OPEN ACCESS
International Journal of Neural Systems, Vol. 33, No. 9 (2023) 2350046 (16 pages)

DOI: 10.1142/S0129065723500466

World Scientific
www.worldscientific.com

A Class-Imbalance Aware and Explainable Spatio-Temporal Graph Attention Network for Neonatal Seizure Detection

Khadijeh Raeisi* and Mohammad Khazaei†
Department of Neuroscience, Imaging and Clinical Sciences
Universita Gabriele d'Annunzio
Chieti 66100, Italy
**khadijeh.raeisi@unich.it*
†*mohammad.khazaei@unich.it*

Gabriella Tamburro‡, Pierpaolo Croce§ and Silvia Comani¶
Department of Neuroscience, Imaging and
Clinical Sciences-Behavioral Imaging and Neural Dynamics Center
Universita Gabriele d'Annunzio
Chieti 66100, Italy
‡*g.tamburro@unich.it*
§*pierpaolo.croce@unich.it*
¶*silvia.comani@unich.it*

Filippo Zappasodi‖
Department of Neuroscience, Imaging and Clinical Sciences-Behavioral
Imaging and Neural Dynamics Center-Institute for
Advanced Biomedical Technologies
Universita Gabriele d'Annunzio
Chieti 66100, Italy
‖*filippo.zappasodi@unich.it*

Accepted 19 June 2023
Published Online 28 July 2023

Seizures are the most prevalent clinical indication of neurological disorders in neonates. In this study, a class-imbalance aware and explainable deep learning approach based on Convolutional Neural Networks (CNNs) and Graph Attention Networks (GATs) is proposed for the accurate automated detection of neonatal seizures. The proposed model integrates the temporal information of EEG signals with the spatial information on the EEG channels through the graph representation of the multi-channel EEG segments. One-dimensional CNNs are used to automatically develop a feature set that accurately represents the differences between seizure and nonseizure epochs in the time domain. By employing GAT, the attention mechanism is utilized to emphasize the critical channel pairs and information flow among brain regions. GAT coefficients were then used to empirically visualize the important regions during the seizure and nonseizure epochs, which can provide valuable insight into the location of seizures in the neonatal brain. Additionally, to tackle the severe class imbalance in the neonatal seizure dataset using under-sampling and focal loss techniques are used. Overall, the final Spatio-Temporal Graph Attention Network (ST-GAT) outperformed previous

‖Corresponding author.

benchmarked methods with a mean AUC of 96.6% and Kappa of 0.88, demonstrating its high accuracy and potential for clinical applications.

Keywords: Electroencephalography (EEG); neonatal seizure detection; attention mechanism; focal loss; Convolutional Neural Networks (CNN); Graph Attention Networks (GAT).

1. Introduction

Seizures are the most common neurological emergency in newborns and are linked to poor neurodevelopmental outcomes in hypoxic-ischemic encephalopathy patients[1] as well as in other critical care situations.[2] Hence, the effective detection and diagnosis of seizures are crucial for reducing the risk of long-term adverse consequences. Electroencephalography (EEG) provides a noninvasive tool for monitoring the cortical activity of the neonatal brain and, thus, for the prediction or identification of seizures in adults and neonates.[3–16] Nonetheless, seizure detection on the neonatal EEG — typically performed by visual inspection of the EEG traces — can be very challenging because the interpretation of the continuous EEG must be performed by an on-site highly trained expert in the Neonatal Intensive Care Unit (NICU) to be reliable.[17] Therefore, seizure detection on neonatal EEG is highly laborious, time-consuming, and subject to unavoidable personal interpretation.

Under these circumstances, several algorithms have been proposed to automate the detection of neonatal seizures from multi-channel EEG signals and to support clinical decision-making. Some existing methods employ heuristic rules or complex features to extract temporal information from the EEG recordings over a time span and combine them with machine learning algorithms to detect seizures.[18–25] These hand-engineered and sometimes physiologically meaningful rules and features are expected to capture a wide range of EEG dynamics characteristics associated with seizures in order to achieve proper classification performances.

As a promising alternative to classical machine learning approaches, deep learning methods, and more specifically, Convolutional Neural Networks (CNNs), have demonstrated state-of-the-art results in several research fields[26–29] including neonatal seizure detection.[9,10,30,11,31] A hybrid architecture was developed by Ansari *et al.*,[10]: in the first step, the feature extraction was done within a CNN and then the seizures were classified using a decision tree. This hybrid method improved the classification performance of previous methods, achieving an area under the curve (AUC) of 88%. O'Shea *et al.*,[11] proposed two deep learning algorithms containing 1-dimensional convolution filters. They reached an AUC of 98.5% and 95.5% for seizure detection, respectively, on the neonatal EEG data from Cork University and the Helsinki University public databases. However, these CNN-based approaches might not be well suited to concurrently leverage both spatial and temporal information to produce rich representations of brain activity from the raw multi-channel EEG recordings. Similarly, most published seizure detection algorithms work on independent single-channel EEG recordings, ignoring the valuable spatial information that can be extracted from the topological organization of the EEG channels. If we consider the brain as a network in which spatially segregated areas are synchronously co-activated,[32] information on the EEG channels that aggregate and share multiple sources of neural information continuously can be of great value to identify events such as seizures.[33] A graph representation can be employed to model the spatial dependencies among the EEG channels, and these spatial features should be thoroughly investigated to develop an ideal seizure classifier. In this regard, Graph Neural Networks (GNNs) would be a suitable method.[34] We recently developed a model based on GNN for the automated detection of neonatal seizures,[35] where the temporal information, including features in both time and frequency domains, was embedded as node features in the graph representation of the EEG signal epochs. The spatial information was represented as the functional connectivity among the EEG channels or as maps of Euclidean distances. On a publicly available Helsinki dataset, the model performance achieved a median AUC of 99.1% and a median AUC for specificity values greater than 90% (AUC90) of 96%.

Advances in generalizing CNN to the graph domain can be categorized as spectral[34,36] and spatial[37] (nonspectral) approaches. Spectral approaches work with a spectral representation of the graphs, which needs potentially intense computations and depends on the graph structure. Spatial methods, on the other hand, define convolutions directly on the graph, operating on groups of spatially close neighbors, and therefore do not rely on the graph structure.

Recently, attention mechanisms have gained popularity in many sequence-based tasks.[38] Inspired by these mechanisms, Graph ATtention networks (GATs) were introduced by Petar *et al.*[39] to perform node classification of graph-structured data. The idea of this spatial approach is to compute each node's representation by aggregating the representations from its neighboring nodes through an attention strategy. By doing so, the more relevant information flow among the nodes of a graph would be highlighted to perform a specific classification task. Moreover, this model is directly applicable to tasks where the model has to generalize to completely unseen input data and graph structure, making it appealing for developing generalizable EEG classifiers. More importantly, GAT is explainable, meaning that it provides an effective way to describe and visualize the intrinsic relationship between neighboring graph nodes using the learned attention coefficients. This new approach offers the possibility to both explore the relationships among the multiple EEG channels during neonatal seizures and highlight the brain regions with a higher probability of seizure activity.

Apart from the model's architecture when addressing the seizure detection task, one common problem that should also be considered is the severe class imbalance. Class imbalance refers to a situation in which the number of samples of one class (e.g. seizure) is much smaller than the other class (e.g. nonseizure). This can be challenging in training any classifier because algorithms trained on imbalanced datasets tend to perform poorly on the minority class, which is usually the class of interest in seizure detection. To address the class imbalance problem, various techniques such as oversampling and undersampling, or using cost-sensitive learning techniques such as weighted loss functions and focal loss can be used. Several works have applied these techniques to adult seizure detection task from EEG datasets,[5,40,41] but so far, no research group has worked on this issue for neonatal seizure detection.

Motivated by the aforementioned issues, in this study, a class imbalance-aware and explainable Spatio-Temporal Graph Attention Network (ST-GAT) is developed for neonatal seizure detection. First, a 1-D CNN is used for the temporal feature extraction step. By doing so, instead of using hand-crafted features, the end-to-end network automatically and repeatedly refine a feature set that accurately describes the differences between EEG temporal dynamics during seizure and nonseizure epochs.[26] Subsequently, by employing GAT, the attention mechanism in GAT is leveraged to force the model to automatically pay more attention to more essential channel pairs and information flows among brain regions.[39] By stacking GAT layers on top of the CNN layers, a ST-GAT is developed to capture both the temporal dynamics and the spatial dependencies of the neonatal EEG signals that should better differentiate between seizure and nonseizure epochs. The attention coefficients are used to explain how ST-GAT performs when classifying seizure and nonseizure segments. This classification approach can highlight the potential seizure locations in the neonatal brain and therefore it speeds up the review process by a clinical expert if needed. Moreover, a 2-step approach is employed to minimize the effect of class imbalance in the dataset: (1) in the preprocessing phase the seizure segments are under-sampled using longer overlaps during segmentation; (2) in the training phase standard cross-entropy is replaced with focal loss to put more emphasis on misclassified epochs. In summary, the following are the main contributions of our work:

- An end-to-end ST-GAT is proposed for the first time to fully exploit both the temporal dynamics and the spatial dependencies of the neonatal EEG for seizure detection.
- Extensive experiments and ablation studies are conducted to demonstrate the effectiveness of our proposed model architecture.
- The severe class imbalance between seizure and nonseizure segments is addressed with a 2-step approach.

- ST-GAT achieves superior performance over the state-of-the-art models on neonatal EEG data from the Helsinki database.
- GAT coefficients are used to explain and empirically visualize the importance of neighboring electrode pairs, which is useful to highlight potential seizure locations in the neonatal brain, hence facilitating the review process by clinical experts.

2. Materials and Methods

2.1. *Neonatal EEG dataset*

The proposed method was applied to the publicly available neonatal EEG dataset recorded at the University of Helsinki (Children's Hospital, the University of Helsinki Central Hospital, Finland) from 79 neonates.[42] EEG recordings from 39 full-term neonates that had seizures by consensus of three clinical experts were included in the present study. On average, each EEG recording in this dataset has a duration of 85 min and contains approximately nine seizures, resulting in a total of 342 consensus seizures. As only neonates with consensus seizure were involved in the study, the final dataset comprised approximately 10 h of seizure data and 53 h of nonseizure data. The neonatal EEG was recorded using a 19-channel EEG system with the international 10–20 system layout and 256 Hz sampling frequency. To be consistent with the previously published papers on the same dataset, a standard longitudinal bipolar montage was employed for further analysis (Fp2-F4, F4-C4, C4-P4, P4-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F7, F7-T3, T3-T5, T5-O1, Fz-Cz, and Cz-Pz).

2.2. *Filtering and segmentation*

The neonatal EEG recordings were filtered between 0.5 and 16 Hz using a combination of seventh-order low-pass and high-pass noncausal Butterworth filters in both forward and backward directions. Afterwards, flat lines and high amplitude fluctuations ($>$ 200 μV) that last for several seconds were removed manually. The filtered signals were down-sampled to 32 Hz to decrease the computational load. The filtered and down-sampled EEG signals were segmented into epochs of 12 s with an overlap of 11 s for seizure and 10 s for nonseizure segments to handle the class imbalance problem. In our classification task, class imbalance occurs since there are comparatively fewer seizure class segments in the Helsinki dataset than nonseizure segments. Such class imbalance leads to a bias in the model's predictions of the majority class (nonseizure class) and therefore hinders the classifier's performance. Choosing a larger overlap for seizure segmentation will address this issue to some extent.[43]

2.3. *Proposed method*

The proposed ST-GAT learning model for neonatal seizure detection is based on CNN and GAT. The overall pipeline of the proposed method for classifying the EEG epochs that contain seizure events is outlined in Fig. 1. The method is composed of the following modules:

(i) A CNN architecture for the extraction of temporal features from EEG signals;
(ii) A graph representation of EEG signals using the output of CNN;

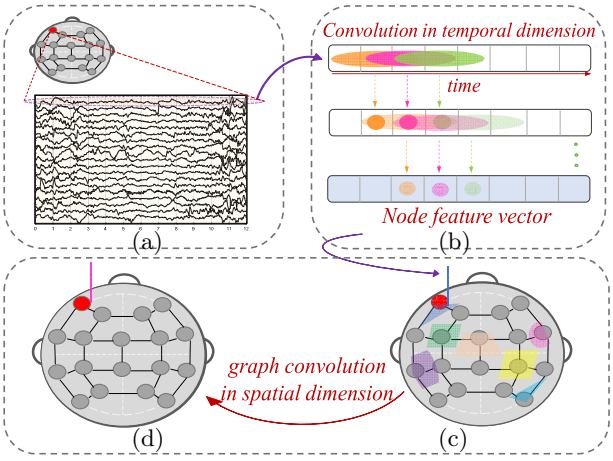


Fig. 1. The overall pipeline of the proposed ST-GAT method for neonatal seizure detection. Twelve-s EEG epochs (a) were applied to 1-D CNN to extract temporal features (b). These features were then used as graph signals (or nodes' features) to complete the graph representation of the input EEG epoch (c). Graph convolution layers were then used to explore spatial information (d). After the average pooling of the node features, two fully connected layers perform the classification task.

(iii) GAT blocks to explore the spatial information contained in the constructed graphs and extract spatial features.

Each module has been thoroughly covered in the remainder of the section.

2.3.1. *CNN architecture*

A CNN architecture was proposed for extracting the temporal features of the EEG signals and, consequently, for capturing their temporal dynamics. CNN has the advantage of fast training and simple structures.[44] Here, the output of convolutions constructs the features of each node in the EEG network (i.e. the features of the EEG signal segments at each channel) to complete the graph representation of the EEG segments. The employed CNN comprises four types of layers: convolutional, rectifying, pooling, and normalization layers. Each convolutional block contains 1-D convolutions with a kernel size of K_l (kernel size of layer l), followed by a rectified linear unit (ReLU) as the nonlinearity function[45] (please see the schematic in Table 1 in Sec. 2.4). ReLU was chosen as the activation function for several reasons: first, ReLU maintains the positive values of its input and replaces the negative values with zero. This property introduces nonlinearity into the model, allowing it to learn complex and nonlinear features from the input signal. ReLU is also computationally efficient, as it only requires a simple thresholding operation, which leads to faster convergence during training. Finally, it has been widely used in CNNs and has demonstrated strong performance in various applications.[46–48]

In each block, 1-D convolutions and ReLU layers are followed by a max-pooling layer with a stride of two for downsampling and Batch-normalization layers for the normalization of the previous layer's output. Model predictions are based on information collected from a reasonable receptive field size through four stacked CNN blocks.

2.3.2. *Graph representation*

To generate the graph representation of the EEG epochs, three elements are required: graph nodes, graph edges and node features. Here, graph nodes correspond to EEG channels and the graph edges to the functional connections among the nodes that model the brain network (Fig. 2). The node features are sets of temporal features extracted from EEG signals recorded at the nodes, and corresponded to the output of the CNN blocks.

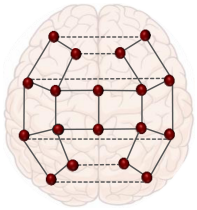


Fig. 2. Global inter-hemisphere (dashed lines) and local connections (solid lines) used for constructing graph representations of the EEG epochs. Red points represent the position of the middle point of each pair of electrode used for bipolar montage.

Bullmore and Sporns[49] claim that the brain network is structured to strike a balance between reducing wiring costs and increasing overall effectiveness. Local connections ensure that wiring costs, or the cost of creating anatomical connections between brain regions, are kept to a minimum. To replicate this aspect of brain structure, the proposed model takes into account the spatial relationships between pairs of EEG channels, in the bipolar montage. Specifically, the Euclidean distance between the middle points of the 3D coordinates of each pair of electrodes on the head surface is calculated using a standard head model based on the 10–20 system. Then, only the connections with the lower distances, which correspond to the closest pairs of electrodes, are kept to construct the graph. This enables the simulation of local connections in the brain network by connecting neighboring nodes. In Fig. 2, these local connections are shown in solid black lines.

However, maximizing brain network efficiency, i.e. the overall capacity for information transfer across brain regions, can be achieved through global connections. As shown in Fig. 2, some global interhemispheric connections are also added to the graph to improve the network efficiency. The following justifications provide credibility to the choice of interhemispheric connections as global connections (shown in Fig. 2 in dotted lines): (1) Prior studies on neonatal EEG showed that the degree of interhemispheric symmetry in the neonatal brain is informative in assessing background EEG activity.[50,51] (2) The selected symmetric channels covered interhemispheric homologous areas that have several

direct local connections that ease the information flow and thus maximize the network efficiency while keeping wiring costs low. (3) Several configurations were assessed to select the best-performing graph, and the one adopted here showed slightly higher performance.

Once the nodes, edges, and node features have been defined, the graph representation is complete and can be fed into the GAT blocks for exploring the spatial information.

2.3.3. *GAT blocks*

After constructing the graph representation using the output of CNNs and the generated adjacency matrix, GAT blocks, as recently proposed in Ref. 39 were applied to explore the spatial information embedded in the constructed graphs. GATs are a specialized type of neural network designed to handle graph-structured data. Specifically, GATs are a variant of Graph Convolutional Networks[52] (GCNs) that incorporate an attention mechanism to enhance their functionality. GCNs are neural networks capable of learning representations for nodes and edges in a graph. Analogous to the way CNNs extract features from image data using filters and execute message passing between pixels to propagate information throughout the network, GCNs employ filters to extract features from graphs and perform message passing between nodes to explore and propagate information.

However, GATs differ from traditional GCNs by utilizing an attention mechanism to allocate weights to messages originating from neighboring nodes (i.e. features) during the message passing process. The attention mechanism allows the model to selectively attend to and learn from task-relevant features in the graph. By assigning higher weights to more relevant messages, the GAT network is able to focus on the most important information, leading to better performance on the task at hand.

Figure 3 visually demonstrates the process by which each GAT layer leverages message passing to learn and update individual node features based on those of neighboring nodes. The attention weights can be learned as follows. Given the j-th node, the attention weight to node i can be calculated as

$$\alpha_{ij} = \frac{e_{ij}}{\sum_{k \epsilon N_i} e_{ik}} \tag{1}$$

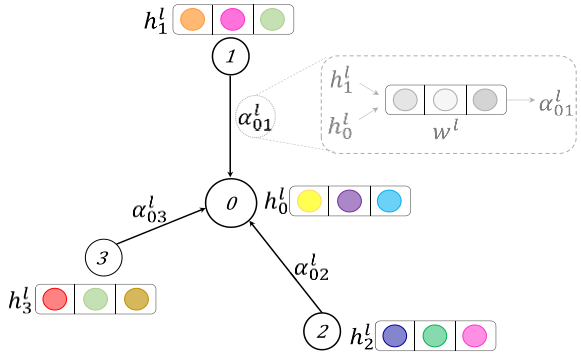


Fig. 3. Illustration of message passing and aggregation of features for updating a sample node 0 features (h_0^l) of layer l, using attention weights (α_{01}^l, α_{02}^l, α_{03}^l). Different colors indicate different feature values for each node.

where e_{ij} is node j's relevance to node i and N_i denotes the direct neighbors for node i. In GAT, this relevance is defined as:

$$e_{ij} = \text{Leaky ReLU}(a^l[W^l h_i^l \| W^l h_j^l]) \tag{2}$$

W^l is the learnable weight matrix of layer l, which maps the input features space to a hidden space, $\|$ is the concatenation operation and a^l is a learnable attention vector that is trained jointly with other network parameters within the optimization process. Finally, the attention weights α_{ij} of layer l are used to compute a linear weighted sum of features of a given node's neighbors. The updated feature vector of node i can be calculated as:

$$h_i^{l+1} = \sigma\left(\sum_{j \in N_i} \alpha_{ij}^l W^l h_j^l\right) \tag{3}$$

where σ is the nonlinearity function.

In the proposed model, three GAT layers are applied. The main reason for choosing three layers of GAT is to achieve an optimal spatial receptive field that facilitates the flow of information between every two nodes of the graph.

In each layer, graph nodes aggregate the first-order node features, updating their own node embeddings. Therefore, by stacking three GAT layers, aggregation from third-order neighbors is facilitated. After each GAT layer, the information is aggregated from the direct (first-order) neighbors of each node. The output of the first layer is updated based on the node's features in the input as well as the features from its direct neighbors. This process is repeated in the second and third layers, ultimately

allowing the model to aggregate information from third-order neighbors by stacking three GAT layers. Figure 4 illustrates the spatial receptive field that was achieved after stacking these layers. Dark, medium, and light blue in the adjacency matrix indicate each node's first, second, and third-order neighborhood. As Fig. 4 clearly shows, the adjacency network employed in this work and the architecture of ST-GAT provide an efficient and low-cost exploration of all spatial information included in multi-channel EEG epochs.

2.4. *Classification and training methodology*

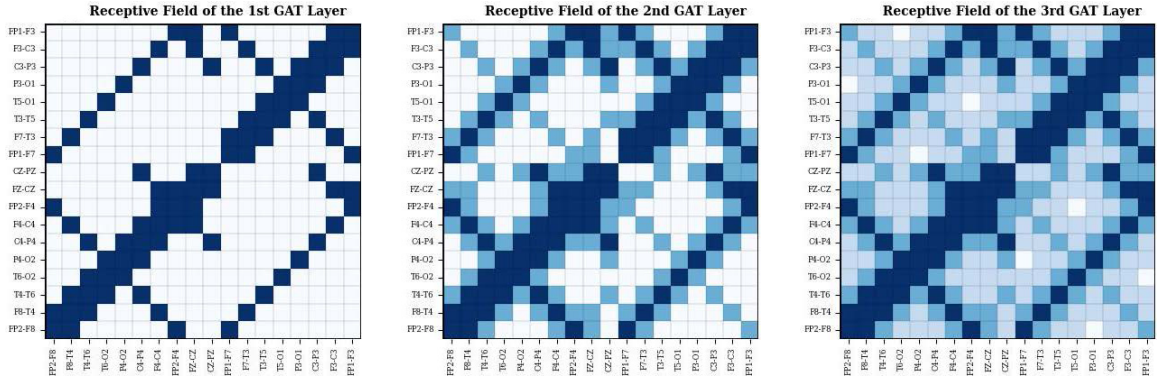
Following the GAT layers, there is a graph readout layer. The readout layer collapses node representations of each graph into a unique graph representation. Then, by applying two fully connected and softmax layers, the classification results are obtained: a scalar indicating the probability of each input being a seizure event. After two fully connected layers, for computing the classification error focal loss (FL[53]), is utilized instead of standard cross-entropy loss function (L_{CE}). Focal loss is defined as:

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t), \tag{4}$$

where p_t is the model's predicted probability of the true class and $(1-p_t)^{\gamma}$ is the modulating factor.

Focal loss is designed to address the problem of class imbalance in tasks — such as seizure detection — in which one class (nonseizure) is more prevalent than the other classes (seizure). In such scenarios, the model may prioritize the majority class and perform poorly on the minority class. Focal loss attempts to address this problem by down-weighting the contribution of easy examples (those that are already well-classified: true positives and true negatives) and focusing more on the hard examples (those that are misclassified: false positives and false negatives). The focal loss function modifies the cross entropy loss by the modulating factor. In general, a smaller value of λ places more emphasis on easy examples, while a larger value of λ places more emphasis on hard examples. A value of $\lambda = 0$ corresponds to the standard cross-entropy loss function. To train our proposed model, we utilized the Adam optimization approach,[44] a popular algorithm for training deep learning models.[7,54] Adam is known for its computational efficiency and convergence speed, making it suitable for training the proposed deep learning model. The learning rate was set to 0.001 for training the model. The learning rate is a crucial hyperparameter that controls the step size of the weight updates during optimization. Generally, a smaller learning rate results in slower convergence but may achieve better accuracy, while a larger learning rate speeds up convergence but may cause the model to overshoot the optimal solution.

In this study, a learning rate of 0.001 provided a good balance between convergence speed and model performance. The convolutional blocks and fully connected layers used dropout[45] with a probability of 0.2. Dropout works by randomly setting a fraction of the input units to zero during training at each



(a) (b) (c)

Fig. 4. (Color online) The spatial receptive field of the first, second, and third stacked GAT layers. Dark, medium, and light blue in the adjacency matrix indicate each node's first, second, and third-order neighborhood.

update. This process helps prevent the model from relying too heavily on any single input feature (which can be due to noise) and promotes the learning of more robust and generalizable features.

Weight decay is another regularization technique that helps prevent overfitting by adding a penalty term to the loss function. These two techniques have been widely used in the literature.[9,14,46,54] In this study, a weight decay coefficient of 0.0001 is used. This value was chosen after experimentation to provide sufficient regularization without significantly reducing the model's capacity to learn from the data. Combined with dropout, this weight decay value contributes to preventing overfitting and improving the model's generalization performance on unseen data.The configuration of the proposed ST-GAT is detailed in Table 1.

2.5. *Performance evaluation*

To evaluate the performance of the ST-GAT model, training and testing were carried out within a leave-one-subject-out cross-validation (LOSO) scheme. Based on the 39 neonates in the Helsinki dataset, the recordings from 38 neonates were used to train the model, and then the left-out subject was used to test the performance of the trained network on unseen data. This process was repeated until every subject had been tested.

2.5.1. *Evaluation metrics*

The following metrics were calculated to evaluate the performance of the proposed ST-GAT model:

- AUC: The area under the receiver operating characteristic (ROC) curve is a metric that measures the performance of a binary classifier. AUC equal to 0.5 denotes random chance classification, whereas AUC equal to 1 denotes an ideal classifier.
- FPR: False Positive Rate is the average number of false seizure detection per hour of EEG recording, calculated as the number of predicted seizure events in 1 h that have no overlap with actual reference seizures.
- Kappa: Cohen's Kappa is a statistical measure of the degree of agreement between the classifier output and the clinical labels. Cohen's Kappa

Table 1. Structure of the proposed ST-GAT model.

	Layer	Parameters
Input		(18×384)
Conv1	Convolution	$K = (1, 5)$, $S = 1$, $N = 4$
	Convolution	$K = (1, 5)$, $S = 1$, $N = 8$
	Max pooling	$K = (1,2)$
	Batch normalization	—
Conv2	Convolution	$K = (1, 5)$, $S = 1$, $N = 16$
	Convolution	$K = (1, 5)$, $S = 1$, $N = 32$
	Max pooling	$K = (1,2)$
	Batch normalization	—
Conv3	Convolution	$K = (1, 5)$, $S = 1$, $N = 8$
	Convolution	$K = (1, 5)$, $S = 1$, $N = 8$
	Max pooling	$K = (1,2)$
	Batch normalization	
Conv4	Convolution	$K = (1, 5)$, $S = 1$, $N = 1$
	Convolution	$K = (1, 5)$, $S = 1$, $N = 1$
	Max pooling	$K = (1,2)$
GAT1	Graph Attention	Out: (18×37)
GAT2	Graph Attention	Out: (18×32)
GAT3	Graph Attention	Out: (18×16)
	Average pooling along the spatial dimension	Out: (1×32)
Classifier	Fully Connected	Out: (1×16)
	Fully Connected	Out: (1×1)
Output	Softmax	Out: (1×1)

Note: K: Kernel size, S: Stride, N: Number filters in each convolution.

accounts for the class imbalance problem, making it an appropriate metric for evaluating the performance of seizure detection systems.

- GDR: Good Detection Rate quantifies the percentage of the seizures that are correctly detected. A seizure event is correctly detected if the model identifies at least one seizure epoch during the event.

2.5.2. *Ablation study*

An ablation study was carried out to assess the effectiveness of the proposed architecture, including each CNN or GAT block as well as the effect of the receptive field on seizure classification.[55] The receptive field is an important factor when developing CNNs for a particular task and is related to kernel size and the depth of the CNNs. The kernel size in a specific convolutional layer decides how much of the input region is processed in that layer. This region in the first layer's input is considered as the receptive field of that particular layer. For example, for two stacked convolutional layers, both with a 1×3 kernel size, the receptive field of the first and second layers are 1×3 and 1×5, respectively. A wider receptive field guarantees that the network is focused on more global patterns. In CNN-based studies of EEG signals, calculating the receptive field concerning the first layer's input and evaluating the effectiveness of each specific block can help to interpret the network dynamics.[56,57]

For the ablation study, the outputs of each layer were used as input of a simple classifier block, including graph readout (in the case of CNN layers), fully connected, and softmax to classify seizure/nonseizure EEG epochs. To train this classifier block, the same training data used for the ST-GAT model were used. The rest of the network remained fixed during this phase. This analysis demonstrates how each GAT or CNN layer with a specific receptive field contributes to the final performance of the ST-GAT model. Moreover, looking at the receptive field and the results of the ablation study simultaneously would lead to a better interpretation and optimal choice for window length and depth of the CNN blocks in the seizure classification task. Finally, to assess the appropriateness of GAT for our proposed network, a study in which GAT layers were replaced by two other graph convolutional layers (GraphSage[37] and GCNN[34]) is also conducted.

3. Results

Table 2 shows the results of ST-GAT using both standard cross entropy and focal loss and compares them with benchmark models on the Helsinki dataset. Additionally, we have included the results for training the ST-GAT model on a different level of imbalanced input data (ST-GAT ImB). In this scenario, the segmentation of both seizure and nonseizure segments maintains the same overlap, specifically, 12-s segments with a 10-s overlap. This analysis not only demonstrates the performance and robustness of the ST-GAT model when dealing with

Table 2. The overall comparison of different methods' performance on the Helsinki dataset (39 patients with seizures).

	AUC (%)				
Method	Median (IQR)	Mean ± std	FPR (/h)	GDR (%)	Kappa
Heuristic[a] (Deburchgraeve *et al.*[20])	68.3 (50.0–81.8)	66.0	**0.66**	—[b]	—
SVM (Temko *et al.*[21])	96.1 (86.9–99.0)	92.3	1.0	89.0	—
SVM (Tapani *et al.*[24])	98.8 (93.1–99.8)	95.7	0.86	—	—
CNN (O'Shea *et al.*[11])	—	95.6	—	—	—
MSC-GCNN (Raeisi *et al.*[35])	99.1 (96.8, 99.6)	94.7 ± 10.9	1.10	96.71	0.80
PLV-GCNN (Raeisi *et al.*[35])	99.0 (95.2, 99.7)	94.1 ± 10.5	0.89	95.3	0.79
Spatial Dis.-GCNN (Raeisi *et al.*[35])	97.3 (86.3, 99.6)	90.9 ± 13.5	0.92	96.68	0.71
ST-GAT (ImB)	99.0 (96.3, 99.6)	95.6 ± 8.5	0.70	94.20	0.79
ST-GAT (CE)	99.1 (96.9, 99.6)	95.9 ± 8.5	0.71	96.70	0.80
ST-GAT (FL)	**99.3 (96.4, 99.5)**	**96.6 ± 8.9**	0.86	**98.0**	**0.88**

Notes: Highest Value of each performance metric is shown in bold.

[a]As reported in Ref. 24.

[b]Missed values are not reported in the references.

imbalanced datasets, a common challenge in seizure detection tasks, but also highlights the impact of balancing the training data on the performance of our proposed model.

As it can be seen from the table, our proposed model with focal loss achieves the best performance in terms of AUC, kappa and GDR and outperforms our prior feature-based GCNN (i.e. MSC-GCNN, PLV-GCNN, Spatial Dis.-GCNN) in all evaluation metrics. In the best case, the ST-GAT (FL) model achieved a Kappa value of 0.98, primarily due to a high true negative rate and a low false positive rate, whereas in the worst case, the Kappa value dropped to 0.29, mainly attributable to an increase in false negatives. The standard deviation of the AUCs for ST-GAT is lower than that of feature-based models, demonstrating the relatively higher robustness of ST-GAT with respect to the inter-subject variability of neonatal seizures and to the effect of background EEG recordings.

The comparison of the FPR values shows that the ST-GAT model has the lowest number of false detections with respect to our prior feature-based models, with a relative FPR decrease of 0.24, 0.03, and 0.06 false detections per hour as compared to MSC-GCNN, PLV-GCNN, and SpatialDis-GCNN, respectively.

However, the lowest value of FPR is still reported by Deburchgraeve *et al.*[20] Moreover, when comparing the results obtained by ST-GAT utilizing standard cross entropy and focal loss, the higher effectiveness of focal loss in our classification task is clearly highlighted. The experiments with a different level of data imbalance (ST-GAT ImB) revealed a decline in GDR. This outcome aligns with our expectations: given the diverse types of seizures present in the dataset, balancing the dataset allows the network to be exposed more to the broad range of neonatal seizure types and consequently leading to an increased GDR.

Table 3 presents the results of the ablation study conducted on the ST-GAT architecture. As evidenced by the AUC and kappa values, each component of ST-GAT plays a crucial role in achieving the final performance, and removing any one of these blocks substantially degrades the model's performance. These findings not only demonstrate the significance of each ST-GAT block but also validate the carefully considered design choices made in constructing the ST-GAT architecture for the task of neonatal seizure detection. Table 4 shows the results of the comparison between the ST-GAT model and two other models where the GAT layers were replaced by GraphSage and GCNN. These results indicate that the model's performance is reduced when replacing GAT with any alternative layers. The training time for one training sample was comparable among the three networks, although the model with GraphSAGE showed a slightly higher training speed.

Table 3. Ablation study of the proposed ST-GAT. The network's name indicates the last block included in the ablation study.

	Results		
Network	AUC (%)	Kappa	Receptive field
Conv2	77.3 ± 8.3	0.50 ± 0.13	27 (0.8 s)
Conv3	84.7 ± 7.1	0.64 ± 0.16	63 (2.0 s)
Conv4	89.3 ± 9.3	0.72 ± 0.30	135 (4.2 s)
GAT1	94.1 ± 8.0	0.77 ± 0.19	135 (4.2 s)
GAT2	94.9 ± 8.9	0.81 ± 0.10	135 (4.2 s)
ST-GAT	96.6 ± 8.9	0.88 ± 0.13	135 (4.2 s)

To visualize and explain what GAT layers have learned, in Fig. 5 an example of the results obtained for a single neonate (subject #4) is shown, summarizing the learned attention coefficient between each directly connected channel pair in seizure and nonseizure segments. The top panel illustrates the true labels and the predicted probabilities of each sample (or time instant). The bottom panel shows the heatmap of attention coefficient trends across time. Figure 5 clearly shows how the attention coefficients in different connections differ between seizure and nonseizure segments of the EEG signals. In nonseizure areas (NS1, NS2, NS3, and NS4), the

Table 4. Comparison of models based on different GCNs.

	Results		
Network	AUC (%)	Kappa	Training time (s)
GCNN	94.6 ± 9.0	0.75 ± 0.10	3.89×10^{-6}
GraphSAGE	95.2 ± 8.8	0.76 ± 0.17	3.27×10^{-6}
ST-GAT	96.6 ± 8.9	0.88 ± 0.13	3.61×10^{-6}

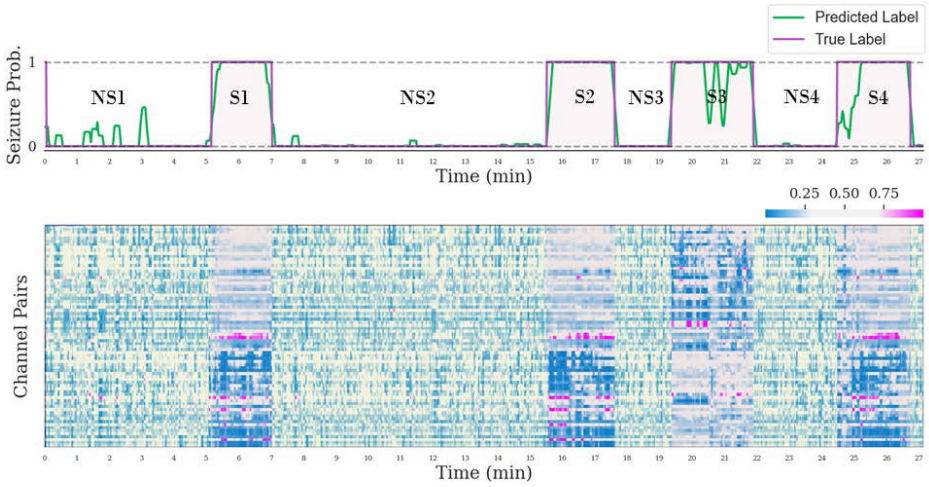


Fig. 5. (Color online) Visualization of the GAT coefficients for subject #4. The top panel represents the actual labels (purple) and predicted probabilities (green). The bottom panel represents the attention coefficients during the seizures and non-seizure epochs.

attention coefficients among different channel pairs are randomly distributed in mid-range values, i.e. between 0.2 and 0.3, suggesting that during regular brain activity, each channel pair connection has the same importance and that no particular pattern emerges. On the other hand, the heatmap in the seizure areas (S1, S2, S3, and S4) shows completely different and nonrandom patterns: the attention coefficients of the top half regions in S1, S2, and S4 vary between 0.2 and 0.3, whereas the bottom half regions include mainly smaller values — approximately around 0.1 — but also high values exceeding 0.9 (dark blue and dark pink).

The pattern for S3 shows the same dichotomy as the patterns of the other three seizure segments but with reversed values of the attention coefficients.

In Fig. 6, the circular graphs for attention coefficients of connections in the left and right brain

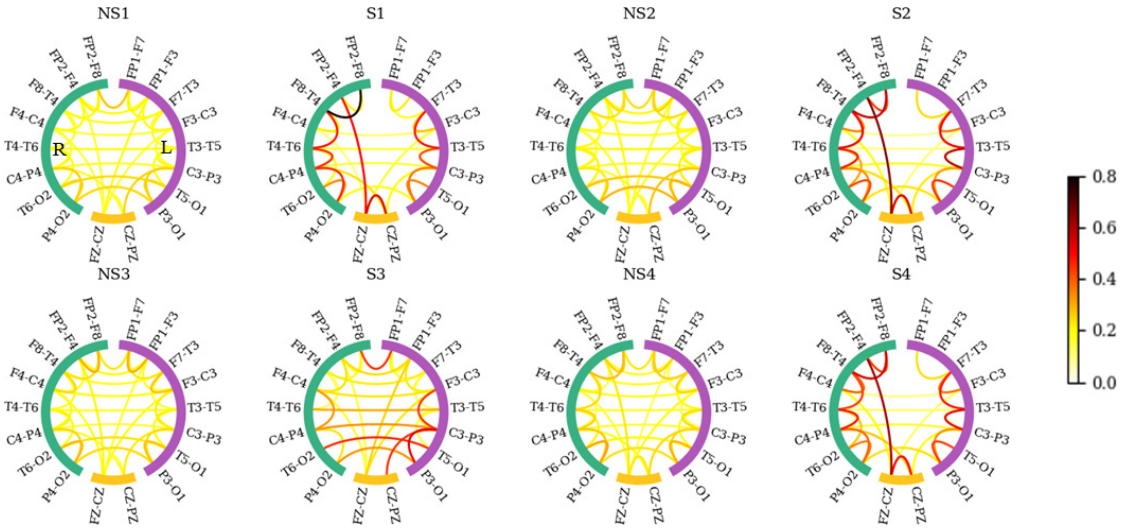


Fig. 6. (Color online) GAT coefficient between seizure and non-seizure EEG segments in both the left and right hemispheres of the neonate's brain for subject 4. During non-seizure epochs (NS1, NS2, NS3, and NS4), attention is relatively equal across all connections (as indicated by the yellow color). However, during seizure epochs, attention values significantly shift, with a greater emphasis, placed on the more important hemisphere of the brain (as indicated by red color connections).

hemispheres have been outlined to illustrate the location of connections that form the two halves of seizure segments in Fig. 5. Purple and green colors represent the electrodes in left and right hemispheres, respectively. It can be seen that, unlike nonseizure segments, during seizures the attention coefficient of connections in the left and right hemispheres are asymmetric, and one hemisphere contributes more to the output of the classifier, with some very high values (>0.8). This outcome is in line with the clinical information on the location of each seizure segment; meaning that during seizure segments attention mechanism puts more emphasis on the potential channels including seizure events. Some other visualization samples can be found in supplementary materials.

4. Discussion

In this paper, a novel end-to-end explainable ST-GAT model is proposed for the detection of neonatal seizures. Previous methods did not incorporate the spatial topology of multi-channel EEG and have modeled them as independent single-channel recordings, which ignores the precious spatial domain information. Differently, stemming from spatial topology modeling of the EEG channels and graph representation of each EEG epoch, our model can exploit the information in both time and space domains and use it for training the seizure classifier. The fact that our approach outperforms several other neonatal seizure detectors demonstrates the ability of ST-GAT to learn more meaningful and rich representations of the raw multi-channel EEG recordings.

As with any other classification task, defining a feature set able to properly and exhaustively represent the time domain information contained in neonatal seizures is challenging. These features are typically very complex, and their computation can be inefficient. Unlike our previous feature-based GCNN,[35] ST-GAT is an end-to-end framework that optimizes the extraction of features and the classifier performance simultaneously through a back-propagation process. The CNN layers used to complete the graph representation of the EEG epochs, which replace the *a-priori* definition of the features designed for the detection of neonatal seizure, are more effective in leveraging all the time information contained in the raw EEG signals and in making rich representations of them. The results indicate that our proposed architecture learns features with sufficient complexity to outperform the engineered features that were utilized to train our previous model. Our ST-GAT performs well in terms of both AUC and Cohen's Kappa, which is critical for a seizure detection model to be clinically applicable. Furthermore, several studies have reported values of Cohen's Kappa ranging from 0.4 to 0.93 among different experts.[58–60] Comparing the Cohen's Kappa obtained for ST-GAT model, which ranges between 0.29 and 0.98, with these values, a high agreement was observed between the ST-GAT output and the human expert, which is desired in clinical practice. This good performance of ST-GAT is even higher when using focal loss instead of standard cross entropy for training ST-GAT, suggesting a higher efficiency of focal loss in reducing the impact of the severe class imbalance problem that occurs when the EEG recorded in neonates with suspicious seizures is analyzed. This class imbalance problem derives from the fact that normally the duration of regular nonseizure brain activity is longer than intervals including seizure events.

The other significant difference between ST-GAT and prior GCNN models is that GAT layers are used instead of GCNN to extract spatial information from graph representations. This enhancement offers several benefits. First, GAT is more computationally efficient than GCNN and does not rely on costly spectral analysis and matrix operations as GCNN does. Second, as shown in Table 4, despite the simple architecture of GAT, the performance of ST-GAT is higher than that of models based on two other graph convolutions. Third, the attention mechanism in GAT layers is designed to assign a level of importance to each node within a node neighborhood. This characteristic of GAT provides a potential solution to model interpretability, which is a typical issue with deep learning models. Although it is not the primary goal of the current study, the importance level of each EEG channel for the output of the seizure classifier was visualized. The empirical findings suggest an asymmetric contribution of connections' importance in the right and left hemispheres of the brain during the seizure epochs. These results are in

line with the clinical information on the seizure locations: the level of importance was higher in the hemisphere that included the channels with seizure annotation. Although no spatial information on the seizure location is provided to the model during the training phase, the fact that the attention mechanism included in our model embeds information on the seizure location in an unsupervised manner makes our ST-GAT model a potential support tool for clinicians. By suggesting the most probable time and location of the seizure occurrences, the decision-making and evaluation process will be accelerated by the ST-GAT outcome.

Our proposed ST-GAT model offers other advantages for clinical applications. The GAT architecture allows to deal with different graph sizes, thus making the ST-GAT model insensitive to the number of input channels. It is well known that each clinical center generally uses a different number and layout of EEG channels. Therefore, the fact that the ST-GAT model does not depend on the number of electrodes would be a bonus for clinical settings. In fact, the electrode positions in an EEG headset are fixed, and the graph representations in our ST-GAT model are location-based, thus allowing for the generation of graph representations that can be easily adapted with any type of EEG recording.

Finally, it should be noted that a balance between receptive field and network characteristics was considered when designing the ST-GAT architecture. As shown in Table 3, most of the improvement in both AUC and Cohen's Kappa is achieved using wider temporal receptive fields. However, expanding the receptive field needs stacking more CNN blocks and increasing the number of network parameters to be tuned. A trade-off should be considered between the receptive field and the network parameters, as more network parameters need more computational resources and specialized hardware to train the model. Moreover, improving a neural network performance is not necessarily guaranteed by increasing its depth, which, on the other hand, would also increase the chance of overfitting. In fact, by adding more CNN blocks to the ST-GAT model, a minor decline in performance was observed. Hence, by employing the current architecture of ST-GAT, the desired performance was achieved for neonatal seizure detection using a relatively low number of parameters and a sufficiently large receptive field.

In terms of limitations, it should be mentioned that the applicability of our study is affected by the relatively small sample size employed for training the ST-GAT model. However, it should be underlined that this problem is common to all deep learning models, especially if developed for the classification of neonatal EEG recordings. In fact, deep learning models generally require many more training data to guarantee high performance and generalizability, but a large number of annotated neonatal EEG recordings are typically unavailable at single clinical centers. The future direction would be to include neonatal EEG datasets from other centers to improve the generalizability — and therefore the applicability — of our proposed ST-GAT model. The other limitation of the proposed method is the limited effectiveness of GAT when the number of input channels is very low. In such cases, the benefits of using GAT may be diminished due to the reduced complexity of the input data. The proposed neonatal seizure detection method is more suitable for EEG data with a higher number of channels, as this allows for more effective utilization of the relationships between channels.

4.1. *Conclusion*

The proposed neonatal seizure detection model integrates the temporal information of individual EEG signals with the spatial information on the EEG channels, which are embedded in the graph representation of the multi-channel EEG data. Here, by replacing hand-crafted features with 1D-CNNs, the resulting end-to-end network is capable of automatically and repeatedly developing a feature set that accurately characterizes the differences between seizure and nonseizure epochs in the time domain. By employing GAT instead of standard spectral graph convolutional layers, the attention mechanism in GAT was utilized to identify the essential channel pairs and information flow among brain regions. By stacking GAT layers on top of the CNN layers, the final ST-GAT model achieved high performance in detecting neonatal seizures, outperforming previous classifiers. Additionally, GAT coefficients were used to empirically visualize the importance levels of

neighboring node pairs during the seizure and non-seizure epochs, which can provide insight into the location of seizures in the neonatal brain.

Acknowledgments

This work was supported by the European Union under the H2020-EU.1.3.1 Program (European Training Networks Funding Scheme; INFANS Project "INtegrating Functional Assessment measures for Neonatal Safeguard"- G.A. nr. 813483).

Declaration of competing interest

The authors claim they have no conflict of interest.

References

1. L. Kharoshankaya *et al.*, Seizure burden and neurodevelopmental outcome in neonates with hypoxic–ischemic encephalopathy, *Develop. Med. Child Neurol.* **58**(12) (2016) 1242–1248, doi: 10.1111/dmcn.13215.
2. R. M. Pressler and L. Lagae, Why we urgently need improved seizure and epilepsy therapies for children and neonates, *Neuropharmacology* **170** (2020) 107854, doi: 10.1016/j.neuropharm.2019.107854.
3. P. Peng, L. Xie and H. Wei, A deep Fourier neural network for seizure prediction using convolutional neural network and ratios of spectral power, *Int. J. Neur. Syst.* **31**(8) (2021) 2150022, doi: 10.1142/S0129065721500222.
4. J. M. Rennie *et al.*, Characterisation of neonatal seizures and their treatment using continuous EEG monitoring: A multicentre experience, *Archiv. Dis. Childhood Fetal Neonatal Ed.* **104**(5) (2019) F493–F501, doi: 10.1136/archdischild-2018-315624.
5. Y. Zhao, G. Zhang, C. Dong, Q. Yuan, F. Xu and Y. Zheng, Graph attention network with focal loss for seizure detection on electroencephalography signals, *Int. J. Neur. Syst.* **31**(7) (2021) 2150027, doi: 10.1142/S0129065721500271.
6. A. T. Tzallas, M. G. Tsipouras and D. I. Fotiadis, Epileptic seizure detection in EEGs using time–frequency analysis, *IEEE Trans. Inform. Technol. Biomed.* **13**(5) (2009) 703–710, doi: 10.1109/TITB.2009.2017939.
7. M. A. Ozdemir, O. K. Cura and A. Akan, Epileptic EEG classification by using time-frequency images for deep learning, *Int. J. Neur. Syst.* **31**(8) (2021) 2150026, doi: 10.1142/S012906572150026X.
8. H. Daoud and M. A. Bayoumi, Efficient epileptic seizure prediction based on deep learning, *IEEE Trans. Biomed. Circ. Syst.* **13**(5) (2019) 804–813, doi: 10.1109/TBCAS.2019.2929053.
9. A. O'Shea *et al.*, Deep learning for EEG seizure detection in preterm infants, *Int. J. Neur. Syst.* **31**(8) (2021) 2150008, doi: 10.1142/S0129065721500088.
10. A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos and S. Van Huffel, Neonatal seizure detection using deep convolutional neural networks, *Int. J. Neur. Syst.* **29**(4) (2019) 1850011, doi: 10.1142/S0129065718500119.
11. A. O'Shea, G. Lightbody, G. Boylan and A. Temko, Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture, *Neural Netw.* **123** (2020) 12–25, doi: 10.1016/j.neunet. 2019.11.023.
12. Y. Wang *et al.*, A spatiotemporal graph attention network based on synchronization for epileptic seizure prediction, *IEEE J. Biomed. Health Inform.* **27**(2) (2023) 900–911, doi: 10.1109/JBHI.2022.3221211.
13. P. Thangavel *et al.*, Time–frequency decomposition of scalp electroencephalograms improves deep learning-based epilepsy diagnosis, *Int. J. Neur. Syst.* **31**(8) (2021) 2150032, doi: 10.1142/S0129065721500325.
14. X. Wang, G. Zhang, Y. Wang, L. Yang, Z. Liang and F. Cong, One-dimensional convolutional neural networks combined with channel selection strategy for seizure prediction using long-term intracranial EEG, *Int. J. Neur. Syst.* **32**(2) (2022) 2150048, doi: 10.1142/S0129065721500489.
15. G. Liu, L. Tian and W. Zhou, Patient-independent seizure detection based on channel-perturbation convolutional neural network and bidirectional long short-term memory, *Int. J. Neur. Syst.* **32**(6) (2022) 2150051, doi: 10.1142/S0129065721500519.
16. Y. Zhao *et al.*, Automatic seizure identification from EEG signals based on brain connectivity learning, *Int. J. Neur. Syst.* **32**(11) (2022) 2250050, doi: 10.1142/S0129065722500502.
17. G. Boylan, L. Burgoyne, C. Moore, B. O'Flaherty and J. Rennie, An international survey of EEG use in the neonatal intensive care unit, *Acta Paediatr.* **99**(8) (2010) 1150–1155, doi: 10.1111/j.1651-2227.2010.01809.x.
18. J. Gotman, D. Flanagan, J. Zhang and B. Rosenblatt, Automatic seizure detection in the newborn: Methods and initial evaluation, *Electroencephalogr. Clin. Neurophysiol.* **103**(3) (1997) 356–362, doi: 10.1016/S0013-4694(97)00003-9.
19. P. Celka and P. Colditz, A computer-aided detection of EEG seizures in infants: A singular-spectrum approach and performance comparison, *IEEE Trans. Biomed. Eng.* **49**(5) (2002) 455–462, doi: 10.1109/10.995684.
20. W. Deburchgraeve *et al.*, Automated neonatal seizure detection mimicking a human observer reading EEG, *Clin. Neurophysiol.* **119**(11) (2008) 2447–2454, doi: 10.1016/j.clinph.2008.07.281.
21. A. Temko, E. Thomas, W. Marnane, G. Lightbody and G. Boylan, EEG-based neonatal seizure detection

with support vector machines, *Clin. Neurophysiol.* **122**(3) (2011) 464–473, doi: 10.1016/j.clinph.2010.06.034.

22. K. T. Tapani, P. Nevalainen, S. Vanhatalo and N. J. Stevenson, Validating an SVM-based neonatal seizure detection algorithm for generalizability, non-inferiority and clinical efficacy, *Comput. Biol. Med.* **145** (2022) 105399, doi: 10.1016/j.compbiomed.2022.105399.
23. A. M. Pavel *et al.*, A machine-learning algorithm for neonatal seizure recognition: A multicentre, randomised, controlled trial, *Lancet Child Adolesc. Health* **4** (10) (2020) 740–749, doi: 10.1016/S2352-4642(20)30239-X.
24. K. T. Tapani, S. Vanhatalo and N. J. Stevenson, Time-varying EEG correlations improve automated neonatal seizure detection, *Int. J. Neur. Syst.* **29**(4) (2019) 1850030, doi: 10.1142/S0129065718500302.
25. E. M. Thomas, A. Temko, G. Lightbody, W. P. Marnane and G. B. Boylan, Gaussian mixture models for classification of neonatal seizures using EEG, *Physiol. Meas.* **31**(7) (2010) 1047–1064, doi: 10.1088/0967-3334/31/7/013.
26. A. Craik, Y. He and J. L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: A review, *J. Neural Eng.* **16**(3) (2019) 031001, doi: 10.1088/1741-2552/ab0ab5.
27. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan and H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, *Comput. Biol. Med.* **100** (2018) 270–278, doi: 10.1016/j.compbiomed.2017.09.017.
28. H. S. Nogay and H. Adeli, Detection of epileptic seizure using pretrained deep convolutional neural network and transfer learning, *Eur. Neurol.* **83**(6) (2020) 602–614, doi: 10.1159/000512985.
29. O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih and U. R. Acharya, Deep learning for healthcare applications based on physiological signals: A review, *Comput. Methods Progr. Biomed.* **161** (2018) 1–13, doi: 10.1016/j.cmpb.2018.04.005.
30. A. Caliskan and S. Rencuzogullari, Transfer learning to detect neonatal seizure from electroencephalography signals, *Neural Comput. Appl.* **33**(18) (2021) 12087–12101, doi: 10.1007/s00521-021-05878-y.
31. A. Gramacki and J. Gramacki, A deep learning framework for epileptic seizure detection based on neonatal EEG signals, *Sci. Rep.* **12**(1) (2022) 13010, doi: 10.1038/s41598-022-15830-2.
32. A. Fornito, A. Zalesky and E. Bullmore, *Fundamentals of Brain Network Analysis* (Academic Press, 2016).
33. A. K. Abbas, G. Azemi, S. Ravanshadi and A. Omidvarnia, An EEG-based methodology for the estimation of functional brain connectivity networks: Application to the analysis of newborn EEG seizure, *Biomed. Signal Process. Control* **63** (2021) 102229, doi: 10.1016/j.bspc.2020.102229.
34. T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, arXiv:1609.02907 [cs, stat], Feb. 2017, Accessed: Sep. 03, 2021. [Online]. Available: http://arxiv.org/abs/1609.02907.
35. K. Raeisi, M. Khazaei, P. Croce, G. Tamburro, S. Comani and F. Zappasodi, A graph convolutional neural network for the automated detection of seizures in the neonatal EEG, *Comput. Methods Programs Biomed.* **222** (2022) 106950, doi: 10.1016/j.cmpb.2022.106950.
36. M. Defferrard, X. Bresson and P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2016). Accessed: Sep. 03, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2016/hash/04df4d434d481c5bb723be1b6df1ee65-Abstract.html.
37. W. L. Hamilton, R. Ying and J. Leskovec, Inductive representation learning on large graphs (2018), doi:10.48550/arXiv.1706.02216.
38. D. Bahdanau, K. Cho and Y. Bengio, Neural machine translation by jointly learning to align and translate (2016), doi:10.48550/arXiv.1409.0473.
39. P. Velicković, G. Cucurull, A. Casanova, A. Romero, P. Liò and Y. Bengio, Graph attention networks, (2018), doi:10.48550/arXiv.1710.10903.
40. C. Sun, H. Cui, W. Zhou, W. Nie, X. Wang and Q. Yuan, Epileptic seizure detection with EEG textural features and imbalanced classification based on easy-ensemble learning, *Int. J. Neur. Syst.* **29**(10) (2019) 1950021, doi: 10.1142/S0129065719500217.
41. Y. Zhao *et al.*, EEG-based seizure detection using linear graph convolution network with focal loss, *Comput. Methods Progr. Biomed.* **208** (2021) 106277, doi: 10.1016/j.cmpb.2021.106277.
42. N. J. Stevenson, K. Tapani, L. Lauronen and S. Vanhatalo, A dataset of neonatal EEG recordings with seizure annotations, *Sci Data*, **6**(1) (2019) 190039, doi: 10.1038/sdata.2019.39.
43. E. Lashgari, D. Liang and U. Maoz, Data augmentation for deep-learning-based electroencephalography, *J. Neurosci. Methods* **346** (2020) 108885, doi: 10.1016/j.jneumeth.2020.108885.
44. H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar and P.-A. Muller, Deep learning for time series classification: A review, *Data Min Knowl. Disc* **33**(4) (2019) 917–963, doi: 10.1007/s10618-019-00619-1.
45. C. Nwankpa, W. Ijomah, A. Gachagan and S. Marshall, Activation functions: comparison of trends in practice and research for deep learning, (2018), doi:10.48550/arXiv.1811.03378.
46. J. Thomas *et al.*, Automated detection of interictal epileptiform discharges from scalp

electroencephalograms by convolutional neural networks, *Int. J. Neur. Syst.* **30**(11) (2020), 2050030, doi: 10.1142/S0129065720500306.
47. X. Wang and W. Q. Yan, Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory, *Int. J. Neur. Syst.* **30**(1) (2020) 1950027, doi: 10.1142/S0129065719500278.
48. O. Reyes and S. Ventura, Performing multi-target regression via a parameter sharing-based deep network, *Int. J. Neur. Syst.* **29**(9) (2019) 1950014, doi: 10.1142/S012906571950014X.
49. E. Bullmore and O. Sporns, The economy of brain network organization, *Nat. Rev. Neurosci.* **13**(5) (2012), Art. no. 5, doi: 10.1038/nrn3214.
50. J. R. Castro Conde *et al.*, Assessment of neonatal EEG background and neurodevelopment in full-term small for their gestational age infants, *Pediatr. Res.* **88**(1) (2020) Art. no. 1, doi: 10.1038/s41390-019-0693-0.
51. O. Räsänen, M. Metsäranta and S. Vanhatalo, Development of a novel robust measure for interhemispheric synchrony in the neonatal EEG: Activation Synchrony Index (ASI), *NeuroImage* **69** (2013) 256–266, doi: 10.1016/j.neuroimage.2012.12.017.
52. T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks. arXiv, Feb. 22, 2017, doi:10.48550/arXiv.1609.02907.
53. T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, Focal loss for dense object detection, presented at the *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988. Accessed: Mar. 06, 2023. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html.
54. G. Liu, W. Zhou and M. Geng, Automatic seizure detection based on S-transform and deep convolutional neural network, *Int. J. Neur. Syst.* **30**(4) (2020) 1950024, doi: 10.1142/S012906571 9500242.
55. R. Girshick, J. Donahue, T. Darrell and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. Accessed: Nov. 11, 2022. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html.
56. A. O'Shea, G. Lightbody, G. Boylan and A. Temko, Investigating the impact of CNN depth on neonatal seizure detection performance, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC*), Jul. 2018, pp. 5862–5865, doi: 10.1109/EMBC.2018.8513617.
57. H. Chen, Y. Song, and X. Li, A deep learning framework for identifying children with ADHD using an EEG-based brain network, *Neurocomputing* **356** (2019) 83–96, doi: 10.1016/j.neucom.2019.04.058.
58. P. J. Cherian *et al.*, Validation of a new automated neonatal seizure detection system: A clinician's perspective, *Clin. Neurophysiol.* **122**(8) (2011) 1490–1499, doi: 10.1016/j.clinph.2011.01.043.
59. D. K. Shah *et al.*, Accuracy of bedside electroencephalographic monitoring in comparison with simultaneous continuous conventional electroencephalography for seizure detection in term infants, *Pediatrics*, **121**(6) (2008) 1146–1154, doi: 10.1542/peds.2007-1839.
60. N. J. Stevenson, R. R. Clancy, S. Vanhatalo, I. Rosén, J. M. Rennie and G. B. Boylan, Interobserver agreement for neonatal seizure detection using multichannel EEG, *Ann. Clin. Transl. Neurol.* **2**(11) (2015) 1002–1011, doi: 10.1002/acn3.249.