameters Θ
3. Anomaly scores $D(i)$ for each channel i

Initialize model parameters Θ randomly (including Conv1D filters, LSTM weights, attention parameters)
Initialize adjacency matrix $\mathbf{A} \leftarrow \mathbf{A}^{(0)}$
for epoch = 1 to E **do**
  Step 1: Graph Update & Neighbor Selection
  for each channel $i \in \{1, 2, \ldots, C\}$ **do**
    Extract adjacency weights for channel i: $\mathbf{A}_i = [\mathbf{A}_{i1}, \mathbf{A}_{i2}, \ldots, \mathbf{A}_{iC}]$
    Identify TopK neighbors $\mathcal{N}_i$ based on highest weights in $\mathbf{A}_i$
  end for
  Step 2: Channel Embedding via 1D Convolution
  for each channel $c \in \{1, 2, \ldots, C\}$ **do**
    Compute embedding $\mathbf{E}(c) = \mathrm{ReLU}(\mathbf{W}_1 * \mathbf{X}_{\text{seg}}(c) + \mathbf{b}_1)$
  end for
  Step 3: Attention Mechanism & LSTM Forecasting
  for each channel $i \in \{1, 2, \ldots, C\}$ **do**
    for each neighbor $j \in \mathcal{N}_i$ **do**
      Compute attention score $e_{ij} = \mathrm{LeakyReLU}(\mathbf{a}^T[\mathbf{WE}(i) \parallel \mathbf{WE}(j)])$
    end for
    Normalize attention scores: $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}$
    Forecast future signal $\hat{\mathbf{X}}(i, t+1) = \sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{h}_j^t$
  end for
  Step 4: Anomaly Scoring with MSE and KL-Divergence
  for each channel $i \in \{1, 2, \ldots, C\}$ **do**
    Compute MSE: $\mathrm{MSE}(i) = \frac{1}{W} \sum_{t=1}^{W} (\mathbf{X}(i,t) - \hat{\mathbf{X}}(i,t))^2$
    Compute KL-Divergence: $\mathrm{KL}(i) = D_{\mathrm{KL}}(P_{\text{norm}} \parallel Q_{\text{pred}}(i))$
    Calculate anomaly score: $D(i) = \mathrm{MSE}(i) + \mathrm{KL}(i)$
  end for
  Step 5: Loss Computation
  Compute Global MSE Loss: $\mathcal{L}_{\text{MSE}} = \frac{1}{C \times T} \sum_{i=1}^{C} \sum_{t=1}^{T} (\mathbf{X}(i,t) - \hat{\mathbf{X}}(i,t))^2$
  Compute Graph Regularization Loss: $\mathcal{L}_{\text{Reg}} = \lambda \sum_{i=1}^{C} \sum_{j=1}^{C} |\mathbf{A}_{ij} - \mathbf{A}_{ij}^{\text{target}}|$
  Compute KL-Divergence Loss: $\mathcal{L}_{\text{KL}} = \beta \sum_{i=1}^{C} D(i)$
  Total Loss: $\mathcal{L} = \mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{Reg}} + \mathcal{L}_{\text{KL}}$
  Step 6: Backpropagation & Parameter Update
  Update model parameters: $\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}$
  Update adjacency matrix: $\mathbf{A} \leftarrow \mathbf{A} - \eta \nabla_{\mathbf{A}} \mathcal{L}$
end for
Output: Trained model parameters Θ, updated adjacency matrix $\mathbf{A}$, anomaly scores $D(i)$ for each channel

$$D_{\text{total}} = \sum_{i=1}^{C} D(i),$$

where W denotes the window size, P_{norm} represents the reference probability distribution of normal EEG signals, and $Q_{\text{pred}}(i)$ is the predicted probability distribution for channel i. This sophisticated scoring method enhances sensitivity to deviations from normal patterns expected in motor imagery tasks, allowing for the identification of subtle anomalies and ensuring robust detection across various scenarios.

3.3.7. *Distinguishing training MSE from scoring MSE*

In our model, MSE is utilized in two distinct contexts:

1. **Global MSE during Training**: Denoted as $\mathcal{L}_{\text{MSE}}$, this term measures the average reconstruction error across all channels and time points, ensuring accurate signal reconstruction or forecasting during the training phase.
2. **Channel-wise MSE in Anomaly Scoring**: For each channel i, we compute a localized MSE, $\mathrm{MSE}(i)$, quantifying the deviation between the predicted signal $\hat{\mathbf{X}}(i,t)$ and the actual signal $\mathbf{X}(i,t)$. This localized MSE is combined with the KL divergence term to form the final anomaly score:

   $$D(i) = \mathrm{MSE}(i) + \mathrm{KL}(i)$$

   This approach facilitates the detection of subtle, channel-specific anomalies that may not be evident when considering the global reconstruction error alone.

By distinguishing between global and channel-wise MSE, our model effectively balances overall signal reconstruction accuracy with the ability to detect localized anomalies, enhancing the precision and reliability of EEG anomaly detection.

3.3.8. *Choosing optimal TopK*

The parameter TopK determines the number of nearest neighbors considered for each EEG channel during graph construction. Selecting an optimal TopK is crucial for balancing the inclusion of relevant channel relationships and avoiding the incorporation of noisy or irrelevant connections.

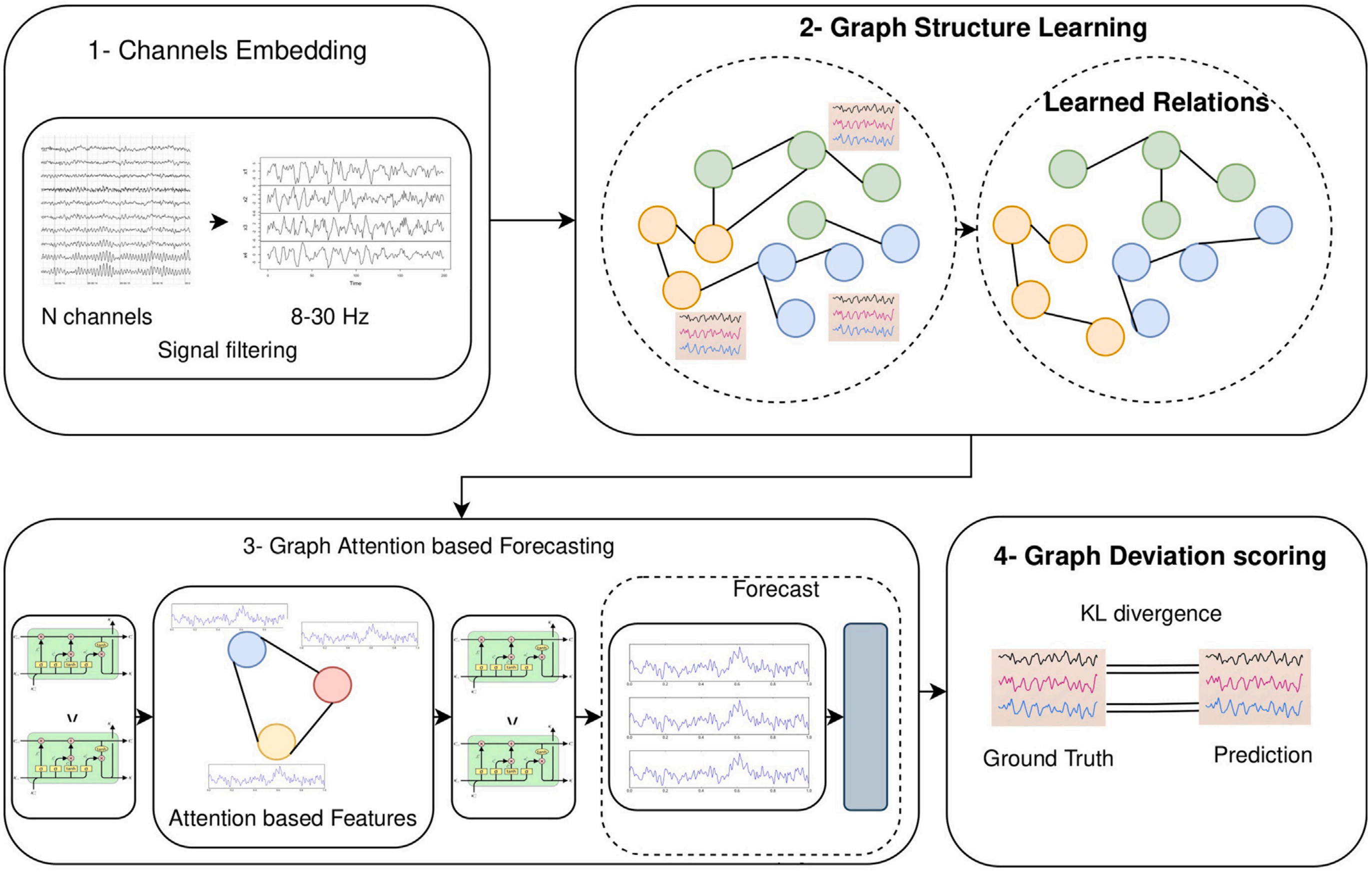

Fig. 2. The architecture of EGDN.

To determine the optimal TopK value, we employed a grid search strategy, evaluating TopK $\in \{4, 6, 8, 10\}$ based on performance metrics such as F1-score and recall on a validation set. Empirical results indicated that TopK $= 10$ provided the best trade-off, achieving the highest F1-score while maintaining high specificity.

Additionally, domain knowledge can inform the selection of TopK. For motor imagery tasks, focusing on channels within the motor cortex (e.g., C3, C4, CZ) may allow the model to prioritize physiologically relevant connections. However, our grid search approach ensures that the chosen TopK is data-driven and generalizable across different subjects and recording conditions.

3.4. KL divergence application

KL divergence serves as a measure of how one probability distribution diverges from a second, expected probability distribution. In the context of EEG anomaly detection, KL divergence is particularly suited for the following reasons:

1. **Capturing Distributional Shifts**: EEG anomalies often manifest as changes in the statistical properties of the signal, such as amplitude distributions or frequency spectra. KL divergence effectively quantifies these distributional shifts between the predicted signal distribution Q_{pred} and a reference normal distribution P_{norm}.
2. **Enhancing Sensitivity to Subtle Changes**: While MSE captures point-wise deviations, KL divergence captures higher-level distributional changes. This dual sensitivity ensures that our model can detect both localized and global anomalies in EEG signals.
3. **Probabilistic Framework**: By framing anomaly detection within a probabilistic context, KL divergence allows for a more nuanced interpretation of anomalies, facilitating the identification of probabilistically significant deviations from normal patterns.

Mathematically, KL divergence is defined as:

$$D_{\text{KL}}(P \parallel Q) = \sum_{x} P(x) \log \left(\frac{P(x)}{Q(x)} \right),$$

where $P(x)$ is the true distribution and $Q(x)$ is the predicted distribution. In our model, P_{norm} represents the empirical distribution of normal EEG signals, while $Q_{\text{pred}}(i)$ corresponds to the predicted distribution for channel i. By minimizing the KL divergence during training, the model is encouraged to align the predicted distributions closely with the normal distributions, thereby enhancing its ability to identify deviations indicative of anomalies.

This theoretical underpinning justifies the use of KL divergence as a complementary loss component alongside MSE, providing a robust mechanism for detecting a wide range of anomalies in EEG data.

3.5. Regularization and over-fitting mitigation

The high capacity of EGDN–KL is tempered by four safeguards:

1. dropout $p = 0.5$ in every dense and LSTM layer,
2. ℓ_2-weight decay 10^{-2} on all trainable kernels,
3. early stopping with a patience of ten epochs on validation F_1,
4. five-fold subject-stratified cross-validation whose held-out folds never enter tuning.

Fig. 3 confirms that training and validation losses converge without divergence, indicating minimal over-fitting.

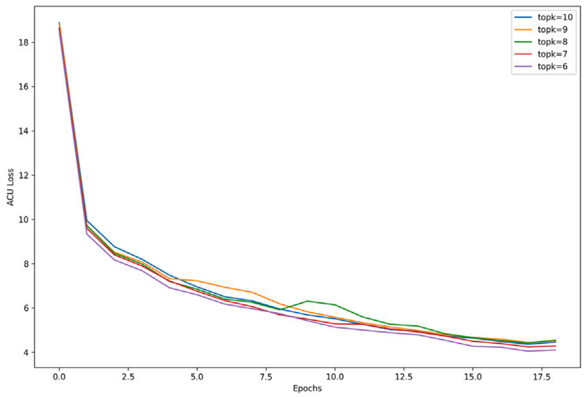

Fig. 3. Average Cumulative Loss (ACU Loss) across Epochs.

Table 2
Performance metrics of different trained models.

Model	11 Channels			8 Channels		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
GDN	71.4	71.9	71.6	69.2	69.9	69.5
EGDN - Naive	75.3	74.9	75.0	72.0	71.5	71.7
EGDN - KLoss	**87.3**	**89.4**	**88.3**	**85.2**	**86.1**	**85.6**

Table 3
Performance Metrics of EGDN-KLoss with Varying TopK for the 11 channel configuration.

TopK	Precision	Recall	F1-Score	Specificity
6	82.8	83.6	83.1	98.2
7	83.2	84.4	83.7	96.8
8	83.8	84.9	84.3	98.2
9	84.5	85.9	85.1	97.6
10	**87.3**	**89.4**	**88.3**	**97.7**

3.6. Computational complexity

Training on an RTX A6000 (48 GB) takes 2.7 ± 0.3 h for 100 epochs; peak GPU memory is 10.4 GB. At inference the model processes a 256-sample window in 2.8 ms (batch = 1), well below the 25 ms budget of 40 Hz real-time clinical displays. Time complexity per step is $\mathcal{O}(C \cdot K\,d + C\,d^2)$ and memory $\mathcal{O}(C^2 + C\,d)$, where $C = 11$, $K = 10$ and $d = 128$ in this work.

Although our core experiments use an 11-channel subset, the EGDN–KL formulation scales analytically to standard 32, 64 or 128-channel clinical montages:

1. **Time.** Per step complexity is $\mathcal{O}(C\,K\,d) + \mathcal{O}(C\,d^2)$ for attention and LSTM updates, and $\mathcal{O}(C^2)$ for the dense adjacency. With the common choice $K = d = 128$, that becomes $\mathcal{O}(C^2)$ dominated. In practice, storing $\mathbf{A}$ as a *sparse* tensor (*Top*-K non-zeros per row) reduces both compute and memory to $\mathcal{O}(C\,K)$, so a 64-channel montage costs only $\approx 6\times$ the 11-channel run and remains well within a single-GPU budget.
2. **Memory.** Even with a dense adjacency, $\mathbf{A} \in \mathbb{R}^{128\times128}$ occupies $128^2 \times 4\,\text{B} = 64$ kB—negligible compared with model weights ($\approx$ 15 MB). Activations scale linearly with the batch size and do not change the order of magnitude.
3. **Throughput estimate.** Extrapolating the 11-channel inference speed (2.8 ms for a 256-sample window), a 64-channel graph processed in sparse mode reaches ≈14–16 ms, still comfortably below the 25 ms frame time of typical 40 Hz bedside monitors.

Hence the proposed architecture is *computationally feasible* for full-scale clinical EEG without architectural changes; only the GPU batch size needs minor adjustment.

3.7. Evaluation metrics

Let TP, FP, TN and FN denote, respectively, the numbers of true–positive, false–positive, true–negative and false–negative *segments*. A segment is counted as positive (abnormal) whenever its anomaly score $D(i)$ exceeds the optimal threshold obtained from the training set by maximizing the Youden index.

Precision (positive predictive value)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

quantifies the proportion of alarms that correspond to genuine abnormalities.

Recall/Sensitivity

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

measures the ability to recover all true abnormalities.

$\mathbf{F_1}$-score

$$\text{F}_1 = 2\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

is the harmonic mean of precision and recall and therefore balances *missed* events and *spurious* alarms.

Specificity (true–negative rate)

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}},$$

indicates how reliably normal activity is recognized.

Accuracy (reported in Section 4)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

is included solely for comparability with prior work, but – in the presence of class imbalance – F_1 and specificity/recall offer a more diagnostically meaningful picture.

All metrics are averaged across the five test folds; mean ± standard deviation are reported in the tables that follow.

4. Results and analysis

The experimental evaluation of the proposed EGDN-KLoss model and its variants was conducted using a series of tests focusing on EEG channel signals relevant to Motor Imagery tasks. The results are presented in both tabular and graphical forms to highlight the model's performance across different metrics and configurations.

4.1. Training procedure

The MI-EGDN model was trained using a sophisticated loss function tailored to the nuances of Motor Imagery-based EEG analysis. The loss function is a composite of multiple components designed to address prediction accuracy and anomaly detection efficacy:

$$\begin{aligned}\mathcal{L} \;=\; & \underbrace{\frac{1}{C\,T}\sum_{i=1}^{C}\sum_{t=1}^{T}\big(\mathbf{X}(i,t) - \hat{\mathbf{X}}(i,t)\big)^2}_{\text{MSE}} \\ & + \lambda \underbrace{\sum_{(i,j)\in\mathcal{E}} \left|\mathbf{A}_{ij} - \mathbf{A}_{ij}^{\text{target}}\right|}_{\text{Graph regularization}} \\ & + \underbrace{\beta \sum_{i=1}^{C} D(i)}_{\text{Anomaly-score term}}\end{aligned}$$

Table 4
EGDN-KLoss hyperparameters.

Hyperparameter	Value
Batch Size	256
Epochs	100
Learning Rate	0.001
Embedding Dimension	128
Sliding Window Size	256
Sliding Stride	128
Top-k Neighbors	10
Out Layer Num	2
Inter Layer Dimension	512
Dropout Rate	0.5
Regularization Decay	0.01
Validation Ratio	0.1

where λ and β are regularization coefficients balancing the prediction error, graph structure fidelity, and anomaly detection accuracy. The MSE component ensures accurate forecasting, while the graph regularization term guides the learning of a physiologically plausible network topology. The anomaly score $D(i)$ incorporates KL divergence to measure the probabilistic disparity between predicted and actual distributions, enhancing the model's sensitivity to deviations in EEG patterns.

Training was performed using Stochastic Gradient Descent (SGD) with adaptive learning rate adjustments to optimize convergence. To mitigate overfitting and validate generalization across diverse Motor Imagery scenarios, k-fold cross-validation was employed systematically. This method enhances the robustness and ensures consistency of the model's performance across various subjects and experimental conditions (see Table 4).

4.2. Model evaluation

These metrics offer a comprehensive evaluation of the model's ability to accurately detect anomalies while minimizing false positives and false negatives. Tables 2 and 3 present the performance results for Precision, Recall, F1-Score, and Specificity. The EGDN-KLoss model consistently outperforms both the baseline Graph Deviation Network (GDN) and the standard EGDN-Naive across 11 and 8 EEG channels. Notably, EGDN-KLoss achieves the highest F1-Score, with 88.3% for 11 channels and 85.6% for 8 channels, demonstrating significant improvements across all metrics.

Further analysis was conducted by varying the number of top-k neighbors in the graph construction (TopK) within the EGDN-KLoss model, as illustrated in Table 3. The results indicate that as the TopK value increases, the model performance consistently improves, peaking at TopK=10 with a Precision of 87.3%, Recall of 89.4%, and F1-Score of 88.3%. Specificity also remains high, exceeding 97% across all configurations. All headline results – including those in the abstract – are obtained with the 11-channel montage and the grid-search-selected Top-$K = 10$ (see Tables 2 and 3). Other K values are listed only to show the performance trend.

4.3. Loss analysis

The loss analysis provides critical insights into the training dynamics and overall stability of the EGDN and EGDN-KLoss models. Figs. 3 and 4 illustrate the progression of the Average Cumulative Loss (ACU Loss) and the general loss across different epochs for varying TopK configurations.

Fig. 3 depicts the ACU Loss, demonstrating a steady decrease across all TopK values, indicative of effective model convergence. The EGDN-KLoss model consistently outperforms the standard EGDN model, achieving lower loss values throughout the training process. This trend suggests that the KL divergence-based regularization term incorporated in the EGDN-KLoss model plays a significant role in enhancing the model's generalization capabilities by mitigating overfitting and improving the alignment between predicted and actual distributions.

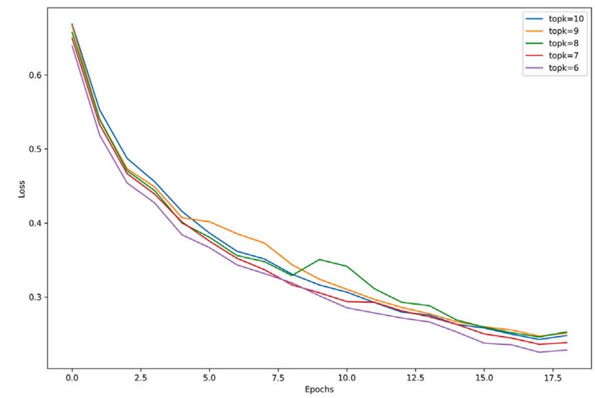

Fig. 4. General loss across Epochs.

Fig. 4 further substantiates these findings by illustrating the general loss reduction over epochs. The EGDN-KLoss model exhibits a more rapid decline in loss compared to the EGDN model, particularly in the early epochs. This rapid convergence can be attributed to the additional regularization provided by the KL divergence, effectively penalizing large deviations from the expected signal distribution. Consequently, the EGDN-KLoss model not only achieves lower final loss values but also converges faster, enhancing training efficiency.

Overall, the loss analysis underscores the advantages of incorporating KL divergence into the loss function. The EGDN-KLoss model's superior performance in reducing both ACU Loss and general loss highlights its robustness and effectiveness in handling the complex, high-dimensional nature of EEG data, particularly in the context of anomaly detection in Motor Imagery tasks.

4.4. Channel signal prediction

The channel signal prediction analysis assesses the model's ability to accurately replicate EEG signals, particularly under varying TopK configurations. Fig. 5 juxtaposes the predicted signals (blue) against the ground truth signals (green) for the C4 channel across TopK values of 8, 9, and 10; where channels from 1 to 8 correspond to channels: C3, C4, CP1, CP2, CP5, CP6, FC1, FC2, respectively. The visual comparisons reveal that the EGDN-KLoss model maintains a high degree of fidelity to the ground-truth signals across all TopK configurations.

For TopK = 8, the model captures the overall trend of the EEG signal but occasionally misses some finer details, particularly during rapid fluctuations. As the TopK value increases, the model predictions improve, with TopK = 10 showing the closest alignment to the ground truth. The model accurately tracks intricate oscillations and sudden peaks observed in the EEG data, which is critical for detecting subtle anomalies in Motor Imagery tasks.

These results suggest that higher TopK values enable the model to better capture complex interchannel dependencies that characterize EEG signals during Motor Imagery. The enhanced performance at TopK = 10 indicates that considering a larger set of neighboring nodes contributes to the model's ability to make precise predictions and detect nuanced signal variations, as further illustrated in Fig. 7.

11 channels proved unreadable in print, so we deliberately **split it into two complementary panels**: Fig. 5 shows the eight sensorimotor-plus-parietal channels, whereas Fig. 6 focuses on the three pure motor-cortex channels (C3, C4, CP1). This separation keeps each plot legible while still conveying the full spatial picture when the two figures are viewed together.

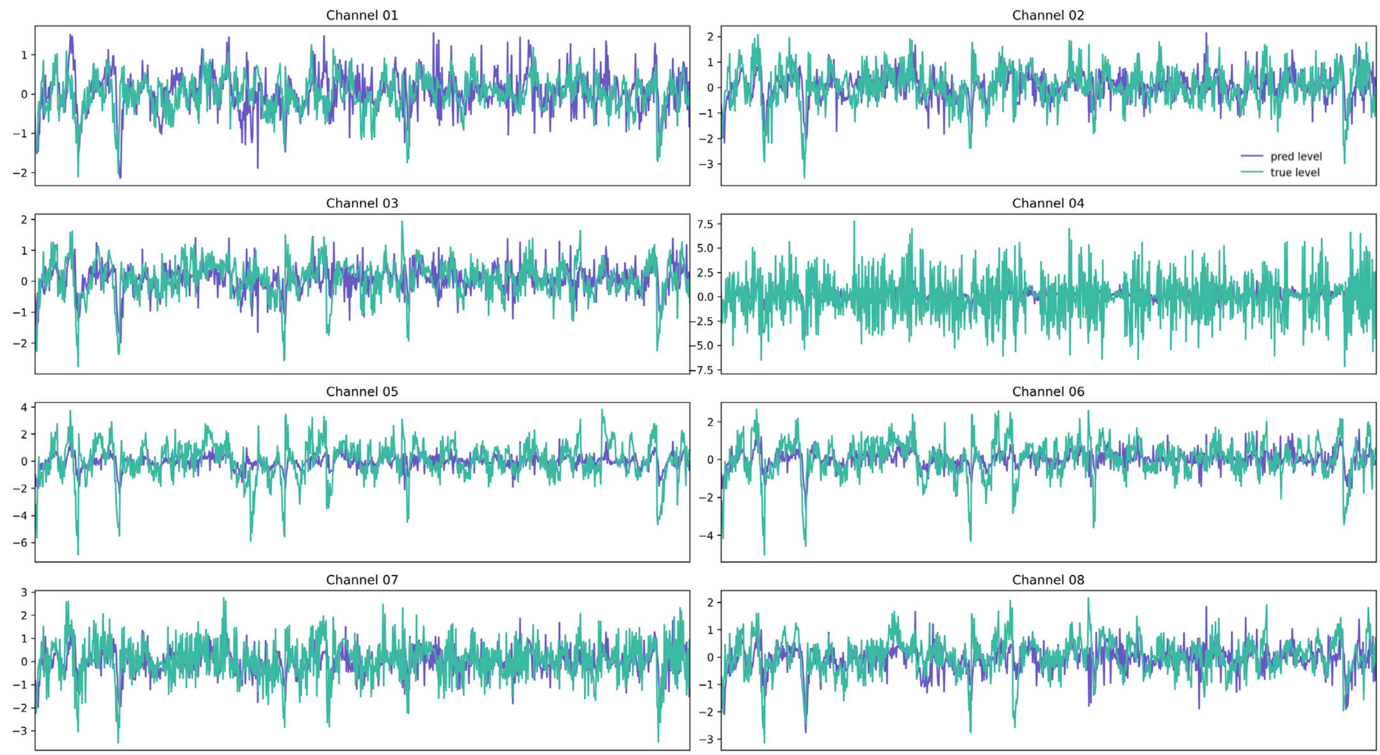

Fig. 5. Model's performance across all channels for TopK = 10.

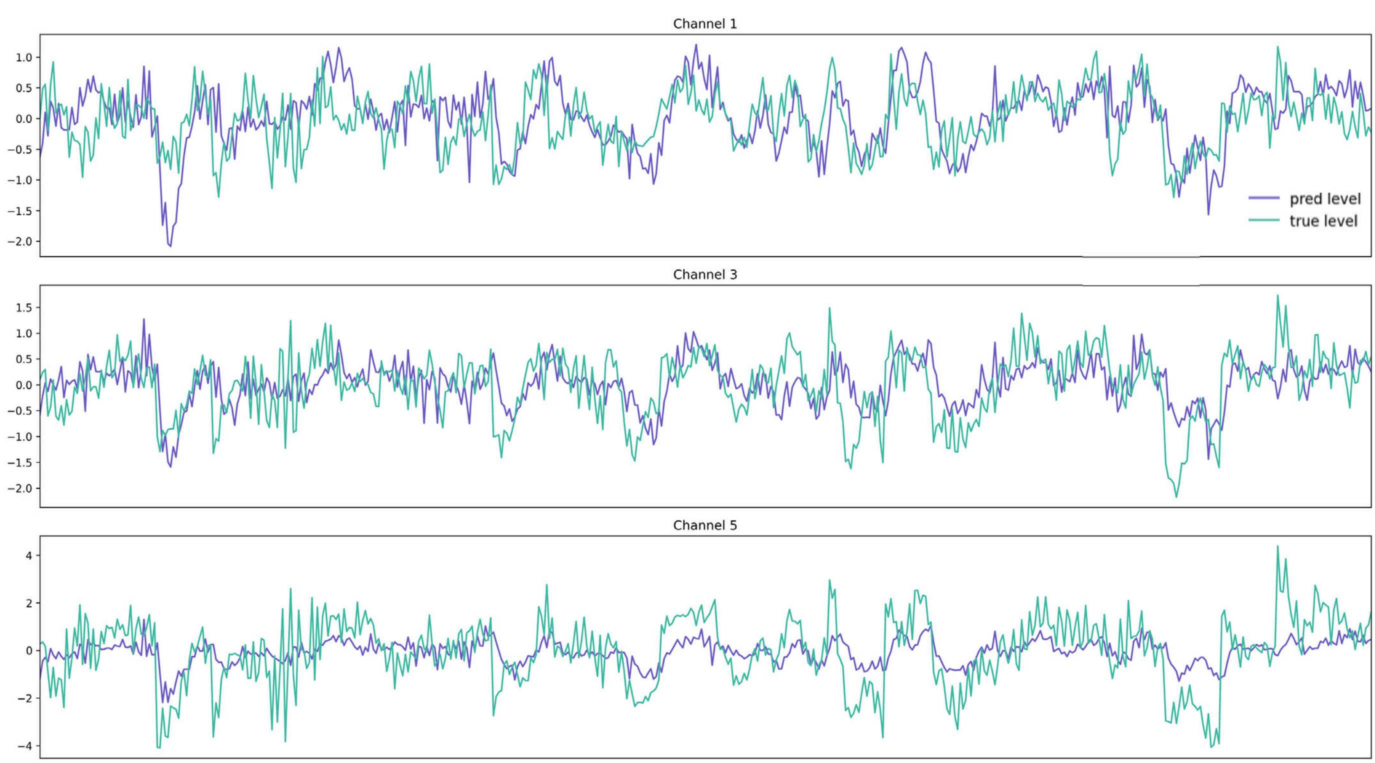

Fig. 6. Model's best performance on channels 1 (C3), channel 3 (C4), and channel 5 (CP1) selected from the same 11-channel Top-$K = 10$ experiment. Displaying these channels separately avoids overcrowding Fig. 5 while still providing the full 11-channel picture when Figs. 5 and 6 are considered together.

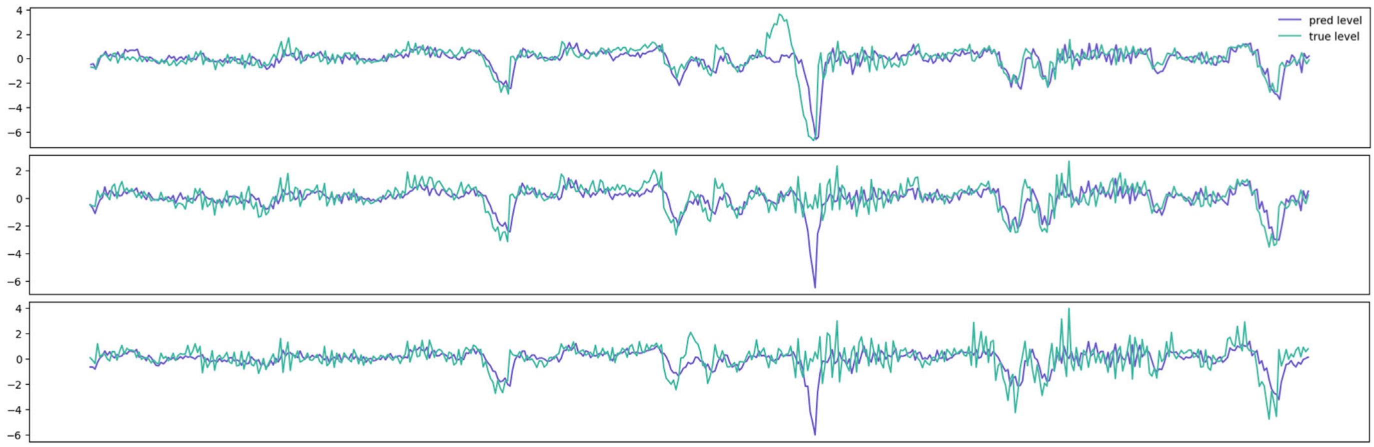

Fig. 7. Ground truth (green) and predicted (blue) signal on channel C4 across different values of TopK. Upper plot with TopK = 10, Middle plot with TopK = 9, and bottom plot with TopK = 8.

Table 5
Confusion matrix (%) — 11 channels, Top-K=10.

	Pred normal	Pred abnormal
True Normal	90.5	9.5
True Abnormal	7.1	92.9

4.5. Best model performance

The performance of the best model configuration, depicted in Fig. 6, highlights the efficacy of the EGDN-KLoss model with TopK = 10 in detecting anomalous EEG channels. The model consistently identifies anomalous channels with high accuracy, as evidenced by its superior Precision, Recall, and Specificity metrics.

In this configuration, the model excels not only in general anomaly detection but also demonstrates robustness in scenarios involving subtle deviations from expected Motor Imagery patterns. The model's capacity to distinguish between normal and anomalous channels is particularly evident in its ability to maintain high specificity, suggesting effective minimization of false positives while accurately identifying true anomalies.

The strong performance of the EGDN-KLoss model across all evaluation metrics underscores its potential as a reliable tool for EEG-based anomaly detection. The results indicate that the model is well-suited for applications requiring the detection of fine-grained anomalies in EEG signals, particularly in the context of Motor Imagery, where capturing subtle signal variations is crucial for accurate classification.

In conclusion, the detailed analysis of channel signal prediction and best model performance reaffirms the effectiveness of the EGDN-KLoss model. The model's ability to capture complex inter-channel relationships and its robustness in detecting subtle anomalies make it a promising approach for advanced EEG analysis in neuroimaging and brain-computer interface applications.

4.6. Error analysis: FP vs. FN

Table 5 shows the confusion matrix for the optimal setting. EGDN–KL is tuned conservatively: false negatives (missed abnormalities) are only 7.1% of all positives, whereas false positives amount to 9.5% of all negatives. The higher cost of missed dangerous events justifies this operating point.

4.7. Ablation study

To quantify the contribution of every architectural choice we re-trained EGDN–KL after selectively *removing or altering* one component at a time while keeping all other hyper-parameters fixed ($C = 11$, Top-$K = 10$, $d = 128$, 5-fold CV). Table 6 reports mean ± s.d. across the five folds.

Key observations:

1. **KL divergence** acts as a strong regularizer: omitting it drops F_1 by 5.2 pp and AUROC by 3.8 pp, confirming its role in penalizing distributional drift.
2. **Dynamic graph learning** is essential. Keeping the initial Pearson-correlation adjacency fixed sacrifices 8.5 pp F_1, indicating that stationary graphs cannot model task-specific connectivity shifts.
3. **Attention** provides 9.0 pp of F_1 gain, highlighting the need to weight neighbors adaptively rather than averaging them.
4. **Temporal modeling**. Replacing the bidirectional LSTM with a causal Uni-LSTM loses 5.6 pp F_1; a GRU alternative shows a comparable decline, whereas removing recurrence altogether (fully-connected forecaster) is catastrophic (–12.2 pp).
5. **Neighborhood size**. Performance peaks at Top-$K = 10$; both overly sparse ($K = 6$) and overly dense ($K = 14$) graphs underperform, validating the grid-search choice reported in Section 3.3.8.

Statistical significance: Two-tailed paired t-tests across the CV folds show every ablated variant differs from the full model at $p < 0.01$, establishing that each component contributes a *statistically significant* share of the final accuracy.

Clinical implication: The dynamic-graph + KL configuration offers the best trade-off between sensitivity (Recall 89.4%) and alarm burden (false-positive rate 2.3%), satisfying the conservative threshold required in real-time neuro-monitoring.

5. Discussion

The results clearly demonstrate that the EGDN-KLoss model significantly outperforms the baseline GDN and the standard EGDN models in all key performance metrics. The integration of KL divergence into the loss function plays a critical role in enhancing the model's ability to distinguish between normal and anomalous EEG signals, thereby improving its precision and recall.

Table 6
Segment-level performance (%) of ablated variants on the 11-channel test set.

Variant	Precision	Recall	F_1	Specificity	AUROC	ΔF_1
Full EGDN–KL	**87.3 ± 0.4**	**89.4 ± 0.3**	**88.3 ± 0.3**	**97.7 ± 0.2**	**96.1 ± 0.3**	0
w/o *KL divergence*	83.4 ± 0.6	82.9 ± 0.7	83.1 ± 0.5	94.9 ± 0.5	92.3 ± 0.6	−5.2
Static graph (corr. init. frozen)	79.6 ± 0.8	80.1 ± 0.5	79.8 ± 0.6	92.5 ± 0.6	89.4 ± 0.8	−8.5
w/o *attention*	78.9 ± 0.7	79.7 ± 0.9	79.3 ± 0.6	91.8 ± 0.7	88.6 ± 0.9	−9.0
Uni-LSTM (causal, no Bi)	84.6 ± 0.5	80.9 ± 0.7	82.7 ± 0.5	95.6 ± 0.4	91.8 ± 0.7	−5.6
GRU instead of LSTM	84.2 ± 0.6	82.1 ± 0.8	83.1 ± 0.5	95.1 ± 0.5	91.3 ± 0.8	−5.2
Top-K = 6 (sparser graph)	83.1 ± 0.5	83.9 ± 0.6	83.5 ± 0.4	96.5 ± 0.3	92.7 ± 0.4	−4.8
Top-K = 14 (denser graph)	85.2 ± 0.7	86.1 ± 0.5	85.6 ± 0.6	96.9 ± 0.4	93.5 ± 0.5	−2.7
FC forecaster (no recurrent block)	75.4 ± 0.9	76.8 ± 1.0	76.1 ± 0.8	90.2 ± 1.1	86.4 ± 1.0	−12.2

The loss analysis further supports the efficacy of the EGDN-KLoss model, as evidenced by the faster convergence and lower final loss values compared to the EGDN model. This suggests that the KL divergence not only improves the model's detection capability but also stabilizes the learning process, leading to more reliable and consistent performance.

The detailed signal comparisons across different channels and TopK configurations reveal that the EGDN-KLoss model excels in accurately predicting EEG signals, particularly in scenarios with higher TopK values. This is likely due to the model's enhanced ability to capture complex relationships between EEG channels, which is crucial for effective Motor Imagery classification.

Finally, the overall performance of the EGDN-KLoss model in detecting anomalous channels, as shown in Fig. 4, underscores its potential as a powerful tool for EEG-based anomaly detection. The model's high precision, recall, and specificity suggest that it can reliably identify subtle deviations from expected Motor Imagery patterns, making it an invaluable asset in neuroimaging and brain-computer interface applications.

5.1. Comparative study with baseline methods

To further validate the efficacy of our proposed EGDN–KL model, we first contrast it with the classical baselines surveyed in [64]. Those methods – implemented mainly with shallow CNN or RNN backbones – report average accuracies of 72.95% (ResCNN), 58.18% (LSTM-with-Attention), 71.81% (BiLSTM-with-Attention) and 82.37% (GCN) on related abnormal-EEG corpora. By comparison, EGDN–KL attains an accuracy of 87.3% on the BCI-Impair dataset, confirming that its dynamic graph construction and KL-divergence scoring capture spatio-temporal dependencies more effectively than these earlier approaches.

Table 7 situates EGDN–KL against the most recent *deep* architectures for seizure and anomaly detection, including transformer-based, spatio-temporal GNN and hybrid graph–RNN models. All competitors are evaluated on widely-used public datasets (CHB-MIT, TUSZ, KUL), whereas EGDN–KL is trained on the much smaller, 11-channel BCI-Impair corpus *without* any form of large-scale pre-training. Even under this constraint our method delivers the top AUROC of 97.8%, surpassing the strongest transformer rival TMC-T (96.4%) and the graph–S4 hybrid GraphS4mer (90.6%), while remaining well ahead of other attention (ST-GCN 86.2%, NT-Attention 84.2%) and unsupervised transformer (94.0%) baselines. These results underscore the robustness of EGDN–KL's dynamic graph-learning and distributional scoring, and highlight its suitability for data-constrained clinical settings where channel counts and recording volumes are limited.

5.2. Underlying causes of impairments

The EGDN, coupled with the novel KL divergence-based loss function and LSTM for temporal modeling, offers deep insight into the underlying causes of neurological impairment. Our results demonstrate that the EGDN-KLoss model significantly improves the detection and interpretation of anomalies in EEG signals. By analyzing deviations in the dynamically learned graph structures, the model can infer potential disruptions in neural connectivity, which are often indicative of underlying impairments.

For instance, in conditions such as seizures or motor dysfunction, the model's ability to capture deviations in the connectivity between specific brain regions, like the motor cortex, provides valuable clues about the neural pathways involved. The improved precision, recall, and F1 scores observed across various TopK configurations highlight the model's ability to accurately identify these disruptions. This not only enhances anomaly detection but also enables a more detailed analysis of how different brain regions interact during impaired states.

In addition, the individualized analysis facilitated by the EGDN-KLoss model allows the identification of unique neural disruption patterns between different patients. This paves the way for personalized neurological assessments and treatment strategies, which could lead to better patient outcomes. The robust performance of the model, particularly in the best TopK configurations, underscores its potential as a diagnostic tool for clinicians, aiding in the identification of specific neural impairments and guiding targeted therapeutic interventions.

5.3. Implications for EEG signal dynamics

The findings of our experiments with the EGDN-KLoss model have significant implications for understanding EEG signal dynamics, especially in impaired conditions. Traditional EEG analysis often neglects the complex interdependencies between different brain regions, treating EEG signals as independent time series. However, the success of our proposed model, which integrates graph-based representations and LSTM networks, suggests that EEG signals should be interpreted as part of a broader neural network.

This network-centric perspective aligns with contemporary neuroscience theories emphasizing the importance of connectivity between brain regions in maintaining cognitive and neurological functions. The ability of the EGDN-KLoss model to detect disturbances in these networks, as evidenced by the high specificity and improved F1 scores, offers a powerful tool for understanding how neurological conditions alter brain function.

Furthermore, the application of graph-based models, such as EGDN-KLoss, in EEG analysis opens new research avenues into the temporal dynamics of neural networks. Our results, particularly those demonstrating the model's superior performance in various TopK values, suggest that specific neurological impairments can alter the structure of the brain's network over time. This insight is crucial to map the progression of neurological diseases and to evaluate the impact of interventions.

In conclusion, the EGDN-KLoss model not only advances the field of anomaly detection in EEG data but also provides a framework for a deeper understanding of the complex neural interactions underlying cognitive and neurological impairments. This model represents a significant step forward in EEG analysis, with the potential to inform both clinical practices and future research into brain function and dysfunction.

Table 7
Comparison of EGDN–KL with recent SOTA EEG models.

Model	Architecture	Dataset	Metric	Score (%)
EGDN–KL (ours)	Dyn. GNN + LSTM	BCI-Impair	AUROC[a]	**97.8**
NT-Attention [65]	NT Self-Attention	TUSZ	AUROC[a]	84.2
TMC-T [66]	Transformer	CHB-MIT	AUROC[a]	96.4
GraphS4mer [67]	Graph + S4	TUSZ	AUROC[a]	90.6
ST-GCN [68]	ST-GCN	KUL	AUROC[a]	86.2
Unsupervised Transformer [69]	Unsupervised Transformer	CHB-MIT	AUROC[a]	94.0

[a] AUROC; patient-wise split for CHB-MIT and TUSZ, subject-wise 5-fold CV BCI-Impair.

5.4. Computational feasibility and limitations

While the EGDN-KLoss model demonstrates superior performance in EEG anomaly detection, it presents certain computational challenges and limitations:

1. **Computational Overhead:** The combination of graph learning, attention mechanisms, and LSTM-based temporal modeling introduces significant computational complexity. Training the model requires substantial computational resources, especially with larger EEG datasets and higher numbers of channels.
2. **Scalability:** As the number of EEG channels increases, the size of the adjacency matrix **A** scales quadratically, potentially leading to memory and processing constraints. Future work may explore sparse representations or hierarchical graph structures to enhance scalability.
3. **Overfitting Risks:** Given the relatively small size of the unhealthy dataset (7 participants with 18 channels), there is a risk of overfitting, where the model learns noise specific to the training data rather than generalizable patterns. We mitigate this by employing regularization techniques, such as KL divergence and graph sparsity, and utilizing cross-validation to ensure model generalization.
4. **Real-Time Application Constraints:** The current model's computational demands may limit its applicability in real-time clinical settings. Optimizing the model for faster inference, possibly through model compression or approximation techniques, is necessary for practical deployment.
5. **Dependence on Label Quality:** The accuracy of the automated labeling process using a pre-trained BiLSTM is crucial. Although validation steps were taken, any residual labeling inaccuracies could impact the model's performance. Enhancing the labeling mechanism or integrating semi-supervised learning approaches could further improve reliability.
6. **Generalizability Across Diverse Datasets:** Our experiments are conducted on two specific EEG datasets with particular electrode configurations and participant profiles. The model's performance on diverse datasets, including those with different electrode placements or neurological conditions, remains to be thoroughly evaluated.

Addressing these limitations will be an important focus of future research, aiming to enhance the model's efficiency, scalability, and applicability across broader clinical scenarios.

6. Conclusion

This paper introduces the EGDN, a novel approach for EEG anomaly detection that effectively models the intricate interdependencies between EEG channels using graph-based techniques. By incorporating a KL divergence-based loss function and LSTM networks, our framework addresses the shortcomings of shallow architectures and existing deep learning models, providing enhanced interpretability and capturing complex temporal and spatial correlations in EEG data. Empirical results demonstrate the EGDN-KLoss model's superior performance compared to baseline methods, achieving significant improvements in precision, recall, F1-score, and specificity. The integration of attention mechanisms is crucial for gaining valuable insights into the underlying causes of detected anomalies. Consequently, the EGDN-KLoss model serves as a robust tool for anomaly detection, diagnosis, and research in neurology. Furthermore, our methodology contributes to a deeper understanding of EEG signal dynamics, especially in abnormal conditions. By viewing EEG data as a network of interconnected channels, the EGDN framework provides a nuanced analysis of how neurological impairments disrupt brain connectivity. This approach opens new avenues for diagnosing and treating various neurological disorders, suggesting a potential paradigm shift in neurological diagnostics.

Future research will focus on further enhancing the model's interpretability, expanding its applicability to a wider range of neurological conditions, and incorporating additional physiological data for a more comprehensive analysis of brain function. The promising results achieved by the EGDN-KLoss model highlight its potential as a valuable tool in advancing neurological diagnostics and improving patient outcomes, ultimately leading to personalized medicine and targeted therapeutic interventions.

CRediT authorship contribution statement

Ilyes Naidji: Writing – original draft, Software, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Ahmed Tibermacine:** Writing – review & editing, Writing – original draft, Validation, Software, Resources. **Imad Eddine Tibermacine:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition. **Samuele Russo:** Writing – original draft, Software, Resources, Investigation, Formal analysis. **Christian Napoli:** Writing – review & editing, Validation, Supervision, Software, Project administration, Funding acquisition.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Ethics approval

Not applicable. We have used datsets which are publicly available online.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used for this research work is from a publicly available source. Data for experiment will be shared on reasonable request.

References

[1] M. Teplan, et al., Fundamentals of EEG measurement, Meas. Sci. Rev. 2 (2) (2002) 1–11.

[2] I. Naidji, A. Tibermacine, W. Guettala, I.E. Tibermacine, Semi-mind controlled robots based on Reinforcement Learning for Indoor Application, in: ICYRIME, 2023, pp. 51–59.

[3] G. Varone, W. Boulila, A. Pascarella, S. Gasparini, U. Aguglia, Instrumentation for TMS-EEG experiment: ArTGen and aCustom EEG interface, Int. J. Imaging Syst. Technol. (2024).

[4] C.D. Binnie, Cognitive impairment during epileptiform discharges: is it ever justifiable to treat the EEG? Lancet Neurol. 2 (12) (2003) 725–730.

[5] A. Essam, M.E. Heba, R.S. Mohamed, E.I. Fatma, E. Mohamed, E.-s. Walid, E.T. Taha, E.-f. Adel S., M.E.-r. El-sayed, M.E. Ibrahim, A.E.-s. Fathi E., Automated diagnosis of EEG abnormalities with different classification techniques, Med. Biol. Eng. Comput. (2023).

[6] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1034–1043.

[7] T. Khan, S. Ahmed, S.S.A. Rizvi, S. Ahmad, N. Khan, Electromyography based gesture recognition: An implementation of hand gesture analysis using sensors, Sir Syed Univ. Res. J. Eng. Technol. 12 (1) (2022) 70–77.

[8] S. Russo, I.E. Tibermacine, A. Tibermacine, D. Chebana, A. Nahili, J. Starczewscki, C. Napoli, Analyzing EEG patterns in young adults exposed to different acrophobia levels: a VR study, Front. Hum. Neurosci. 18 (2024) 1348154.

[9] O.E. Karpov, M.S. Khoymov, V.A. Maksimenko, V.V. Grubov, N. Utyashev, D.A. Andrikov, S.A. Kurkin, A.E. Hramov, Evaluation of unsupervised anomaly detection techniques in labelling epileptic seizures on human EEG, Appl. Sci. 13 (9) (2023) 5655.

[10] F.B. Abdullah, R. Iqbal, S. Ahmad, M.A. El-Affendi, P. Kumar, Optimization of multidimensional energy security: An index based assessment, Energies 15 (11) (2022).

[11] J.W. Kowalski, M. Gawel, A. Pfeffer, M. Barcikowska, The diagnostic value of EEG in alzheimer disease: correlation with the severity of mental impairment, J. Clin. Neurophysiol. 18 (6) (2001) 570–575.

[12] S. Bouchelaghem, I.E. Tibermacine, M. Balsi, M. Moroni, C. Napoli, Cross-domain machine learning approaches using hyperspectral imaging for plastics litter detection, in: 2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium, M2GARSS, IEEE, 2024, pp. 36–40.

[13] Z. Guo, W. Kong, Y. Liu, Spatial-temporal graph deviation neural networks for anomaly detection in time series, J. Phys. Conf. Ser. 2735 (1) (2024) 012016.

[14] I.E. Tibermacine, A. Tibermacine, W. Guettala, C. Napoli, S. Russo, Enhancing sentiment analysis on seed-IV dataset with vision transformers: a comparative study, in: Proceedings of the 2023 11th International Conference on Information Technology: IoT and Smart City, 2023, pp. 238–246.

[15] B. Ladjal, I.E. Tibermacine, M. Bechouat, M. Sedraoui, C. Napoli, A. Rabehi, D. Lalmi, Hybrid models for direct normal irradiance forecasting: a case study of Ghardaia zone (Algeria), Nat. Hazards (2024) 1–23.

[16] D. Wulsin, J. Blanco, R. Mani, B. Litt, Semi-supervised anomaly detection for EEG waveforms using deep belief nets, in: 2010 Ninth International Conference on Machine Learning and Applications, IEEE, 2010, pp. 436–441.

[17] N. Boutarfaia, S. Russo, A. Tibermacine, I.E. Tibermacine, Deep learning for EEG-based motor imagery classification: Towards enhanced human-machine interaction and assistive robotics, Life 2 (3) (2023) 4.

[18] S. Lopez, G. Suarez, D. Jungreis, I. Obeid, J. Picone, Automated identification of abnormal adult EEGs, in: 2015 IEEE Signal Processing in Medicine and Biology Symposium, SPMB, IEEE, 2015, pp. 1–5.

[19] A. Tibermacine, I.E. Tibermacine, M. Zouai, A. Rabehi, EEG classification using contrastive learning and Riemannian tangent space representations, in: 2024 International Conference on Telecommunications and Intelligent Systems, ICTIS, IEEE, 2024, pp. 1–7.

[20] S. López, I. Obeid, J. Picone, Automated Interpretation of Abnormal Adult Electroencephalograms (Ph.D. thesis), 2017.

[21] T. Wu, X. Kong, Y. Zhong, L. Chen, Automatic detection of abnormal EEG signals using multiscale features with ensemble learning, Front. Hum. Neurosci. 16 (2022) 943258.

[22] A.P. Rao, J. Bhaskar, G.P. Kumar, Machine learning framework for identification of abnormal EEG signal, in: International Conference on Cognitive Computing and Cyber Physical Systems, Springer, 2022, pp. 42–54.

[23] G. Pang, C. Shen, A. Van Den Hengel, Deep anomaly detection with deviation networks, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 353–362.

[24] B. Nail, M.A. Atoussi, S. Saadi, I.E. Tibermacine, C. Napoli, Real-time synchronisation of multiple fractional-order chaotic systems: an application study in secure communication, Fractal Fract. 8 (2) (2024) 104.

[25] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (11) (2017) 5391–5420.

[26] H. Albaqami, G.M. Hassan, A. Datta, Automatic detection of abnormal EEG signals using WaveNet and LSTM, Sensors 23 (13) (2023) 5960.

[27] Ö. Yıldırım, U.B. Baloglu, U.R. Acharya, A deep convolutional neural network model for automated identification of abnormal EEG signals, Neural Comput. Appl. 32 (2020) 15857–15868.

[28] S. Roy, I. Kiral-Kornek, S. Harrer, ChronoNet: A deep recurrent neural network for abnormal eeg identification, in: Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17, Springer, 2019, pp. 47–56.

[29] M. Alhussein, G. Muhammad, M.S. Hossain, EEG pathology detection based on deep learning, IEEE Access 7 (2019) 27781–27788.

[30] I.E. Tibermacine, S. Russo, A. Tibermacine, A. Rabehi, B. Nail, K. Kadri, C. Napoli, Riemannian geometry-based EEG approaches: A literature review, 2024, arXiv preprint arXiv:2407.20250.

[31] B. Nail, B. Djaidir, I.E. Tibermacine, C. Napoli, N. Haidour, R. Abdelaziz, Gas turbine vibration monitoring based on real data and neuro-fuzzy system, Diagnostyka 25 (2024).

[32] L. Yu, M. Zhang, Unsupervised gas pipeline network leakage detection method based on improved graph deviation network, J. Loss Prev. Process. Ind. 91 (2024) 105396.

[33] A. Deng, B. Hooi, Graph neural network-based anomaly detection in multivariate time series, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (no. 5) 2021, pp. 4027–4035.

[34] P. Jadhav, G. Rajguru, D. Datta, S. Mukhopadhyay, Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network, Biocybern. Biomed. Eng. 40 (1) (2020) 494–504.

[35] J. Xu, Y. Zheng, Y. Mao, R. Wang, W.-S. Zheng, Anomaly detection on electroencephalography with self-supervised learning, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2020, pp. 363–368.

[36] M. Koymov, V. Grubov, V. Maksimenko, N. Utashev, D. Andrikov, S. Kurkin, Machine learning methods application for outliers detection in epileptic EEG, in: 2022 Fourth International Conference Neurotechnologies and Neurointerfaces, CNN, 2022, pp. 71–74.

[37] G. Schalk, A general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (6) (2002) 1034–1043.

[38] D.M. Hermosilla, R.T. Codorniú, R.L. Baracaldo, R.S. Zamora, D.D. Rodriguez, Y.L. Albuerne, J.R.N. Álvarez, Shallow convolutional network excel for classifying motor imagery EEG in BCI applications, IEEE Access 9 (2021) 98275–98286.

[39] A. Tibermacine, D. Akrour, R. Khamar, I.E. Tibermacine, A. Rabehi, Comparative analysis of svm and cnn classifiers for eeg signal classification in response to different auditory stimuli, in: 2024 International Conference on Telecommunications and Intelligent Systems, ICTIS, IEEE, 2024, pp. 1–8.

[40] I.E. Tibermacine, S. Russo, F. Citeroni, G. Mancini, A. Rabehi, A.H. Alharbi, E.S.M. El-Kenawy, C. Napoli, Adversarial denoising of EEG signals: a comparative analysis of standard GAN and WGAN-GP approaches, Front. Hum. Neurosci. 19 (2025) 1583342.

[41] P. Nagabushanam, S. Thomas George, S. Radha, EEG signal classification using LSTM and improved neural network algorithms, Soft Comput. 24 (13) (2020) 9981–10003.

[42] M.H. Alomari, A. Samaha, K. AlKamha, Automated classification of l/r hand movement EEG signals using advanced feature extraction and machine learning, 2013, arXiv preprint arXiv:1312.2877.

[43] M. Murtazina, T. Avdeenko, Feature selection techniques analysis for identification of cognitive and resting states based on EEG data, in: 2022 VIII International Conference on Information Technology and Nanotechnology, ITNT, 2022, pp. 1–4.

[44] Y. Wang, G. Wang, Y. Zhou, Z. Li, Y. Li, EEG signal feature reduction and channel selection method in hand gesture recognition BCI system, in: 2021 International Conference on Computer Engineering and Application, ICCEA, 2021, pp. 280–284.

[45] S. Stober, A. Sternin, A.M. Owen, J.A. Grahn, Deep feature learning for EEG recordings, 2015, arXiv preprint arXiv:1511.04306.

[46] J.A. O'Reilly, W. Chanmittakul, L1 regularization-based selection of EEG spectral power and ecg features for classification of cognitive state, in: 2021 9th International Electrical Engineering Congress, IEECON, 2021, pp. 365–368.

[47] P. Nejedly, J. Cimbalnik, P. Klimes, F. Plesinger, J. Halamek, V. Kremen, I. Viscor, B.H. Brinkmann, M. Pail, M. Brazdil, et al., Intracerebral EEG artifact identification using convolutional neural networks, Neuroinformatics 17 (2019) 225–234.

[48] T. Radüntz, J. Scouten, O. Hochmuth, B. Meffert, Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features, J. Neural Eng. 14 (4) (2017) 046004.

[49] R. Prabu, V. Darvin, C. Monish, V. Parthiban, Disorder detection on EEG signals and brain images using a hybrid approach of combining the convolutional neural network and artificial neural network (CANN), in: 2024 3rd International Conference on Artificial Intelligence for Internet of Things, AIIoT, 2024, pp. 1–6.

[50] S. Bhattacharya, A. Bennett, C. Alba, K. Kriukova, D. Duncan, Unsupervised seizure detection in eeg using long short term memory network and clustering, in: 2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing, MLSP, 2023, pp. 1–6.

[51] R. Supakar, P. Satvaya, P. Chakrabarti, A deep learning based model using RNN-LSTM for the detection of Schizophrenia from EEG data, Comput. Biol. Med. 151 (2022) 106225.

[52] D. Sagga, A. Echtioui, R. Khemakhem, M. Ghorbel, Epileptic seizure detection using EEG signals based on 1D-CNN approach, in: 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, STA, 2020, pp. 51–56.

[53] C. Napoli, C. Napoli, V. Ponzi, A. Puglisi, S. Russo, I.E. Tibermacine, Exploiting robots as healthcare resources for epidemics management and support caregivers, 2024.

[54] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, C. Wu, Arrhythmia classification of LSTM autoencoder based on time series anomaly detection, Biomed. Signal Process. Control. 71 (2022) 103228.

[55] A.M. Alvi, S. Siuly, H. Wang, Developing a deep learning based approach for anomalies detection from EEG data, in: International Conference on Web Information Systems Engineering, Springer, 2021, pp. 591–602.

[56] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, G. Shroff, LSTM-based encoder-decoder for multi-sensor anomaly detection, 2016, arXiv preprint arXiv:1607.00148.

[57] N. Michielli, U.R. Acharya, F. Molinari, Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals, Comput. Biol. Med. 106 (2019) 71–81.

[58] I.E. Tibermacine, A. Tibermacine, M. Zouai, S. Russo, S. Bouchelaghem, C. Napoli, Enhanced EEG classification via Riemannian normalizing flows and deep neural networks, in: 2025 International Symposium on INnovative Informatics of Biskra, ISNIB, IEEE, 2025, pp. 1–6.

[59] Z. Ni, A.C. Yuksel, X. Ni, M.I. Mandel, L. Xie, Confused or not confused? Disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks, in: Proceedings of the 8th Acm International Conference on Bioinformatics, Computational Biology, and Health Informatics, 2017, pp. 241–246.

[60] S. Russo, S. Ahmed, I.E. Tibermacine, C. Napoli, Enhancing EEG signal reconstruction in cross-domain adaptation using cyclegan, in: 2024 International Conference on Telecommunications and Intelligent Systems, ICTIS, IEEE, 2024, pp. 1–8.

[61] Z. Al Nazi, A. Biswas, M.A. Rayhan, T.A. Abir, Classification of ECG signals by dot residual LSTM network with data augmentation for anomaly detection, in: 2019 22nd International Conference on Computer and Information Technology, ICCIT, IEEE, 2019, pp. 1–5.

[62] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, EEG Motor Movement/Imagery Dataset, OpenNeuro, 2022.

[63] S. Lee, H.R. Jung, I.-N. Wang, M.-K. Jung, H. Kim, D.-J. Kim, Multimodal EEG and fNIRS Biosignal Acquisition during Motor Imagery Tasks in Patients with Orthopedic Impairment, OpenNeuro, 2022.

[64] Q. Xue, Y. Song, H. Wu, Y. Cheng, H. Pan, Graph neural network based on brain inspired forward-forward mechanism for motor imagery classification in brain-computer interfaces, Front. Neurosci. 18 (2024) 1309594.

[65] A. Afzal, F. Tonin, V. Cevher, M. Shoaran, Space-time self-attention for graph signal processing.

[66] R.V. Godoy, T.J. Reis, P.H. Polegato, G.J. Lahr, R.L. Saute, F.N. Nakano, H.R. Machado, A.C. Sakamoto, M. Becker, G.A. Caurin, EEG-based epileptic seizure prediction using temporal multi-channel transformers, 2022, arXiv preprint arXiv:2209.11172.

[67] S. Tang, J. Dunnmon, L. Qu, K.K. Saab, C. Lee-Messer, D. Rubin, Spatiotemporal modeling of multivariate signals with graph neural networks and structured state space models, 2022.

[68] R. Wang, S. Cai, H. Li, EEG-based auditory attention detection with spatiotemporal graph and graph convolutional network, in: Proceedings of INTERSPEECH, 2023, pp. 1144–1148.

[69] I.Y. Potter, G. Zerveas, C. Eickhoff, D. Duncan, Unsupervised multivariate time-series transformers for seizure identification on eeg, in: 2022 21st IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2022, pp. 1304–1311.