ROBUST EEG CLASSIFICATION VIA GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalogram (EEG) classification has gained prominence due to its applications in medical diagnostics and brain-computer interfaces. However, EEG data is known to have a low signal-to-noise ratio, resulting in high variance in predictions across similar instances. To overcome this issue, we introduce *RoGra*, a novel approach leveraging residual graph convolutional networks for robust EEG classification. Our model incorporates dynamic time warping (DTW) to align temporal information and capture meaningful neighborhood relationships, enhancing robustness against artifacts. Experiments on three well-established EEG datasets demonstrate that *RoGra* outperforms baseline methods by up to 25%, marking the largest improvement in EEG classification accuracy since the introduction of the seminal EEGNet. Our code is publically available[1].

1 INTRODUCTION

Electroencephalogram (EEG) classification is a key task in time-series analysis, both from a theoretical and practical perspective, due to its inherent complexity and significant challenges (Parbat & Chakraborty, 2021; Pan et al., 2022). One of the main problems is that the EEG sequences are often contaminated by various artifacts, such as eye blinks, eye movements, and muscle activity (Kotte & Dabbakuti, 2020; Delorme, 2023), which can significantly degrade the quality of the recorded data. These artifacts, which originate from sources other than brain activity, obscure relevant neural signals and reduce the signal-to-noise ratio (Johnson, 2006), thereby impairing the performance of downstream applications such as classification and clinical analysis.

Existing EEG classification approaches typically focus on denoising as a preprocessing step, aiming to remove specific types of artifacts, as demonstrated in (Pan et al., 2022). However, the optimal denoising strategy remains an open question. Recent findings by Delorme (2023) suggest that automated denoising can, in fact, reduce classification performance, or at best, have no significant impact, raising concerns about the efficacy of such preprocessing steps.

In this work, we model EEG sequences as a graph, where edges represent the relationships between different EEG instances, transforming the EEG classification task into a node classification problem. We introduce a neural architecture for EEG classification, built on graph neural networks (GNNs) (Scarselli et al., 2008; Bresson & Laurent, 2017; Kipf & Welling, 2022). Leveraging the inherent denoising capabilities of GNNs, due to the smoothness properties of graph-based architectures (Ma et al., 2021), we demonstrate that our approach is resilient to noise and enhances classification accuracy. Our method, termed **Ro**bust EEG classification via **Gra**ph Neural Networks (*RoGra*), integrates an InceptionTime (Ismail Fawaz et al., 2020) module to capture high-level temporal features, which are further refined through a residual GNN layer (Bresson & Laurent, 2017; Liu et al., 2021). This layer processes information from neighboring nodes using a similarity-based adjacency matrix (Zha et al., 2022), treating each edge as a functional dependency between connected nodes. To account for time shifts between different time series, we compute similarities of data points with the help of Dynamic Time Warping (DTW) (Sakoe & Chiba, 1978). We argue that the similarity-based adjacency matrix introduces a beneficial inductive bias, improving classification performance for EEG data. Our method achieves a significant improvement of approximately 25% across various EEG datasets while maintaining stable performance. These results indicate that

[1]Git link redacted for double-blind review. Please check the .zip file in the supplementary materials.

GNNs can effectively mitigate the impact of diverse artifacts without compromising the integrity of the underlying neural signals.

Our contributions in this work are the following:

1. We introduce **Ro**bust EEG Classification via **Gra**ph Neural Networks, (*RoGra*), which exploits similarity-based graph neural networks within the context of EEG classification.
2. We formulate the EEG classification task as an inductive node classification problem, where weighted edges are constructed from the data matrix with the help of dynamic time warping.
3. We demonstrate the effectiveness of *RoGra* on both EEG classification and general time-series classification tasks. Our model achieves up to a 25% improvement in accuracy across three benchmark EEG datasets, representing the largest performance gain in EEG classification since the introduction of EEGNet! Additionally, *RoGra* proves to be generalizable, outperforming state-of-the-art time-series classification models by approximately 2% on four multivariate non-EEG time-series datasets.
4. We empirically show that *RoGra* is especially robust to noise. By adding successively more and more noise to the data, we show that *RoGra* is able to classify in highly noisy settings, where *RoGra* is able to maintain its performance.

2 Related Work

In the EEG literature, most models integrate various types of convolutions (spatial, temporal, and hybrid), along with normalization, pooling, and a final linear layer. Some pioneering models include EEGNet (Lawhern et al., 2018) and ShallowConvNet (Schirrmeister et al., 2017), both of which employ temporal convolutions and convolutional blocks. SCCNet (Wei et al., 2019) extends this by incorporating spatiotemporal convolution to learn spectral filtering. FBCNet (Mane et al., 2021) follows a similar approach to EEG-TCNet but incorporates spectral filtering in the initial stage. EEG-TCNet (Ingolfsson et al., 2020) utilizes causal convolutions, while TCNet-Fusion (Musallam et al., 2021) enhances this and concatenates the outputs of the first and second layers before the final classification step. MBEEGSE (Altuwaijri et al., 2022) is one of the first transformers and uses EEG blocks (Riyad et al., 2020) in combination with SE attention blocks (Altuwaijri & Muhammad, 2022). MAtt (Pan et al., 2022) introduces a novel approach by utilizing manifold attention layers in Riemann space instead of the standard Euclidean space. Furthermore, Burchert et al. (2024) propose ResNet (Kachuee et al., 2018) and Inception (Ismail Fawaz et al., 2020) as robust baselines, along with a different training protocol for joint subject training. One of the main challenges in EEG classification literature is the large variance observed when training models on different subjects for the same task. Additionally, model performance is highly dependent on the choice of architecture and hyperparameters, leading to high standard deviations. As a result, it becomes difficult to identify the most suitable models for a given task, as many results lack statistical significance.

Signals in EEG datasets inherently contain brain activity alongside various sources of noise and artifacts (Zhang et al., 2021). Numerous studies have addressed these noise issues, originating from sources such as ocular movements (Croft & Barry, 2000; Chan et al., 2010) and muscle activity (McMenamin et al., 2010; Nekrasova et al., 2022). Traditional denoising techniques, such as regression and linear filtering methods, have been widely used to mitigate noise in EEG signals (Lai et al., 2018; Grobbelaar et al., 2022). However, these approaches often risk removing or distorting important physiological information, thereby reducing classification performance (Lai et al., 2018). Advanced methods, including blind source separation (Taha & Abdel-Raheem, 2022) and empirical mode decomposition (Soler et al., 2020), have also been explored but struggle with non-linear or overlapping artifacts. Furthermore, signal decomposition methods like Wavelet Transform (Borse, 2015; Alyasseri et al., 2019) have gained popularity, though they rely heavily on the selection of an appropriate wavelet basis and thresholding, which can be challenging and may lead to the loss of significant signal components (Lai et al., 2018; Grobbelaar et al., 2022). Therefore, mitigating the impact of diverse noise sources without compromising the integrity of the neural signals remains an open question.

Graph neural networks (GNNs) capture dependencies in a graph by facilitating information exchange between nodes (Zhou et al., 2020). The problem of time series classification can be approached from two distinct perspectives: as a graph classification task, referred to as Series-as-Graph, or as a node classification task, referred to as Series-as-Node (Jin et al., 2024). In the Series-as-Graph approach, time series classification was first explored by Time2Graph (Cheng et al., 2020), which extracts time-aware shapelets to build a shapelet-based graph for classification. This was extended to Time2Graph+ (Cheng et al., 2021), introducing time-level attention to capture shapelet evolution. MTS2Graph (Younis et al., 2024) combines CNNs and clustering to extract patterns and build graphs for classification, while TodyNet (Liu et al., 2024) avoids predefined graphs, using temporal graph pooling to capture spatio-temporal dependencies. From the Series-as-Node perspective, SimTSC (Zha et al., 2022) represents each time series as a node in a graph, with edges weighted by similarity. GNN operations generate node embeddings, which are then classified. In this work, we formulate EEG classification task from a Series-as-Node perspective as in (Zha et al., 2022) where edges are constructed from the data matrix.

GNNs have recently gained attention for EEG classification, with various GNN architectures being tailored to specific EEG tasks (Klepl et al., 2024) such as emotion recognition (Song et al., 2018; Zhou et al., 2023), epilepsy diagnosis (Wang et al., 2023), seizure detection (Ho & Armanfard, 2023; Tang et al., 2021), sleep staging (Eldele et al., 2021), and motor imagery (Jin et al., 2021). Many state-of-the-art GNN-based approaches leverage pre-defined structural connectivity to represent the physical connections between EEG sensors (Zhong et al., 2020; Lin et al., 2021), and rely on feature extraction techniques involving convolutional neural networks (CNNs) (Jia et al., 2021), long short-term memory (LSTM) (Hou et al., 2020), or multi-layer perceptrons (MLPs) (Sun et al., 2022). These features are then fed into different GNN variants, such as simplified Graph Neural Networks (GCNs) (Klepl et al., 2022), first-order ChebConv (Raeisi et al., 2022), and Graph Attention Networks (GATs) (Priyasad et al., 2022). However, current state-of-the-art graph-based EEG classification methods have several limitations. First, they are only tailored to specific tasks, lacking a generalizable framework for diverse EEG classification problems. Additionally, these models typically construct graphs based on the physical locations of EEG sensors, which fails to capture the functional dependencies between EEG signals. Furthermore, like other EEG-based approaches, these models often show only moderate performance, as they struggle to fully leverage the complex information inherent in EEG signals.

3 Preliminaries

3.1 Problem Setting: EEG Time Series Classification

Given a set of EEG recordings from a subject and their classification into distinct categories (e.g., by an expert), the objective is to classify new EEG recordings from this subject into these categories based on patterns in the signals. Each EEG recording is represented by a time series $X \in \mathbb{R}^{C \times T}$, where C is the number of EEG channels (electrodes) and T the length of the recording in time steps, each class label by a number $y \in \{1, \dots, K\}$. Given N such labeled EEG recordings, i.e., pairs $(X_1, y_1), \dots, (X_N, y_N)$ from an unknown distribution p (e.g., representing a subject), the task is to find a model $\hat{y}$ that maps EEG signals X to the correct class (where X and its ground truth class y are from the same distribution p). Correctness is measured simply by a loss function, typically accuracy or for problems with imbalanced class distribution area under the curve (AUC; see Appendix A.2). To achieve this goal, the model has to capture both spatial (across channels C) and temporal (across time points T) patterns.

We focus on the standard **inductive** problem setting, where the model must make predictions for each test instance independently, without access to the features of other test instances. In contrast, in the **transductive** setting, models have access to all test features collectively, providing additional information. We emphasize this distinction to differentiate our work from transductive approaches in the literature.

3.2 EEG time series classification as node classification

We represent EEG data as a graph from a novel perspective. Instead of using channels as nodes and treating EEG time series classification as a graph classification task, we construct a graph $G = (V, E)$ where each node $v_i \in V$ corresponds to a full EEG sequence X_i. Edges E denote relationships between these time series, based on a similarity matrix A. Each EEG sequence X_i has

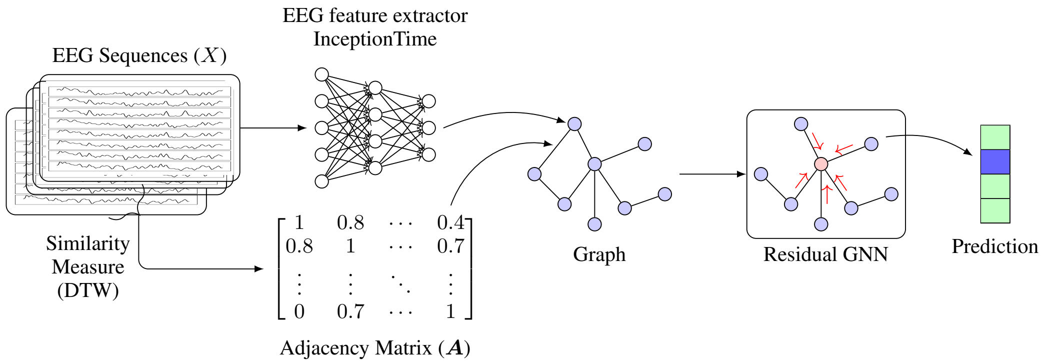

Figure 1: Our proposed model *RoGra*

a label $y_i \in \{1, 2, \ldots, K\}$, where K is the number of distinct classes representing different brain states or conditions. The objective is to predict these labels $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ for all nodes in the graph.

4 METHODOLOGY

While the vast majority of recent EEG classification models follow one common architecture, deep convolutional neural networks, with many different layers and choices in detail (see sec. 2), we are interested in a model than can leverage both, a rich encoder of the EEG recording and a distance measure d between such EEG recordings in an end-to-end learnable way in a graph neural network. Here, the graph neural network is build not over a single instance, e.g., over the different channels, but each EEG instance is a node. Let us have in the following a sequence of EEG training series $\boldsymbol{X} = (X_1, \ldots, X_N) \in (\mathbb{R}^{C \times T})^N$. RoGra contains the steps we discuss in the following. It is depicted in Figure 1 and the training procedure is described in Algorithm 1.

1. DTW Distance and Graph Construction The EEG sequences $X_1, \ldots, X_n$ denote the nodes and the adjacency matrix (and therefore the edges) is constructed as follows. We are using dynamic time warping (DTW; Sakoe & Chiba 1978) to measure distances between two EEG recordings, allowing to re-align patterns that slightly shifted in time between different instances:

$$d^{\text{DTW}}(X_m, X_n) := \min\{\sum_{i=1}^{|w|} d(X_{m,w_{i,1}}, X_{n,w_{i,2}}) \mid w \in (\{1, \ldots, T\}^2)^* \text{ warping path}\}$$

where a sequence w of index pairs is called a warping path if it starts at $(1, 1)$, ends at (T, T) and in each step each index increases by 1 or stays at its previous value. We convert distances into similarity values used for edge weights in a weighted adjacency matrix A via an exponential decay (Zha et al., 2022):

$$A_{m,n} := \exp(-\alpha \cdot d^{\text{DTW}}(X_m, X_n)), \quad m, n \in \{1, \ldots, N\} \tag{1}$$

where $\alpha \geq 0$ controls how quickly the weights decay with increasing DTW distance; in all our experiments we used $\alpha := 0.3$. We sparsify the weighted adjacency matrix A by keeping only the largest J values in every row and setting all others to zero, in effect dropping the edges between these nodes.

2. Instance Encoding. We utilize InceptionTime (Ismail Fawaz et al., 2020) to extract temporal features from each EEG recording X_n. It employs multi-scale convolutional layers with various filter sizes $k \in \{1, 3, 5\}$ to capture temporal patterns at different resolutions and then concatenates all of them to an initial latent representation:

$$Z_n^0 := \text{inception}(X_n; \theta^{\text{enc}}) \in \mathbb{R}^F \tag{2}$$

where $F \in \mathbb{N}$ is the latent feature dimension and θ^{enc} are the parameters of the inception encoder (e.g., its kernel matrices). Z_n^0 now represents the initial latent embedding of instance X_n.

3. Information Fusion with Residual Graph Neural Networks. Once the temporal features Z^0 are extracted, we use Residual Graph Neural Networks (ResGNN; Bresson & Laurent 2017; Liu et al. 2021) to fuse the two types of information: the distance information in the graph G, esp. its weighted adjacency matrix A, and the initial encodings Z_n^0 of the EEG recordings, used as node features in the first GNN layer. Each layer performs one step of message passing between neighboring nodes, transforming the $N \times d^l$ matrix Z^l of all node features to a $N \times d^{l+1}$ matrix Z^{l+1}

$$Z^{l+1} := \phi(A' Z^l W_l) + Z^l W_l' \tag{3}$$

if the latent dimension changes ($d_{l+1} \neq d_l$), and just with a residual link otherwise:

$$Z^{l+1} := \phi(A' Z^l W_l) + Z^l \tag{4}$$

where W_l, W_l' are trainable weight matrices. Here $A' = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ denotes the normalized adjacency matrix, where D is the weighted degree matrix, the diagonal matrix containing the row sums of A, and ϕ an activation function. The output of the last, L-th layer we denote as

$$\text{ResGNN}(A, Z^0; \theta^{\text{GNN}}) := Z^L \tag{5}$$

with parameters $\theta^{\text{GNN}} := (W_l, W_l')_{l=0:L-1}$.

4. Node Classification Head. We choose the number of classes as last embedding dimension ($d_L := K$) and finally apply softmax on the output of the last GNN layer (on top of its activation function), to predict class probabilities:

$$\hat{y}_n := \text{RoGra}(X; \theta)_n := \text{softmax}(Z_n^L) \tag{6}$$

with parameters $\theta := (\theta^{\text{enc}}, \theta^{\text{GNN}})$.

5. Loss and Training Procedure. During training time, we infer the predicted labels with RoGra and train with cross-entropy:

$$\ell(\theta; X, y) := \frac{1}{N} \sum_{n=1}^{N} \text{cross-entropy}(y_n, \hat{y}_n) \quad \text{with } \hat{y} := \text{RoGra}(X; \theta) \tag{7}$$

6. Inference for Test Examples. At inference, RoGra will build a separate graph for each test instance X^{qu}. This is done by building a graph having the training examples and the particular test example as nodes. We then apply RoGra, i.e., step 1 to 4 to $(X_1, \ldots, X_N, X^{\text{qu}})$ and use the predictions for the test instance:

$$\hat{y}(X^{\text{qu}}) := \text{RoGra}((X_1, \ldots, X_N, X^{\text{qu}}); \theta)_{N+1}$$

Equipped with this inference procedure RoGra is a fully inductive model that can predict each instance separately. For performance reasons, we batch the training examples and compute the similarity matrix only in the batch and apply RoGra batch wise. For inferring a test example, we sample a batch of training instances $X_1, ..., X_B$ and compute $\text{RoGra}((X_1, \ldots, X_B, X^{\text{qu}}))$.

7. Delineation from SimTSC and Kernel-based Methods. While *RoGra* shares some conceptual similarities with SimTSC (Zha et al., 2022) in constructing the adjacency matrix, it introduces key distinctions that set it apart. *RoGra* leverages an InceptionTime encoder to process time series data or EEG recordings. It utilizes a ResGNN architecture to dynamically update node representations. Unlike approaches that build a graph over channels, separately for each instance, *RoGra* constructs a graph over all training instances and the test instance being predicted. For each test instance X^{qu}, a graph $G(X^{\text{qu}})$ is formed with nodes $V(X^{\text{qu}}) := \{X_1, \ldots, X_N, X^{\text{qu}}\}$, and edges are created based on nearest neighbors according to the distance measure d. This setup differs from the transductive inference approach used in SimTSC, where a single graph is built over both the training and several test instances. It also contrasts with kernel-based models, as *RoGra* constructs its graph using only the training predictors X_n, without incorporating the class labels y_n.

Algorithm 1 RoGra Training

Require: Training dataset $\mathcal{D}^{\text{train}} = \{(X_1, y_1), \ldots, (X_N, y_N)\}$, scaling factor α, batch size $B \leq N$, number of epochs I
1: Initialize θ^{enc}, θ^{GNN} randomly
2: **for** epoch $i = 1$ to I **do**
3: Partition data $\mathcal{D}^{\text{train}}$ in batches $\{(X_b^{(j)}, y_b^{(j)})_{b=1:B} \mid j = 1, \ldots, \lceil N/B \rceil\}$
4: **for** each batch $(X_b^{(j)}, y_b^{(j)})_{b=1:B}$ **do**
5: $A_{ab} \leftarrow \text{ComputeSimilarity}(X_a^{(j)}, X_b^{(j)}) \quad \forall a, b \in \{1, \ldots, B\}$ ▷ Equation (1)
6: $Z_b^0 \leftarrow \text{inception}(X_b^{(j)}; \theta^{\text{enc}}) \quad \forall b \in \{1, \ldots, B\}$ ▷ Equation (2)
7: $Z^L \leftarrow \text{ResGNN}(A, Z^0; \theta^{\text{GNN}})$ ▷ Equation (5)
8: $\hat{y}_b \leftarrow \text{softmax}(Z_b^L) \quad \forall b \in \{1, \ldots, B\}$ ▷ Equation (6)
9: $\mathcal{L} \leftarrow \frac{1}{B} \sum_{b=1}^{B} \text{cross-entropy}(\hat{y}_b^{(j)}, y_b^{(j)})$ ▷ Equation (7)
10: Update parameters: $\theta^{\text{enc}}, \theta^{\text{GNN}}$ based on $\nabla_\theta \mathcal{L}$
return $\theta^{\text{enc}}, \theta^{\text{GNN}}$

5 EXPERIMENTS

We compare our model *RoGra* in two settings, EEG classification and time-series classification (TSC). For the application domain of EEG, we evaluate our method with the current state-of-the-art models for EEG Classification including Inception and InceptionJoint (Burchert et al., 2024), MAtt (Pan et al., 2022), MBEEGSE (Altuwaijri et al., 2022), FBCNet (Mane et al., 2021), TCNet-Fusion (Musallam et al., 2021), EEG-TCNet (Ingolfsson et al., 2020), SCCNet (Wei et al., 2019), EEGNet (Lawhern et al., 2018), and ShallowConvNet (Schirrmeister et al., 2017). Additionally, we analyze *RoGra* in the broader context of TSC on five multivariate UCR (Dau et al., 2019) datasets. Here we compare against SimTSC(Zha et al., 2022), HIVE-COTE2 (Middlehurst et al., 2021), Hydra-MR (Tan et al., 2022), and H-InceptionTime (Ismail-Fawaz et al., 2022).

5.1 DATASETS

For EEG classification, we experiment on the following three datasets representing three different classification targets, motor imagery, visual stimuli, and error recognition, respectively.

MI – Motor Imagery (Brunner et al., 2008). Originally released for the BCI Competition IV in 2008 as dataset BCIC-IV-2a, it is widely used in EEG-based studies and consists of recordings from 9 subjects. The EEG signals were collected using 22 Ag/AgCl electrodes placed over central and surrounding scalp regions, with a sampling rate of 250 Hz. The motor imagery task in this dataset includes four classes, where subjects were asked to imagine one of four movements: right hand, left hand, feet, or tongue. Standard preprocessing procedures were applied to the 22-channel data, which involved down-sampling the signals from 256 Hz to 128 Hz, followed by band-pass filtering to retain frequencies between 4 Hz and 38 Hz. The signals were then segmented, beginning 0.5 seconds after the onset of the cue and continuing for 4 seconds, resulting in segments containing 438 time points.

SSVEP – Steady-State Visual Evoked Potentials (Nikolopoulos, 2021). Released 2016 by the MAMEM project as dataset II, it includes EEG recordings from 11 subjects, using an EGI 300 Geodesic EEG System (GES 300). During the task, subjects focused on one of five visual stimuli flickering at specific frequencies (6.66, 7.50, 8.57, 10.00, and 12.00 Hz) for a duration of five seconds. Preprocessing of the EEG signals involved applying a band-pass filter between 1 Hz and 50 Hz. Eight channels, located in the occipital region of the brain (PO7, PO3, PO, PO4, PO8, O1, Oz, and O2), where the visual cortex is situated, were selected for analysis. Each trial was divided into four 1-second segments, starting 1 second after the cue onset and continuing for the next four seconds. This produced a total of 500 trials of 1-second, 8-channel SSVEP signals for each subject, with each segment consisting of 125 time points.

Table 1: Performance comparison for the datasets BCIC-IV-2a (MI), MAMEM EEG SSVEP (SSVEP) and the BCI challenge error-related negativity (ERN). We report the average accuracy for MI and SSVEP and the AUC for ERN over 5 runs respectively. The **best** result is highlighted in bold and the <u>second best</u> is underlined. Overall we achieve an average relative increase in performance of 18.78% over the current state-of-the-art.

Model	**MI**	**SSVEP**	**ERN**
ShallowConvNet	61.84±6.39	56.93±6.97	71.86±2.64
EEGNet	57.43±6.25	53.72±7.23	74.28±2.47
SCCNet	71.95±5.05	62.11±7.70	70.93±2.31
EEG-TCNet	67.09±4.66	55.45±7.66	<u>77.05</u>±2.46
TCNet-Fusion	56.52±3.07	45.00±6.45	70.46±2.94
FBCNet	71.45±4.45	53.09±5.67	60.47±3.06
MBEEGSE	64.58±6.07	56.45±7.27	75.46±2.34
MAtt	<u>74.71</u>±5.01	65.50±8.20	76.01±2.28
Inception	62.85±3.21	62.71±2.95	73.55±5.08
InceptionJoint	61.38±1.57	<u>66.00</u>±0.36	76.13±0.95
RoGra (ours)	**92.09**±1.45	**71.33**±1.79	**96.29**±1.57
Increase in %	23.27	8.08	24.98

ERN – Error-Related Negativity (Margaux et al., 2012). Released 2015 as part of the BCI Challenge NER 2015[2], it captures EEG data from 16 subjects from 56 Ag/AgCl electrodes. Subjects are performing a P300-based BCI spelling task, a binary classification challenge, with an inherent class imbalance due to more frequent correct inputs. Preprocessing involved down-sampling the signals from 600 Hz to 128 Hz and applying a band-pass filter between 1 Hz and 40 Hz. After processing, each trial was composed of 56 channels, with 160 time points per trial.

5.2 Experimental Setup

We compare our model *RoGra* against the EEG Classification baselines for these three datasets. For the MI and SSVEP datasets, accuracy is used as the performance metric, while AUC is employed for ERN due to class imbalance. We split the data 80/20 uniform at random for training and testing respectively. The baselines use a slightly different split, where for MI there is a fixed train/test split by instances, and SSVEP as well as ERN are split time-wise by sessions. There is no significant difference between the two protocols for model performance. We show this for SSVEP in the appendix in Table 9, where we apply our split for MAtt. We train *RoGra* for 500 epochs and we repeat the training 5 times. For our model, we did no additional hyperparameter optimization for any of the three datasets and used a learning rate of $1e^{-4}$, weight-decay of $4e^{-3}$, and dropout of 0.5 in the inception backbone for all experiments. Additionally, we set the scaling factor α to 0.3 and assigned each node 3 neighbors (J = 3). We use two layers of ResGNN in all of our experiments for fair comparison. The results for the baselines were aggregated from (Burchert et al., 2024) and (Pan et al., 2022).

5.3 EEG Classification Results

In Table 1, we show the performance of *RoGra* compared to state-of-the-art EEG classification methods. On the first dataset, MI, we achieve a significant 23.27% lift over the second-best model, MAtt, as shown in Table 2. For the second dataset, SSVEP, we also observe a performance increase of 8.08% compared to the InceptionJoint model. However, this model employs a different evaluation protocol, training all subjects jointly. We further compare performance for individual subjects directly in Table 3. Notably, *RoGra* demonstrates consistent performance for challenging subjects; for instance, in the case of subjects 4,5 and 8, where all other models are incapable of learning useful patterns and default to random performance. For the last dataset, ERN, we again achieve a significant lift of 24.98% over the second-best model, EEG-TCNet. The results for individual subjects for ERN can be found in the Appendix in Table 8

[2]https://www.kaggle.com/c/inria-bci-challenge

Table 2: Performance comparison for the **MI** dataset. The **best** result is highlighted in bold and the <u>second best</u> is underlined. Overall we achieve an average increase in performance of 23.3% over the current state-of-the-art for EEG classification models.

Subject	**Inception**	**MAtt**	**RoGra**
1	78.96 ± 1.82	<u>86.94</u> ± 1.36	**94.83** ± 2.18
2	41.25 ± 2.63	<u>56.00</u> ± 3.27	**77.06** ± 1.02
3	82.09 ± 3.05	<u>88.33</u> ± 1.17	**98.28** ± 1.41
4	52.43 ± 2.40	<u>67.85</u> ± 3.42	**91.38** ± 1.02
5	38.75 ± 3.74	<u>61.32</u> ± 1.07	**92.81** ± 1.82
6	48.61 ± 1.78	<u>67.00</u> ± 2.54	**85.06** ± 2.38
7	75.99 ± 4.51	<u>91.18</u> ± 0.89	**98.28** ± 0.86
8	74.31 ± 3.28	<u>83.06</u> ± 2.01	**96.29** ± 1.08
9	73.26 ± 2.14	<u>81.18</u> ± 1.06	**94.83** ± 1.28

Table 3: Performance comparison for the **SSVEP** dataset. The **best** result is highlighted in bold and the <u>second best</u> is underlined. Overall we achieve an average increase in performance of 8.08% over the current state-of-the-art for EEG classification models.

Subject	**Inception**	**MAtt**	**RoGra**
1	<u>80.40</u> ± 2.06	**81.60** ± 2.87	68.01 ± 1.52
2	<u>86.60</u> ± 1.62	**89.40** ± 1.36	71.05 ± 1.74
3	<u>61.60</u> ± 3.07	58.20 ± 5.64	**67.01** ± 0.64
4	25.00 ± 4.00	20.60 ± 3.88	**66.12** ± 2.15
5	25.00 ± 6.72	26.40 ± 4.80	**62.34** ± 3.51
6	79.20 ± 1.72	79.00 ± 2.68	**81.24** ± 1.63
7	69.20 ± 1.72	66.00 ± 2.19	**69.94** ± 1.36
8	23.60 ± 1.74	23.80 ± 2.71	**61.28** ± 1.28
9	<u>79.40</u> ± 2.58	**88.20** ± 2.04	74.62 ± 1.78
10	68.60 ± 3.72	70.60 ± 4.54	**71.36** ± 2.46
11	<u>91.20</u> ± 2.48	90.20 ± 1.47	**91.74** ± 1.67

Overall *RoGra* performs 18.78% better on average than its competitors while also exhibiting strong robustness. We use the same set of hyperparameters across all datasets, whereas the other methods are highly susceptible to the choice of the hyperparameter search space. Furthermore, *RoGra* also has a lower standard deviation compared to the other EEG classification models. A notable exception is InceptionJoint, which employs a subject-conditional evaluation protocol. However, even with the additional training data and static features containing subject information, our model performs significantly better across all datasets.

5.4 Time Series Classification Results

Additionally, we evaluate *RoGra* in the broader domain of Time Series Classification on five multivariate UCR (Dau et al., 2019) datasets against five TSC baselines: SimTSC (Zha et al., 2022), HIVE-COTE2 (Middlehurst et al., 2021), Hydra-MR (Tan et al., 2022), and H-InceptionTime (Ismail-Fawaz et al., 2022). Our model, *RoGra*, also outperforms all other baselines in four out of the five datasets as shown in Table 4. The improvements are statistically significant for the ECG5000 and Handwriting datasets, as determined by a t-test. While we achieve an average performance increase of approximately 2%, the gains are much lower compared to the EEG datasets. Although the addition of the DTW and residual GCN compared to the vanilla H-InceptionTime represents an improvement, the encoding with DTW is especially valuable for noisy datasets, such as those in the application of EEG.

5.5 *RoGra*: Empirical Analysis

To further analyze the performance of *RoGra*, we conducted a series of experiements: (I) we investigated the impact of individual model components on overall performance, (II) we examined the

Table 4: Performance comparison for multivariate TSC datasets. The **best** result is highlighted in bold and the second best is underlined. Overall we achieve an average increase in performance of 2% over the current state-of-the-art for time series classification models.

Datasets	**ECG5000**	**ElectricDevices**	**CharTraj**	**Handwriting**	**PhonemeSpectra**
SimTSC	94.13 ± 0.2	70.33 ± 1.2	98.89 ± 0.1	56.65 ± 1.3	30.54 ± 0.6
HIVE-COTE2	94.58 ± 0.2	74.70 ± 0.7	**99.28** ± 0.1	57.56 ± 0.6	31.46 ± 1.1
Hydra-MR	94.60 ± 0.1	73.93 ± 0.5	99.19 ± 0.2	56.09 ± 0.9	31.93 ± 0.9
H-InceptionTime	94.09 ± 0.2	71.74 ± 1.0	98.95 ± 0.2	56.50 ± 1.9	32.01 ± 1.0
RoGra (ours)	**95.26** ± 0.3	**75.08** ± 0.7	<u>99.09</u> ± 0.2	**58.49** ± 0.6	**32.72** ± 0.7

Table 5: Ablation study for different similarity measures, backbones, and GNNs and their relative impact on accuracy on the MI dataset.

Similarity	**Backbone**	**GNN**	**(MI) Performance**	**Impact**
DTW	**+ Inception**	**+ ResGCN**	**92.09** ± 1.45	**RoGra**
Euclidean	+ Inception	+ ResGCN	63.51 ± 7.53	-31.06%
Euclidean	+ Inception	+ GCN	52.36 ± 8.46	-43.14%
DTW	+ ResNet	+ ResGCN	91.54 ± 2.81	-0.60%
DTW	+ ResNet	+ GCN	88.17 ± 2.87	-4.25%
DTW	+ 1NN	+ None	22.41 ± 9.34	-75.67%
None	+ Inception	+ None	65.56 ± 5.12	-28.81%

robustness of *RoGra* compared to *MAtt* under noisy input conditions, and (III) we demonstrated the denoising effect of the Graph Neural Network (GNN) through output smoothing.

For the first study, we compared the full *RoGra* architecture by systematically replacing its components, including the similarity measure, backbone, and GNN. Our findings, reported in Table 5, indicate that the primary factor driving performance improvement is the use of the DTW similarity measure. The model's predictive capability significantly diminishes when this measure is replaced with Euclidean distance. However, DTW alone is insufficient for accurate prediction, as demonstrated by replacing the backbone and GCN with a simple 1-Nearest-Neighbor approach, which causes accuracy to drop to near-random levels. DTW, when combined with ResNet and the GCN, resembles the SimTSC architecture but uses the entire training set. However, we managed to outperform that with more than 4%. When comparing ResNet with Inception, the latter performs better in EEG classification tasks but, when used alone without additional components, also shows a decline in model accuracy.

Table 6: Impact of noise on *RoGra* and MAtt for the **MI** Dataset. Below each model, we list the degradation of performance in percent.

Model	no Noise	$\gamma = 0.05$	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$
RoGra (ours)	92.09 ± 1.45	92.33 ± 1.30	91.73 ± 1.31	91.64 ± 1.37	91.39 ± 1.52
	-	-0.26%	0.39%	0.49%	0.76%
MAtt	74.71 ± 5.01	73.29 ± 4.95	70.92 ± 5.34	66.82 ± 5.73	61.79 ± 15.30
	-	1.90%	5.07%	10.56%	17.29%

Secondly, we demonstrate the robustness of *RoGra*. In this ablation, we evaluate the model's performance under varying levels of noise in the MI dataset and analyze the resulting degradation. We have plotted the EEG sequences with the noise levels in Figure 2 and for higher values of γ in Figure 3 in the Appendix. The noise is drawn from the following distribution and applied to the full dataset in an additive fashion:

$$\text{noise} = \mathcal{N}(0,1)\text{std}(D^{\text{train}})\gamma$$

Where $\text{std}(D^{\text{train}})$ represents the standard deviation of the training data, and γ denotes the noise level. The mean of the added noise is zero, simulating the behavior of EEG data, which typically

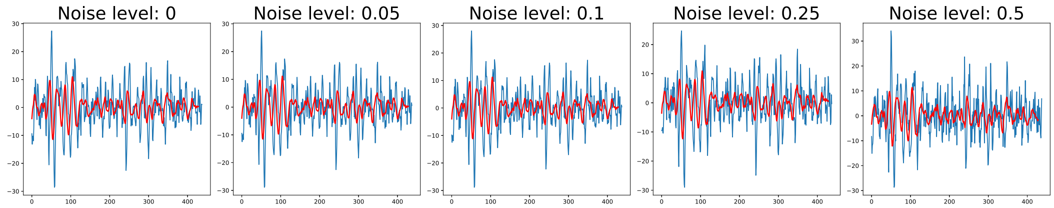

Figure 2: In this figure we show the effect of the additive noise on the EEG sequence for the first channel of MI. We also show the moving average with a window size of 10 in red to highlight the trend. The figure is best viewed in color.

Table 7: Ablation study for Kullback–Leibler divergence (KL) metric between instances within same class in **MI** dataset in comparison to the best-performing baseline **MAtt**.

Model	**Class I**	**Class II**	**Class III**	**Class IV**
MAtt	0.0468	0.0994	0.0366	0.0246
RoGra	0.0011	0.0010	0.0016	0.0008

has a mean of zero and oscillates around this point. For **MI**, the oscillation ranges approximately from -50 to 50, with a standard deviation of 6. Thus, we modify the magnitude of the oscillations and peaks in the data with additive noise. For a value of $\gamma = 0.5$ the trend of the data is identifiable, therefore a model should be able to extract meaningful features. As shown in Table 6, our model, *RoGra*, is an order of magnitude less sensitive to noisy data compared to *MAtt*, which performance degrades significantly. This demonstrates that *RoGra* is still able to extract meaningful patterns even when the peaks and valleys in the data are distorted. For EEG classification tasks, this high resistance to noise is particularly valuable, as the data is inherently noisy due to ocular and myogenic artifacts, as well as misaligned EEG sensors.

Lastly, we demonstrate that *RoGra* also generalizes effectively in the output space, as shown by the Kullback–Leibler (KL) divergence analysis on the MI dataset. For this experiment, we compute the KL divergence between the output distributions of instances within the same class, using the softmax of the final model output. In Table 7, we compare the KL divergence values for each class between *RoGra* and the best performing baseline, *MAtt*. *RoGra* exhibits significantly lower KL divergence across all classes compared to *MAtt*. This lower divergence indicates that the output distributions produced by *RoGra* are much more similar within the same class, suggesting that the model achieves more consistent and well-regularized predictions. As a result, the decision boundaries between different classes are better defined, leading to improved classification performance on EEG data in the MI dataset. Furthermore, we also compute the Shannon Divergence for the output distributions, which can be found in Appendix (Table 12). These results further support that *RoGra* effectively minimizes uncertainty in its predictions, enhancing its robustness for EEG classification tasks.

6 Conclusion

In this paper, we introduce a novel approach to EEG classification by representing EEG sequences as a graph, transforming the task into a node classification problem. Our proposed method, *RoGra*, leverages the denoising capabilities of GNNs to improve classification accuracy and robustness in noisy environments. By integrating an InceptionTime module for high-level temporal feature extraction and refining these features through a residual GNN layer with a similarity-based adjacency matrix, *RoGra* captures functional dependencies between EEG instances. This design introduces a valuable inductive bias, enabling the model to perform consistently across various EEG datasets, achieving up to a notable 25% performance improvement. Additionally, our method demonstrates strong resilience to noise, maintaining classification accuracy even when the data contains significant noise levels. These results underscore the potential of GNNs in enhancing EEG classification, particularly in challenging real-world conditions where noise is prevalent.

7 Reproducibility Statement

We are committed to ensuring the reproducibility of our model, *RoGra*. To support this, we provide the source code, including pre-processing steps, model architecture, and training scripts, in the supplementary material. Upon publication, the code will also be made publicly available on GitHub, along with detailed documentation to guide experiment setup and execution. The datasets used in our experiments are publicly available, and we additionally provide a link to the preprocessed data in a cloud service for easy access. All hyperparameters and model configurations are detailed in both the paper and the code repository to ensure easy replication. The computing environment, including hardware specifications, software dependencies, and package versions, is fully documented, with a "requirements.txt" file provided to facilitate seamless environment setup.

8 Ethics Statement

We are committed to contributing positively to society and human well-being, while respecting the privacy of the subjects involved in our evaluation. Our experiments utilize three established EEG datasets: MI, SSVEP, and ERN, which involve data from human subjects. These datasets are fully anonymized and do not contain any personally identifiable information. Additionally, they are publicly available and widely used within the machine learning community.

References

Ghadir Ali Altuwaijri and Ghulam Muhammad. A multibranch of convolutional neural network models for electroencephalogram-based motor imagery classification. *Biosensors*, 12(1):22, 2022.

Ghadir Ali Altuwaijri, Ghulam Muhammad, Hamdi Altaheri, and Mansour Alsulaiman. A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for eeg-based motor imagery signals classification. *Diagnostics*, 12(4):995, 2022.

Zaid Abdi Alkareem Alyasseri, Ahamad Tajudin Khader, Mohammed Azmi Al-Betar, Ammar Kamal Abasi, and Sharif Naser Makhadmeh. Eeg signals denoising using optimal wavelet transform hybridized with efficient metaheuristic methods. *IEEE Access*, 8:10584–10605, 2019.

Sachin Borse. Eeg de-noising using wavelet transform and fast ica. *IJISET-International Journal of Innovative Science Engineering & Technology*, 2(7):200–205, 2015.

Xavier Bresson and Thomas Laurent. Residual gated graph convnets. *arXiv preprint arXiv:1711.07553*, 2017.

Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. Bci competition 2008–graz data set a. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16:1–6, 2008.

Johannes Burchert, Thorben Werner, Vijaya Krishna Yalavarthi, Diego Coello de Portugal, Maximilian Stubbemann, and Lars Schmidt-Thieme. Are eeg sequences time series? eeg classification with time series models and joint subject training. *arXiv preprint arXiv:2404.06966*, 2024.

Hsiao-Lung Chan, Yu-Tai Tsai, Ling-Fu Meng, and Tony Wu. The removal of ocular artifacts from eeg signals using adaptive filters based on ocular source components. *Annals of Biomedical Engineering*, 38:3489–3499, 2010.

Ziqiang Cheng, Yang Yang, Wei Wang, Wenjie Hu, Yueting Zhuang, and Guojie Song. Time2graph: Revisiting time series modeling with dynamic shapelets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3617–3624, 2020.

Ziqiang Cheng, Yang Yang, Shuo Jiang, Wenjie Hu, Zhangchi Ying, Ziwei Chai, and Chunping Wang. Time2graph+: Bridging time series and graph representation learning via multiple attentions. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):2078–2090, 2021.

Rodney J Croft and Robert J Barry. Removal of ocular artifact from the eeg: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(1):5–19, 2000.

Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

Arnaud Delorme. Eeg is better left alone. *Scientific Reports*, 13(1):2372, 2023.

Emadeldeen Eldele, Zhenghua Chen, Chengyu Liu, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:809–818, 2021.

Maximilian Grobbelaar, Souvik Phadikar, Ebrahim Ghaderpour, Aaron F Struck, Nidul Sinha, Rajdeep Ghosh, and Md Zaved Iqubal Ahmed. A survey on denoising techniques of electroencephalogram signals using wavelet transform. *Signals*, 3(3):577–586, 2022.

Thi Kieu Khanh Ho and Narges Armanfard. Self-supervised learning for anomalous channel detection in eeg graphs: Application to seizure analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7866–7874, 2023.

Yimin Hou, Shuyue Jia, Xiangmin Lun, Shu Zhang, Tao Chen, Fang Wang, and Jinglei Lv. Deep feature mining via attention-based bilstm-gcn for human motor imagery recognition. *arXiv preprint arXiv:2005.00777*, 2020.

Thorir Mar Ingolfsson, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini. Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2958–2965. IEEE, 2020.

Ali Ismail-Fawaz, Maxime Devanne, Jonathan Weber, and Germain Forestier. Deep learning for time series classification using new hand-crafted convolution filters. In *2022 IEEE International Conference on Big Data (Big Data)*, pp. 972–981. IEEE, 2022.

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F Schmidt, Jonathan Weber, Geoffrey I Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020.

Ziyu Jia, Youfang Lin, Jing Wang, Xiaojun Ning, Yuanlai He, Ronghao Zhou, Yuhan Zhou, and H Lehman Li-wei. Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1977–1986, 2021.

Jing Jin, Hao Sun, Ian Daly, Shurui Li, Chang Liu, Xingyu Wang, and Andrzej Cichocki. A novel classification framework using the graph representations of electroencephalogram for motor imagery based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:20–29, 2021.

Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I Webb, Irwin King, and Shirui Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Don H Johnson. Signal-to-noise ratio. *Scholarpedia*, 1(12):2088, 2006.

Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. Ecg heartbeat classification: A deep transferable representation. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 443–444. IEEE, 2018.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2022.

Dominik Klepl, Fei He, Min Wu, Daniel J Blackburn, and Ptolemaios Sarrigiannis. Eeg-based graph neural network classification of alzheimer's disease: An empirical evaluation of functional connectivity methods. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30: 2651–2660, 2022.

Dominik Klepl, Min Wu, and Fei He. Graph neural network-based eeg classification: A survey. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

Shailaja Kotte and JRK Kumar Dabbakuti. Methods for removal of artifacts from eeg signal: A review. In *Journal of Physics: Conference Series*, volume 1706, pp. 012093. IOP Publishing, 2020.

Chi Qin Lai, Haidi Ibrahim, Mohd Zaid Abdullah, Jafri Malin Abdullah, Shahrel Azmin Suandi, and Azlinda Azman. Artifacts and noise removal for electroencephalogram (eeg): A literature review. In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp. 326–332. IEEE, 2018.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

Zhiqiang Lin, Taorong Qiu, Ping Liu, Lingyun Zhang, Siwei Zhang, and Zhendong Mu. Fatigue driving recognition based on deep learning and graph neural network. *Biomedical Signal Processing and Control*, 68:102598, 2021.

Huaiyuan Liu, Donghua Yang, Xianzhang Liu, Xinglei Chen, Zhiyu Liang, Hongzhi Wang, Yong Cui, and Jun Gu. Todynet: temporal dynamic graph neural network for multivariate time series classification. *Information Sciences*, pp. 120914, 2024.

Xiaorui Liu, Jiayuan Ding, Wei Jin, Han Xu, Yao Ma, Zitao Liu, and Jiliang Tang. Graph neural networks with adaptive residual. *Advances in Neural Information Processing Systems*, 34:9720–9733, 2021.

Yao Ma, Xiaorui Liu, Tong Zhao, Yozen Liu, Jiliang Tang, and Neil Shah. A unified view on graph neural networks as graph signal denoising. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 1202–1211, 2021.

Ravikiran Mane, Effie Chew, Karen Chua, Kai Keng Ang, Neethu Robinson, A Prasad Vinod, Seong-Whan Lee, and Cuntai Guan. Fbcnet: A multi-view convolutional neural network for brain-computer interface. *arXiv preprint arXiv:2104.01233*, 2021.

Perrin Margaux, Maby Emmanuel, Daligault Sébastien, Bertrand Olivier, and Mattout Jérémie. Objective and subjective evaluation of online error correction during p300-based spelling. *Advances in Human-Computer Interaction*, 2012:4–4, 2012.

Brenton W McMenamin, Alexander J Shackman, Jeffrey S Maxwell, David RW Bachhuber, Adam M Koppenhaver, Lawrence L Greischar, and Richard J Davidson. Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. *Neuroimage*, 49(3):2416–2432, 2010.

Matthew Middlehurst, James Large, Michael Flynn, Jason Lines, Aaron Bostrom, and Anthony Bagnall. Hive-cote 2.0: a new meta ensemble for time series classification. *Machine Learning*, 110(11):3211–3243, 2021.

Yazeed K Musallam, Nasser I AlFassam, Ghulam Muhammad, Syed Umar Amin, Mansour Alsulaiman, Wadood Abdul, Hamdi Altaheri, Mohamed A Bencherif, and Mohammed Algabri. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. *Biomedical Signal Processing and Control*, 69:102826, 2021.

J Nekrasova, O Bazanova, D Shunenkov, M Kanarskiy, I Borisov, and E Luginina. Problem of myogenic contamination in electroencephalography. *Human Physiology*, 48(4):470–482, 2022.

Spiros Nikolopoulos. Mamem eeg ssvep dataset ii (256 channels, 11 subjects, 5 frequencies presented simultaneously). 2021.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg decoding. *Advances in Neural Information Processing Systems*, 35:31116–31129, 2022.

Debanjan Parbat and Monisha Chakraborty. A novel methodology to study the cognitive load induced eeg complexity changes: Chaos, fractal and entropy based approach. *Biomedical Signal Processing and Control*, 64:102277, 2021.

Darshana Priyasad, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Affect recognition from scalp-eeg using channel-wise encoder networks coupled with geometric deep learning and multi-channel feature fusion. *Knowledge-Based Systems*, 250:109038, 2022.

Khadijeh Raeisi, Mohammad Khazaei, Pierpaolo Croce, Gabriella Tamburro, Silvia Comani, and Filippo Zappasodi. A graph convolutional neural network for the automated detection of seizures in the neonatal eeg. *Computer methods and programs in biomedicine*, 222:106950, 2022.

Mouad Riyad, Mohammed Khalil, and Abdellah Adib. Incep-eegnet: a convnet for motor imagery decoding. In *Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings 9*, pp. 103–111. Springer, 2020.

Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, 1978.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017.

Andres Soler, Pablo A Muñoz-Gutiérrez, Maximiliano Bueno-López, Eduardo Giraldo, and Marta Molinas. Low-density eeg for neural activity reconstruction using multivariate empirical mode decomposition. *Frontiers in Neuroscience*, 14:175, 2020.

Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3): 532–541, 2018.

Xinlin Sun, Chao Ma, Peiyin Chen, Mengyu Li, He Wang, Weidong Dang, Chaoxu Mu, and Zhongke Gao. A novel complex network-based graph convolutional network in major depressive disorder detection. *IEEE Transactions on Instrumentation and Measurement*, 71:1–8, 2022.

Luay Yassin Taha and Esam Abdel-Raheem. Blind source separation: A performance review approach. In *2022 5th International Conference on Signal Processing and Information Security (ICSPIS)*, pp. 148–153. IEEE, 2022.

Chang Wei Tan, Angus Dempster, Christoph Bergmeir, and Geoffrey I Webb. Multirocket: multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery*, 36(5):1623–1646, 2022.

Siyi Tang, Jared A Dunnmon, Khaled Saab, Xuan Zhang, Qianying Huang, Florian Dubost, Daniel L Rubin, and Christopher Lee-Messer. Self-supervised graph neural networks for improved electroencephalographic seizure analysis. *arXiv preprint arXiv:2104.08336*, 2021.

Jialin Wang, Rui Gao, Haotian Zheng, Hao Zhu, and C-J Richard Shi. Ssgcnet: A sparse spectra graph convolutional network for epileptic eeg signal classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

Chun-Shu Wei, Toshiaki Koike-Akino, and Ye Wang. Spatial component-wise convolutional network (sccnet) for motor-imagery eeg classification. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 328–331. IEEE, 2019.

Raneen Younis, Abdul Hakmeh, and Zahra Ahmadi. Mts2graph: Interpretable multivariate time series classification with temporal evolving graphs. *Pattern Recognition*, 152:110486, 2024.

Daochen Zha, Kwei-Herng Lai, Kaixiong Zhou, and Xia Hu. Towards similarity-aware time-series classification. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pp. 199–207. SIAM, 2022.

Haoming Zhang, Mingqi Zhao, Chen Wei, Dante Mantini, Zherui Li, and Quanying Liu. Eegdenoisenet: a benchmark dataset for deep learning solutions of eeg denoising. *Journal of Neural Engineering*, 18(5):056057, 2021.

Peixiang Zhong, Di Wang, and Chunyan Miao. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2020.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020.

Yijin Zhou, Fu Li, Yang Li, Youshuo Ji, Guangming Shi, Wenming Zheng, Lijian Zhang, Yuanfang Chen, and Rui Cheng. Progressive graph convolution network for eeg emotion recognition. *Neurocomputing*, 544:126262, 2023.

A APPENDIX

A.1 INCEPTIONTIME

As our backbone for *RoGra*, we use the InceptionTime (Ismail Fawaz et al., 2020) model, which consists of multiple stacked InceptionTimeBlocks. In Algorithm 2, we show the layout of such a block. Here, 1D convolutions are used with different kernel sizes to extract features and produce latent embeddings at different resolutions. Afterwards, BatchNorm and ReLU, as non-linear activation functions, are applied. Additionally, in C_4, InceptionTime applies MaxPooling to further regularize the embeddings. In the final step of the InceptionTimeBlock, the latent embeddings are concatenated and returned.

Algorithm 2 InceptionTime Block

```
1:  Input: Input time series data X ∈ ℝ^{C×T}       ▷ C: number of channels, T: sequence length
2:  Output: Output feature map Z^0
3:
4:  procedure INCEPTIONTIMEBLOCK(X)
5:      C_1 ← ReLU(BatchNorm(Conv1D(X)))              ▷ Convolution with kernel size 1
6:      C_2 ← ReLU(BatchNorm(Conv1D(X)))              ▷ Done twice with kernel size 1 and 3
7:      C_3 ← ReLU(BatchNorm(Conv1D(X)))              ▷ Done twice with kernel size 1 and 5
8:      C_4 ← ReLU(BatchNorm(Conv1D(MaxPool(X))))     ▷ Convolution with kernel size 1
9:      Z^0 ← Concatenate([C_1, C_2, C_3, C_4], axis=-1)
10:     return Z^0
```

A.2 EVALUATION METRICS

In addition to accuracy for the first two datasets, **MI** and **SSVEP**, we use Area Under the Curve (AUC) as the evaluation metric for the last dataset, **ERN**. This particular dataset is unbalanced, with a ratio of 70/30 for the positive and negative classes, respectively, which makes accuracy unsuitable; therefore, AUC is applied. AUC measures how well a binary classification model can distinguish between two classes. Ranging from 0 to 1, a higher AUC indicates better performance, with 0.5 representing random guessing and 1 indicating perfect classification. The exact calculation is described in the equation below.

$$\text{AUC} = \int_{x=0}^{1} TPR(x)\, d(FPR(x))$$

where $TPR(x)$ is the True Positive Rate (sensitivity) at a given false positive rate $FPR(x)$.

A.3 ADDITIONAL RESULTS AND ABLATION STUDIES

Here we have compiled additional experimental results and further evidence for the ablation studies. For the main results of this paper, we have included the performance for the ERN dataset in Table 8. Here, our model, *RoGra*, achieves an impressive lift of 24.98% compared to the next best baseline. We achieve very high performance for all subjects, with no noticeable decreases.

Furthermore, we compare two different splitting methods for the train and test data. Here, *Inception* and *MAtt* split the dataset timewise, where the first 300 timesteps are used for training, the next 100 for validation, and the final 100 for testing. We adopt a different protocol by splitting not time-wise by session, but instead using a split by instances. This protocol is utilized by *MAtt Inst.* and *RoGra.* The results for this ablation are shown in Table 9, where there is no significant difference in the performance of the two protocols. *MAtt Inst.* shows a higher standard deviation for some subjects, but this is caused by a lower number of runs for this protocol. The original *MAtt* has 8 runs, while we used 3 as an approximation for this study.

We also test the resilience of *RoGra* and *MAtt* against noisy input data, which is common in the domain of EEG classification. In Figure 3, we show the EEG sequences and the corresponding factor γ for the additive noise. For simplicity, we only plot the first channel of the EEG sequences as a representative sample of the data. We also show the moving average with a sliding window of 10

Table 8: Performance comparison for the **ERN** dataset. The **best** result is highlighted in bold and the second best is underlined. Overall we achieve an average increase in performance of 24.98% over the current state-of-the-art for EEG classification models.

Subject	**Inception**	**MAtt**	**RoGra**
2	74.03±3.37	81.86±2.56	**97.64** ± 1.34
6	88.40±2.67	68.34±1.76	**94.58** ± 1.08
7	85.45±7.05	69.28±10.90	**98.76** ± 1.12
11	55.27±9.19	70.50±2.66	**95.44** ± 2.54
12	68.62±9.92	60.62±4.70	**98.04** ± 1.52
13	51.15±4.16	62.92±3.65	**96.14** ± 2.13
14	75.22±3.17	70.42±4.80	**97.46** ± 1.89
16	50.53±3.54	55.71±2.56	**95.84** ± 2.28
17	75.93±4.20	80.60±5.52	**97.17** ± 1.64
18	72.30±1.39	75.11±1.42	**97.67** ± 1.73
20	59.97±2.56	57.60±8.12	**92.51** ± 2.25
21	67.95±7.76	62.43±9.56	**97.05** ± 1.84
22	95.09±0.89	89.33±4.28	**98.53** ± 1.06
23	61.66±2.38	70.03±5.40	**96.91** ± 2.02
24	70.88±4.60	73.71±2.99	**97.81** ± 1.24
26	54.03±4.66	60.08±1.92	**95.52** ± 1.25

Table 9: Performance comparison for the **SSVEP** dataset. The **best** result is highlighted in bold and the second best is underlined. Overall, we achieve an average increase in performance of 8.08% over the current state-of-the-art for EEG classification models. Here, MAtt Inst. is using our train/test split of splitting by instances instead of time.

Subject	Inception	MAtt	MAtt Inst.	RoGra
1	80.40 ± 2.06	**81.60** ± 2.87	75.00 ± 7.35	68.01 ± 1.52
2	86.60 ± 1.62	**89.40** ± 1.36	89.00 ± 2.10	71.05 ± 1.74
3	61.60 ± 3.07	58.20 ± 5.64	52.00 ± 5.90	**67.01** ± 0.64
4	25.00 ± 4.00	20.60 ± 3.88	26.40 ± 5.57	**66.12**± 2.15
5	25.00 ± 6.72	26.40 ± 4.80	27.20 ± 4.79	**62.34** ± 3.51
6	79.20 ± 1.72	79.00 ± 2.68	**85.80** ± 2.48	81.24 ± 1.63
7	69.20 ± 1.72	66.00 ± 2.19	**73.60** ± 7.39	69.94 ± 1.36
8	23.60 ± 1.74	23.80 ± 2.71	22.20 ± 3.19	**61.28** ± 1.28
9	79.40 ± 2.58	88.20 ± 2.04	**90.60** ± 4.96	74.62 ± 1.78
10	68.60 ± 3.72	70.60 ± 4.54	68.20 ± 4.58	**71.36** ± 2.46
11	91.20 ± 2.48	90.20 ± 1.47	84.40 ± 6.83	**91.74** ± 1.67
Summary	62.71	63.09	63.13	**71.34**

in red to better visualize the trends in the data. For a noise level ranging from 0 to 0.5, the patterns, while distorted, remain clearly identifiable, with the peaks and valleys largely unchanged. When applying a $\gamma \geq 1$, the signal begins to deteriorate until it is mostly noise in the case of $\gamma = 100$.

Additionally, we present the impact of noise on the performance of each individual subject for *MAtt* in Table 10 and *RoGra* in Table 11. Here, the same trend of monotonically decreasing performance for larger values of the noise-scaling factor γ can be observed.

For the study of the output distribution of the softmax on the final latent representation of the model, we also use the Jensen-Shannon Divergence. This metric is also based on the KL divergence which takes the following form for two input distributions P and Q:

$$KL(P||Q) = \sum_x P(x) \log(\frac{P(x)}{Q(x)})$$

The Jensen-Shannon Divergence is defined as:

$$JS(P||Q) = \frac{1}{2} KL(P|| \frac{(P+Q)}{2}) + \frac{1}{2} KL(Q|| \frac{(P+Q)}{2})$$

In Table 12 we show the Jensen-Shannon Divergence for each class for *RoGra* and *MAtt* respectively. Like in the case of the KL divergence the Jensen-Shannon similarity measure shows that *RoGra* is well regularized for the output space.

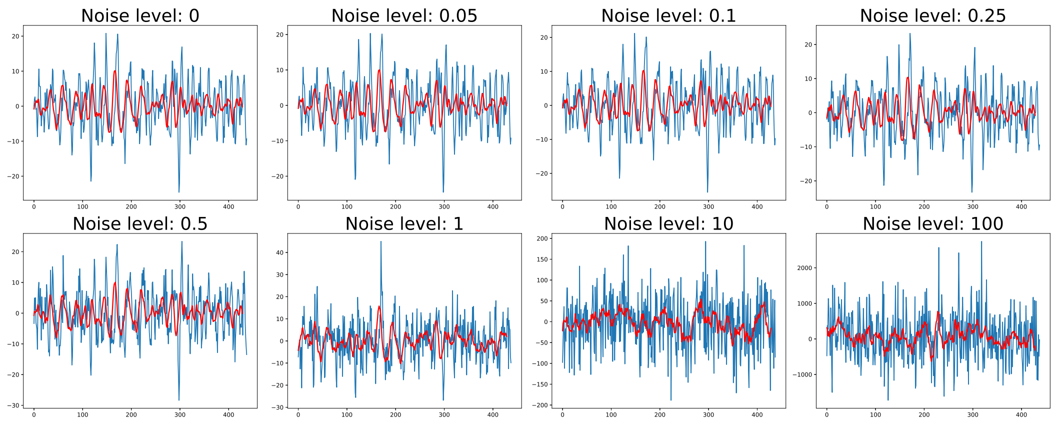

Figure 3: In this figure we show the effect of the additive noise on the EEG sequence for the first channel of MI. We also show the moving average with a window size of 10 in red to highlight the trend. The figure is best viewed in color.

Table 10: Impact of noise on MAtt

Subject	$\gamma = 0.05$	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$
1	86.17 ± 1.67	83.45 ± 2.20	77.31 ± 1.07	70.02 ± 0.91
2	57.81 ± 2.45	55.67 ± 1.07	52.08 ± 3.69	44.68 ± 13.10
3	86.57 ± 1.00	80.67 ± 0.16	78.59 ± 1.61	76.62 ± 2.95
4	67.01 ± 1.01	65.16 ± 3.38	57.87 ± 0.71	51.50 ± 2.09
5	57.52 ± 1.93	50.93 ± 2.36	51.39 ± 0.57	49.19 ± 5.23
6	52.89 ± 1.56	49.07 ± 0.43	47.11 ± 2.57	45.14 ± 0.98
7	88.77 ± 1.93	87.85 ± 1.77	79.17 ± 2.55	68.98 ± 4.05
8	82.06 ± 1.18	83.91 ± 0.71	79.17 ± 1.24	73.38 ± 1.18
9	80.79 ± 1.56	81.60 ± 1.30	78.70 ± 0.16	76.62 ± 1.56
Summary	**73.29 ± 4.95**	**70.92 ± 5.34**	**66.82 ± 5.73**	**61.79 ± 15.30**

Table 11: Impact of noise on RoGra

Subject	$\gamma = 0.05$	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$
1	94.56 ± 1.82	94.24 ± 1.93	94.21 ± 1.86	93.13 ± 2.11
2	77.03 ± 1.17	76.91 ± 1.08	76.73 ± 1.14	76.33 ± 1.44
3	98.13 ± 1.47	98.02 ± 1.57	97.93 ± 1.45	97.67 ± 1.38
4	91.22 ± 1.31	91.04 ± 1.17	91.01 ± 1.23	90.89 ± 1.34
5	92.51 ± 1.09	92.64 ± 1.81	92.46 ± 1.21	92.33 ± 1.48
6	84.92 ± 1.86	84.53 ± 2.13	84.51 ± 2.19	84.47 ± 2.52
7	97.91 ± 0.72	97.82 ± 0.78	97.71 ± 0.83	97.59 ± 1.04
8	96.15 ± 1.06	96.01 ± 1.12	95.92 ± 1.15	95.84 ± 1.12
9	94.66 ± 1.23	94.37 ± 1.31	94.31 ± 1.34	94.28 ± 1.29
Summary	**92.33 ± 1.30**	**91.73 ± 1.31**	**91.64 ± 1.37**	**91.39 ± 1.52**

Table 12: Ablation study for Shannon Divergence metric between instances within the same class in **MI** dataset in comparison to the best-performing baseline **MAtt**.

Model	Class I	Class II	Class III	Class IV
MAtt	0.0221	0.0176	0.0113	0.0163
RoGra	0.0062	0.0033	0.0042	0.0044