Computer Methods and Programs in Biomedicine 222 (2022) 106950

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

A graph convolutional neural network for the automated detection of seizures in the neonatal EEG

Khadijeh Raeisi[a,1,*], Mohammad Khazaei[a,1], Pierpaolo Croce[a,b], Gabriella Tamburro[a,b], Silvia Comani[a,b], Filippo Zappasodi[a,b,c]

[a] Department of Neuroscience, Imaging and Clinical Sciences, University "Gabriele d'Annunzio" of Chieti–Pescara, Chieti, Italy
[b] Behavioral Imaging and Neural Dynamics Center, University "Gabriele d'Annunzio" of Chieti–Pescara, Chieti, Italy
[c] Institute for Advanced Biomedical Technologies, University "Gabriele d'Annunzio" of Chieti–Pescara, Chieti, Italy

ARTICLE INFO

Article history:
Received 6 January 2022
Revised 9 June 2022
Accepted 9 June 2022

Keywords:
Electroencephalography
Neonatal Seizure Detection
Deep Learning
Graph Neural Network
Functional Connectivity

ABSTRACT

Background and Objective: Neonatal seizures are the most common clinical presentation of neurological conditions and can have adverse effects on the neurodevelopment of the neonatal brain. Visual detection of these events from continuous EEG recordings is a laborious and time-consuming task. We propose a novel algorithm for the automated detection of neonatal seizures.

Methods: In this study, we propose a novel deep learning model based on Graph Convolutional Neural Networks for the automated detection of neonatal seizures. Unlike other methods exploiting mainly the temporal information contained in EEG signals, our method also considers long-range spatial information, i.e., the interdependencies across EEG signals. The temporal information is embedded as graph signals in the graph representation of the EEG recordings and includes EEG features extracted from the EEG signals in the time and frequency domains. The spatial information is represented as functional connections among the EEG channels (calculated by the phase-locking value and the mean squared coherence) or as maps of Euclidean distances. These different spatial representations were evaluated to assess their efficiency in providing more discriminative features for an effective detection of neonatal seizures. The model performance was assessed on a publicly available dataset of continuous EEG signals recorded from 39 neonates by means of the area under the curve (AUC) and the AUC for specificity values greater than 90% (AUC90).

Results: After applying post-processing, consisting in smoothing the output of the classifiers, the models based on the mean squared coherence, the phase-locking value, and the Euclidean distance respectively reached a median AUC of 99.1% (IQR: 96.8%-99.6%), 99% (IQR: 95.2%-99.7%), and 97.3% (IQR: 86.3%-99.6%), and a median AUC90 of 96%, 95.7%, and 94.9%. These values are superior or comparable to those reached by methods considered as state-of-the-art in this field.

Conclusions: Our results show that the EEG graph representations drawn from functional connectivity measures can effectively leverage interdependencies among EEG signals and lead to reliable detection of neonatal seizures. Furthermore, our model has the advantage of requiring only temporal annotations on seizures for the training phase, making it more appealing for clinical applications.

1. Introduction

Seizures are the most common clinical presentation of neurological conditions in neonates and could potentially have an adverse effect on neurodevelopment, with the risk of future cognitive and behavioral disabilities or epileptic outcomes [1,2]. Reliable and accurate detection of seizures in neonates is therefore essential to identify a dedicated therapy and prevent the risk of adverse conditions in later life. So far, visual inspection and clinical interpretation of continuous electroencephalographic (EEG) recordings have been considered the best approach to detect neonatal seizures [3]. However, this task is extremely laborious and time-consuming because it requires to be performed by highly trained experts who must inspect the time course of very long EEG recordings. This

* Corresponding author.
E-mail address: khadijeh.raeisi@unich.it (K. Raeisi).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cmpb.2022.106950

process is also prone to subjective interpretation and inter-expert variability. Consequently, the availability of an automated method for the detection of neonatal seizures is a crucial requirement to permit an effective use of EEG monitoring independently from the presence of specifically trained clinicians.

To date, several algorithms for the automated detection of neonatal seizures have been proposed. The earliest ones were based on heuristic rules and thresholds: the basic concept was to detect some clinical characteristics of neonatal seizures on the EEG, such as periodicity, rhythmic discharges, or spikiness. A temporal window of the EEG was then classified as "seizure" or "non-seizure" by comparing these clinical characteristics with pre-defined thresholds [4–7]. For instance, Liu et al. [6] developed a method based on the autocorrelation function, that was able to differentiate seizures from normal EEG background activity on EEG recordings by using specific features of the rhythmic discharges. This method reached a sensitivity of 84% and a specificity of 98%. Gotman et al. [5] proposed a method to detect rhythmic discharges, repetitive spikes, and slow rhythmic patterns occurring during seizures, which was able to detect 71% of the seizures. Rhythmic discharges were identified using spectral analysis, whereas repetitive spikes were extracted by breaking down the EEG into half-waves and assessing their duration, amplitude, and sharpness. Furthermore, Deburchgraeve et al. [8] defined two types of neonatal seizures and developed two detectors to mimic neonatal EEG experts: the first algorithm was based on the correlation between high-energetic segments of the EEG, whereas the second algorithm used autocorrelation to detect increases in the low-frequency EEG activity. The combined use of these two algorithms reached a classification sensitivity of 88%. The main drawback of these methods was their poor classification performance, indicating that they were not sufficiently reliable to be proposed for a routine use in the NICUs.

More recently, automated neonatal seizure detectors based on data-driven machine learning methods have been proposed [9–12]. Their advantage with respect to heuristic classifiers is that the extracted EEG features are fed to more advanced classifiers. For instance, using a support vector machine (SVM), Temko et al. [10] obtained an average detection rate of 89%, with one false seizure detection per hour. Tapani et al. [12] developed a neonatal seizure detection algorithm based on a set of 21 features, defined in the time, frequency, and time-frequency domains, which were fed to an SVM that achieved a median area under the receiver operator characteristics curve (AUC) of 98.8%. Although the overall performance of this type of seizure detectors is satisfying, their main drawback is that in addition to the temporal annotations on the instants when each seizure starts and ends, they also require precise spatial annotations on the EEG channels containing seizures to properly train the classifiers. This labeling scheme is extremely demanding for the clinicians who inspect the EEG. Consequently, automated algorithms that rely only on temporal annotations should be preferred for seizure detection.

In the past decade, deep learning has been employed as a promising alternative to traditional machine learning approaches in several research fields and practical applications [13–15], including the classification of neonatal seizures [16–21]. Ansari et al., [16] developed a hybrid architecture in which features were extracted within a convolutional neural network (CNN) and then seizures were classified with a decision tree to improve the classification performance. This hybrid method achieved an AUC of 88%. O'Shea et al., [17] proposed two deep learning algorithms containing 1-dimensional convolution filters, which implied that the correlations among the EEG signals recorded from different EEG channels were neglected. They reported an AUC of 98.5% for seizure detection on the EEG recordings of the database of Cork University and of 95.5% for the Helsinki University database. The reason for using only 1-dimensional convolutional filters was to develop a seizure detector that was independent of the baby-specific seizure locations. Despite the good results obtained, these methods suffer from the drawback that mainly the within-channel-based features are considered, whereas the long-range interdependencies among the EEG signals recorded at different EEG channels are ignored.

Given that seizures develop a synchronous neuronal activity [22], spatial information on the EEG channels that record the same seizure event would help to improve the performance of a seizure detector. Hence, an ideal seizure classifier should be based on the combination of within-channel-based (time mode) and between-channel-based (channel mode) features. This can be achieved by employing a graph representation of the multi-channel EEG data [23], as it includes both spatial information (the nodes, i.e., the EEG channels) and temporal information (the graph signals, i.e., features extracted from the EEG signals), and complements them with information on the interdependencies among the EEG signals (the edges, i.e., the connections between EEG channels). The edges are a unique feature of a graph representation of multi-channel EEG data and can be quantified in different ways. A simple spatial representation of the edges can be obtained with the pair-wise physical distance among nodes [24], whereas the pair-wise functional connectivity among nodes provides a more advanced quantification of the edges, and can be very useful for a neonatal seizure detector because it provides information on the functional interdependencies among EEG channels [25].

To fully exploit the information contained in a graph representation of multi-channel EEG data, graph convolutional neural networks (GCNNs) would be an appropriate alternative to CNN models [26]. GCNNs can leverage the graph structure and aggregate node information from the neighborhoods in a convolutional manner. GCNN and its variants have a great power to learn the graph representations and have achieved superior performance in a wide range of EEG applications including EEG-based emotion recognition [25,27-30], person identification [31], driver state monitoring [32], automated seizure detection in adults [24,33,34], sleep stage classification [35], autism disorder [36,37], and Alzheimer's disease [38]. Moreover, differently from CNNs that mainly work on image data with regular structures, GCNNs work on irregular non-Euclidean structured data with different graph structures [39]. This characteristic makes GCNNs insensitive to the ordering of the nodes in the graph representations (i.e., to the ordering of the EEG channels) and – consequently – to the location of the seizures. This would be an advantage in the case of neonatal seizure detection because one distinctive feature of neonatal seizures is that they are focal and their location changes from one neonate to another.

Given the aforementioned properties of GCNNs, in the present study, we propose a novel learning model for neonatal seizure detection that is based on GCNN. The structure of the graph representation of the multi-channel EEG data, i.e., the connections (edges) among the EEG channels, is made using two functional connectivity metrics: Phase Locking Value (PLV) [40] and Magnitude Squared Coherence (MSC) [41]. The PLV method can separate the phase component from the amplitude component in EEG recordings, and consequently is a suitable choice for exploring synchronous neuronal activities independently of amplitude changes. On the other hand, MSC quantifies both phase and amplitude coupling in pairs of EEG signals. To compare the effectiveness of these two functional connectivity measures in detecting neonatal seizures using GCNNs, we also used maps of Euclidean distances (spatial distance). Finally, to complete the graph representation of EEG epochs, a set of EEG features previously introduced by Tapani et al. [12] for seizure classification are used as graph signals.

This paper contributes to the literature on neonatal seizure detection in several aspects. To the best of our knowledge, our proposed method is the first GCNN method developed to detect

seizures on neonatal EEG recordings. It has the advantage of relying only on temporal annotations and has shown to be able to leverage information from functional connectivity graphs along with regular temporal information, leading to higher classification performance than existing methods. According to the promising results obtained, our method could be generalized to other tasks in the field of neonatal EEG classification.

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, 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 (for a total of 342 seizures). 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 0.5-30 Hz using a combination of seventh-order low-pass and high-pass non-causal Butterworth filters in both forward and backward directions. To decrease the computational load, the filtered signals were down-sampled to 64 Hz. According to Tapani and colleagues [12], who showed that the use of relatively long epochs slightly improves the ability of features to discriminate between seizure and non-seizure epochs, the filtered and down-sampled EEG signals were segmented into epochs of 32s. To obtain a higher number of training data, subsequent windows were overlapped by 31s.

2.3. Proposed method

The proposed learning model for neonatal seizure detection is based on GCNNs [26,43] that are a generalization of CNNs and are able to extract features from unstructured data like graph representations by performing convolutions for the graph signals defined on graphs. The overall pipeline of the proposed method for classifying the EEG epochs that contain seizure events is outlined in Fig. 1. The method consists of two parts: (1) a graph generator that converts the raw EEG epochs into graph representations (Fig. 1a), and (2) a GCNN-based classifier that extracts and learns discriminative patterns from the EEG-based graph representations (Fig. 1b). A third part of the method, aiming to improve the classification performance, consists in post-processing the outcome of the models.

Part 1: Graph generator. To prepare the information necessary for the GCNN-based classifier to discriminate seizure events, the EEG epochs must be converted into graph representations. Each graph representation is composed of 3 elements: the nodes (i.e., the EEG channels), the edges (i.e., the connections among the nodes), and the graph signals (i.e., the sets of features extracted from the EEG signals recorded at the nodes). To calculate the edges, we alternatively used three different approaches: two measures of functional connectivity and the information on the spatial distance between nodes, calculated as Euclidean distance. Functional connectivity maps were chosen as they capture the functional interactions between EEG channels that can be ascribed to the synchronized firing of neurons from different brain regions that occurs during neonatal seizures [22,44]. On the other hand, the spatial distance between nodes provides simple information on the spatial distribution of the nodes in the graph representation (i.e., on the neonatal head surface).

To calculate functional connectivity maps, we used two measures: PLV MSC. PLV quantifies the strength of the phase synchronization between pairs of EEG signals recorded at different channels. To calculate the PLV for a pair of EEG signals, the Hilbert transform of the EEG signals is first computed. Then, the instantaneous phase of each signal and the relative phase between the two signals are calculated. The PLV is then defined as:

$$PLV_{xy} = \left| \frac{1}{N} \sum_{j=0}^{N-1} e^{i\varphi_{xy}(i\Delta t)} \right| \tag{1}$$

where $\phi_{xy}[t]$ is the relative phase between the EEG signals from channels x and y, $1/\Delta t$ represents the sampling frequency of the EEG signals, and N is the number of samples in one EEG epoch. PLV can vary between 0 (no phase synchronization) and 1 (complete phase synchronization).

On the other hand, MSC quantifies the linear time-invariant relationship between two time series x and y (here, the EEG signals) at a given frequency λ, and is defined as:

$$Coh_{xy}(\lambda) = |R_{xy}(\lambda)|^2 = \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda) f_{yy}(\lambda)} \tag{2}$$

where $R_{xy}(\lambda)$ is the complex-valued coherence of x and y, $f_{xy}(\lambda)$ is the cross-spectrum of x and y, and $f_{xx}(\lambda)$ and $f_{yy}(\lambda)$ are the power spectrum of x and y, respectively [45]. MSC is a positive function, symmetric to x and y (i.e., $Coh_{xy}(\lambda) = Coh_{yx}(\lambda)$), and can vary between 0 and 1, which respectively indicate no coherence and a strong coherence between x and y at frequency λ.

The functional connectivity maps obtained from either PLV or MSC were then converted into adjacency matrices because only the information on the structure of the brain networks is important for the proposed model, which does not consider the strength of the connections between nodes. To calculate an adjacency matrix A, we applied the following rule:

$$A_{xy} = \begin{cases} 1, & if\ PLV_{xy} \geq R\ OR\ Coh_{xy}(\lambda) \geq R, \\ 0, & otherwise. \end{cases} \tag{3}$$

where R is a threshold that is equal to the median plus one standard deviation of all PLV or MSC values [46]. Sample adjacency matrices based on PLV and MSC are shown in Fig. 2b (middle and left panel). The adjacency matrices were then used to train two distinct GCNN classifiers (see *Part 2: GCNN-based classifier* below).

The spatial distances between pairs of EEG channels (nodes) were used as the third estimate of the edges in a graph representation of the EEG epochs. In this case, based on the distance between pairs of channels in the bipolar montage (Fig. 2a), only the neighboring nodes were connected, and the elements of the adjacency matrix were equal to 1 or 0 if two nodes were connected or not, respectively (Fig. 2b, right panel). The distance between two bipolar derivations was calculated as the distance between the centers of the bipolar derivations. A third GCNN classifier was trained using this estimate of the edges, and its performance was compared with that of the classifiers based on PLV and MSC to verify whether incorporating information on the functional connectivity among nodes in the GCNN classifier improved its efficiency with respect to a GCNN classifier where only the spatial distances among the nodes were provided.

Finally, the graph signals were calculated to complete the graph representation of the EEG epochs. To this aim, a set of 18 time-, frequency-, and time-frequency-based features, previously introduced by Tapani et al. [12], was calculated for each node from the

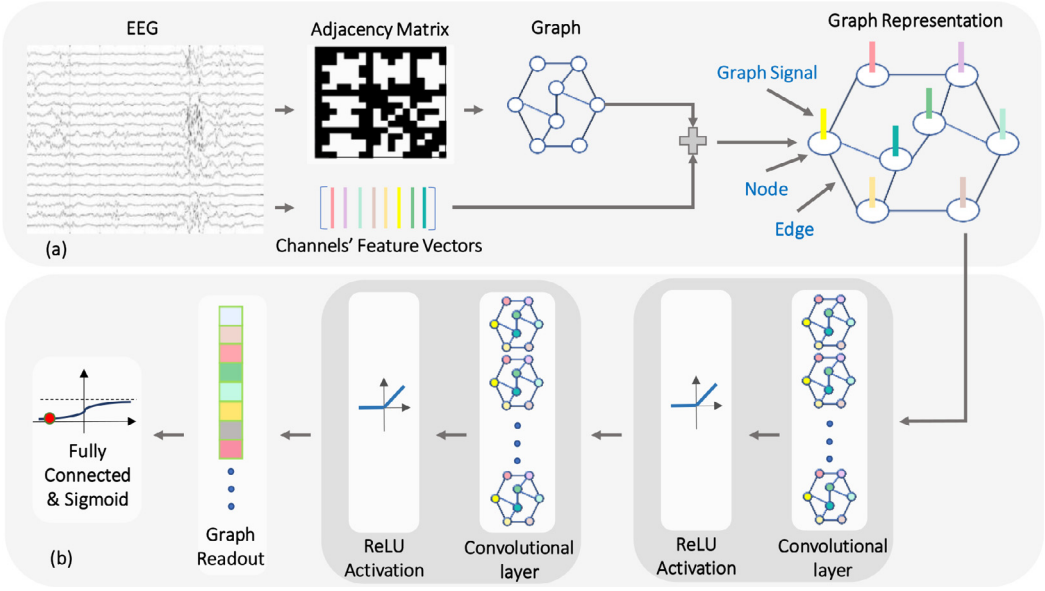

Fig. 1. The architecture of the proposed method for neonatal seizure detection. (a) The graph representation of EEG signals is the first part of this method. It is composed of nodes, edges (the estimated connections between EEG channels), and graph signals (which correspond to the feature vectors extracted from the EEG channels). (b) The GCNN is the second part of the proposed model. It consists of spectral-based convolution layers, the ReLU activation layers, the graph readout, and fully connected layers.

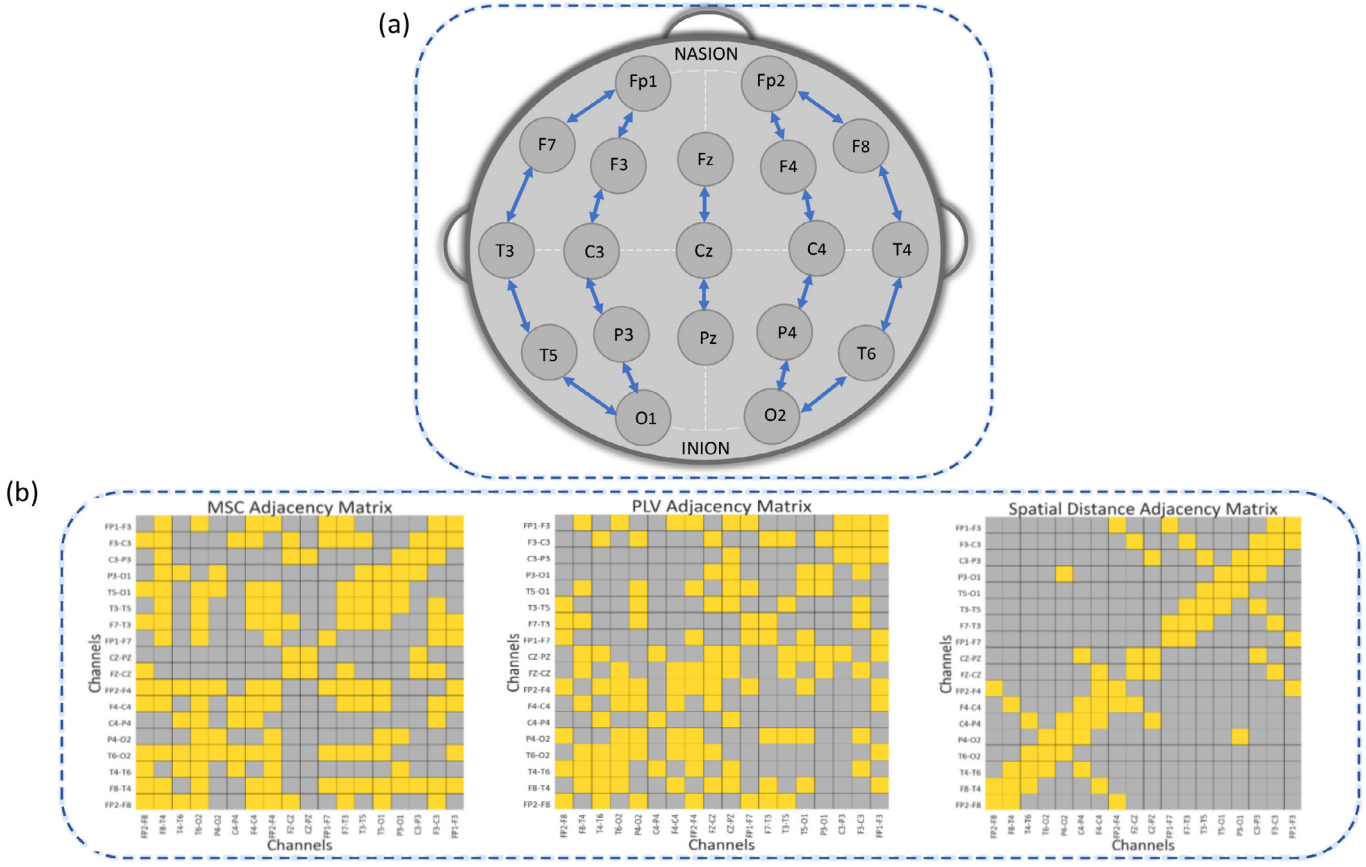

Fig. 2. (a) The bipolar EEG montage and (b) sample adjacency matrices derived from MSC (left), PLV (middle) and spatial distance (right).

EEG signal of that node. In particular, we calculated the following quantities: mean, standard deviation, skewness, and regularity of the smoothed nonlinear energy operator (SNLEO); number, duration, spike-interval, and maximum value of peaks of spikes; mean and standard deviation of spike correlation (SC); power in the first three harmonics, relative power in different frequency bands, and the maximum value of the short-time Fourier transform (STFT). Details on these features can be found in Appendix A and in [12].

Part 2: GCNN-based classifier. The graph representations were used to train three distinct GCNN classifiers. As described above, we obtained three graph representations of the EEG epochs that shared the same information on the nodes and the same graph

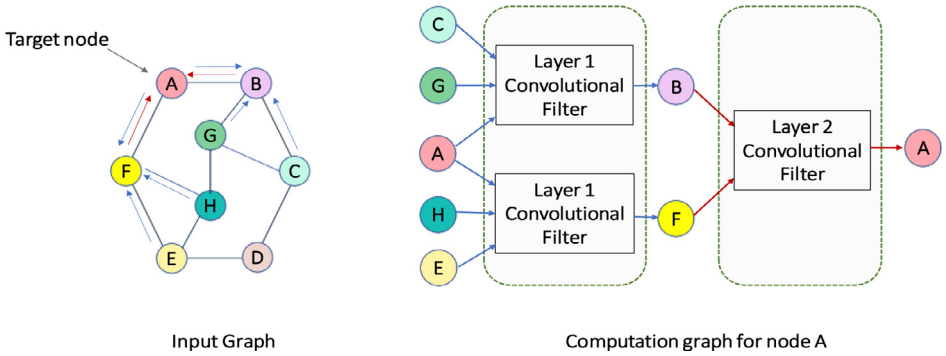

Fig. 3. Visualization of an aggregation of information through a two-layer GCNN model. On the left-hand side, the Input Graph for this example is given. The Computation Graph for updating the features (i.e., graph signal) of a sample node A uses the aggregation of the graph signals of node A and its local neighborhood (i.e., B and F). The graph signals coming from these neighbors (red arrows in the Input Graph) are also based on the aggregated signals of their local neighborhoods (blue arrows in the Input Graph).

signals, but differed for the edges, which were estimated by the adjacency matrices derived from either the PLV or the MSC connectivity maps or the spatial distances between the nodes.

A GCNN extracts features from graph representations by performing convolutions for the graph signals. To define graph convolutions, filters based on spectral graph theory [47] were used. The detailed mathematical description of the spectral-based GCNN is given in Appendix B. First, two layers of graph convolution were stacked. At each step, through the trained convolutional filters, the signal at each node in the adjacency matrix was updated by aggregating its own values with the values of the nodes connected to it by edges. The output of each layer was a graph that had the same structure as the input graph (i.e., the same connected nodes of the adjacency matrix) but updated graph signals. Thus, the convolutional filter of each layer was used to combine the information contained in the structure of the graph (i.e., in the adjacency matrix) with the information provided by the graph signals: this helped generating more informative graph signals. Fig. 3 shows an example of a computation graph for a sample node A through which the aggregation from second-order neighbors is facilitated by stacking two convolutional layers.

The output of these convolutional layers was a graph readout layer (Fig. 1b). Here, the updated graph signals were averaged across nodes to obtain a single and salient feature vector representing the whole graph. Finally, a sigmoid followed three fully connected layers to obtain the classification result: a scalar that indicates the probability of each input being a seizure event. For computing the classification error, the model was trained using the cross-entropy loss function, defined as:

$$loss = \sum_{x}(p(x)\log q(x) + (1-p(x))\log(1-q(x))) \quad (4)$$

where $p(x)$ and $q(x)$ represent the true and predicted label of a sample training set x. $p(x)$ equals 1 when a training set is a seizure event, and $q(x)$ is the probability that the training set is predicted as a seizure.

The optimization routine used mini-batch gradient descent with Adam optimization [48] and a learning rate of 0.001. The two convolutional and fully connected layers used dropout [49] with a probability of 0.3. A weight decay of 0.0001 was also considered to avoid overfitting of the network to the training data [50], and we used the rectified linear unit (ReLU) to increase the nonlinearity and power of the network [51]. Each GCNN was trained for 50 epochs. The structure and hyper-parameters for training the model were chosen by trial and error. Algorithm 1 of Appendix B summarizes the training steps for the proposed GCNN classifier and the configuration of the proposed GCNN is detailed in Table 1.

Table 1
Configuration of the proposed GCNN.

	Layer	Output Size
Input		(18 × 18)*
Block1	Graph Convolution	(18 × 64)
	ReLU	
	Graph Convolution	(18 × 32)
	ReLU	
	average pooling along the spatial dimension	(1 × 32)
Block2	Fully Connected	(1 × 32)
	ReLU	
	Fully Connected	(1 × 16)
	ReLU	
	Fully Connected	(1 × 1)
prediction	Softmax	(1 × 1)

* The input graph representation including 18 graph nodes and 18 EEG features.

Part 3: Post-processing. To maximize the performance of our models, a post-processing scheme was applied to the output of the trained models [10,12,17]. This phase includes smoothing the output of the classifiers using a moving average filter of length 32s. This filter reduced the number of false fast switching between seizure and non-seizure states and increased the temporal precision.

2.5. Performance analysis

To assess the performance of the three GCNN models on unseen data (patient-independent), training and testing were performed within a leave-one-subject-out cross-validation (LOSO) approach, where all subjects except one were used for training the classifier and then the left-out subject was used for testing. This process was repeated until every subject had been tested, for a total of 39 GCNNs for each type of connection estimate between nodes in the graph representation (i.e., PLV, MCS, spatial distance).

We characterized the performance of the proposed GCNN classifiers using AUC, which is a standard in the seizure detection literature [12,17]. The receiver operating characteristic (ROC) curve is a plot of sensitivity versus (1-specificity) and illustrates the performance of a binary classifier as its decision threshold is varied. For seizure detectors, sensitivity and specificity are defined as the proportion of real seizure and non-seizure segments that were correctly labeled by the classifier, respectively. An ideal classifier will have an AUC equal to one. Another collected metric is AUC90, which is defined as the AUC for specificity values greater than 90% (AUC90). This metric is more relevant when a classifier is meant to be used in a clinical setting as a low false detection rate is critically important for clinical applications of seizure classifiers.

Table 2
Performance of the proposed GCNN classifiers using the different representations of the spatial information among EEG channels (IQR: Interquartile Range).

Method	AUCMedian (IQR) (%)	AUCMean ± std (%)	AUC90Median (IQR) (%)	Kappa	GDR (%)	PPV(%)	NPV(%)
MSC + GCNN	0.978 (0.932, 1)	0.940 ± 0.102	0.942 (0.82, 0.98)	0.8	96.71	90	96
PLV + GCNN	0.985 (0.90, 0.99)	0.925 ± 0.110	0.954 (0.81, 0.99)	0.79	95.3	88	95
Spatial Dis. + GCNN	0.974 (0.88, 0.99)	0.898 ± 0.150	0.940 (0.79, 0.97)	0.71	96.68	88	96

Moreover, because the neonatal seizure detection is an imbalanced classification problem, the following measures are also reported:

- Cohen's Kappa: A statistical measure of the agreement between machine labels and true annotations and it corrects for the agreement expected by chance. Cohen's kappa takes into account the class-imbalance problem and therefore is a suitable metric for evaluating the performance of seizure detection systems.
- Good Detection Rate (GDR): The percentage of the seizure events which are correctly detected by the machine. A seizure event is correctly detected if the machine identifies at least one seizure epoch during the event
- Positive Predictive Value (PPV): The probability that epochs with positive machine label are truly seizure epochs.
- Negative predictive value (NPV): The probability that epochs with negative machine label are truly non-seizure epochs.

To assess the overall statistical significance of the differences between the AUCs obtained for the three trained GCNN classifiers, the non-parametric Friedman test [52] was performed. The Friedman test indicates whether there are overall differences among the classifiers but does not highlight the specific differences between pairs of classifiers. To make pair-wise comparisons between the three GCNN classifiers, a post-hoc test was performed by means of the non-parametric Wilcoxon signed-rank test [53], which evaluates whether the paired difference of two distributions comes from a distribution with zero median or not, i.e., whether the two compared distributions are significantly different or not.

3. Results

Table 2 reports the overall performance of the proposed GCNN classifiers. These values were calculated across 39 neonates. The ROC curves for the three GCNN classifiers, from which the AUC and AUC90 were measured, are plotted in Fig. 4. In particular, the median ROC curve (in blue), and the mean ROC curve (in red) are shown for each GCNN classifier. The median ROC curve of the GCNN classifiers using graph representations with the edges estimated from MSC, PLV, and spatial distance has an AUC of 97.8%, 98.5%, and 97.4%, and an AUC90 of 94.2%, 95.4%, and 94%, respectively. The GCNN classifiers based on functional connectivity estimates of the edges achieved more stable results than the GCNN classifier based on spatial distance also in terms of the standard deviation, which, respectively for MSC, PLV, and spatial distance, was equal to 10.2%, 11%, and 15% for AUC and equal to 9.1%, 8% and 8.5% for AUC90. Overall, the GCNN classifiers based on PLV and MSC outperformed the GCNN classifier based on spatial distance.

Table 3 illustrates the performance of the three GCNN classifiers after applying post-processing. As shown in the table, post-processing slightly improved the median and mean AUC and AUC90 for all classifiers except for the mean AUC90 of the GCNN based on spatial distance. When applying post-processing, the GCNN classifier based on graph representations derived from the MSC estimates of the edges had the best performance compared to the other two GCNN classifiers (median AUC = 99.1%, and median AUC90 = 96%).

The Friedman's test comparing the AUCs obtained from the three GCNN classifiers was significant after applying post-processing [$\chi^2(2) = 8.536$, p-value = 0.014] and close to significance before [$\chi^2(2) = 5.485$, p-value = 0.064]. Table 4 provides the results of the Wilcoxon signed-rank test for post-hoc comparisons on each pair of GCNN classifiers. The only statistically significant differences (p-value < 0.05) were observed when comparing the performance of the GCNN classifier based on MSC estimates of the edges with the GCNN classifier based on spatial distance, with the MSC-based classifier having a higher performance for both before and after post-processing conditions. The comparison between the GCNN classifiers based on PLV and spatial distance did not highlight any significant difference, although the AUC showed higher values for PLV (p-value = 0.179 before and p-value = 0.065 after post-processing). No significant differences were observed when comparing the performance of the GCNN classifiers based on the two different functional connectivity measures (p-value = 0.064 before and p-value = 0.117 after post-processing).

Finally, for a better comparison of our developed models and the previously published papers on the Helsinki dataset, the reported AUCs are summarized in Table 5.

4. Discussion

In this study, we propose a novel model for the automated detection of neonatal seizures that is based on GCNN applied to graph representations of multi-channel EEG recordings. One major drawback of neonatal seizure detectors based on deep learning, and more specifically on CNN, is that they are based on 1-dimensional filters, meaning that they mainly consider the temporal information extracted from the EEG signals while ignoring the long- and short-range functional interdependencies among the EEG channels. However, the occurrence of seizure events is generally recorded by more than one EEG channel. Therefore, the interdependencies among EEG channels – expressed through estimates of the functional connections among EEG channels – are worth exploring [22]. Some authors claimed that pairs of neighboring EEG channels (typically defined as an EEG montage) are sufficient to represent the spatial information on the EEG signals and to characterize functional connections [16]. However, the EEG montage can represent the interdependencies among EEG channels (i.e., among the underlying brain areas) imperfectly, because the functional connections between distant brain areas are completely ignored. Moreover, CNNs based on 2-dimensional filters, which analyze local areas using fixed convolutional filters, cannot capture long-range functional interactions among brain areas. Conversely, the GCNN-based model that we propose not only explores the features extracted from individual EEG channels in the time domain, but also captures both short- and long-range spatial connections among EEG channels: these connections are expressed by the elements of the adjacency matrix included in the graph representation of the multi-channel EEG signals.

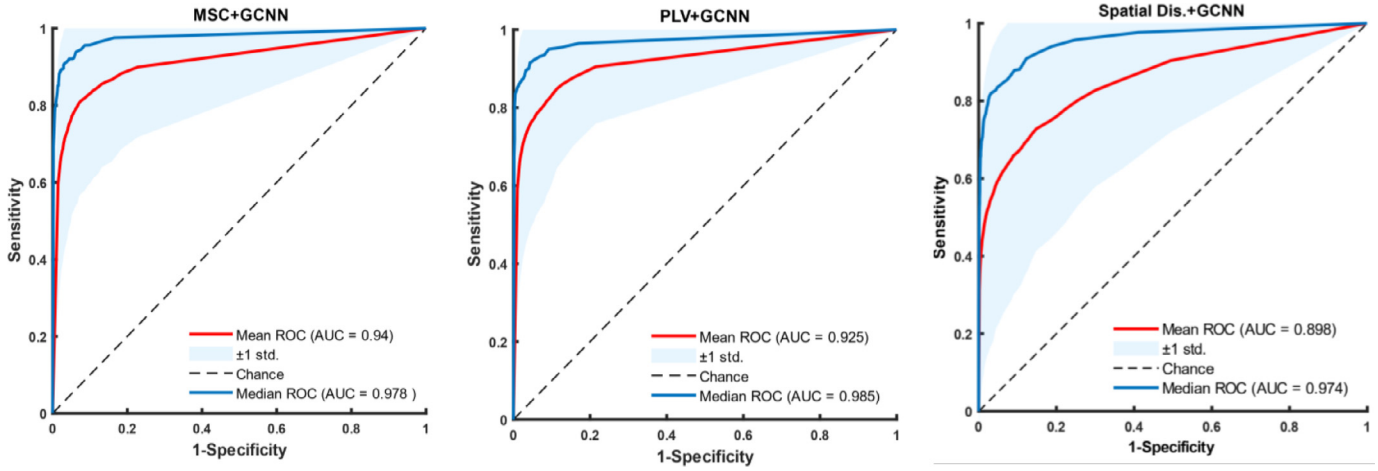

Fig. 4. ROC curves of the proposed models. Mean (red line), standard deviation (blue shaded area), and median (blue line) ROCs are calculated using the obtained ROC curves of each EEG recording for models based on MSC (left), PLV (middle), and spatial distance (right).

Table 3
The performance of proposed models after applying post-processing.

Method	AUCMedian (IQR) (%)	AUCMean ± std (%)	AUC90Median(IQR) (%)	GDR (%)	Kappa	PPV(%)	NPV(%)
MSC + GCNN	99.1 (96.8, 99.6)	94.7 ± 10.9	96.0 (87.5,98.2)	100	0.82	91	98
PLV + GCNN	99 .0 (95.2, 99.7)	94.1 ± 10.5	95.7 (86.9,98.6)	100	0.8	88	99
Spatial Dis. + GCNN	97.3 (86.3,99.6)	90.9 ± 13.5	94.9 (84.6,98.4)	100	0.73	89	98

Table 4
The p-value of statistical comparison between each possible network pair before and after applying post-processing. Asterisks denote statistically significant differences at the $p < 0.05$ level.

	Network Pairs	P-Value	Z-Value
Before post-processing	MSC - PLV	**0.064**	-1.850
	MSC – Spatial Distance	**0.015***	-2.431
	Spatial Distance- PLV	0.179	-1.342
After post-processing	MSC - PLV	0.117	-1.564
	MSC – Spatial Distance	**0.004***	-2.824
	Spatial Distance- PLV	**0.065**	-1.846

Table 5
The Overall Comparison of different methods' performance on the Helsinki dataset.

Author	Method	Validation	AUC (%)
Deburchgraeve et al.[a] [8]	Heuristic Rules	LOSO	68.3
Temko et al.[a] [10]	SVM	LOSO	96.1
Tapani et al. [12]	SVM	LOSO	98.8
O'Shea et al. [17]	2D CNN	LOSO	95.6
Caliskan et al. [19]	Transfer Learning (densenet)	Training (50%) and test (50%)	99.1
Tanveer et al. [54]	Ensembled 2D CNN	10-fold cross-validation	99.3
MSC + GCNN	GCNN	LOSO	99.1
PLV + GCNN	GCNN	LOSO	99.0
Spatial Dis. + GCNN	GCNN	LOSO	97.3

[a] As reported in [12].

In our model, the best performance was obtained when the GCNN model was associated with a spatial representation drawn from functional connectivity maps (MSC and PLV). These models had superior performance than the model based on the spatial distance, demonstrating that pair-wise functional connectivity can more effectively leverage the signal interdependencies contained in multi-channel EEG recordings. This result may be ascribed to the ability of graph representations based on functional connectivity to represent both short- and long-range interdependencies between pairs of EEG signals (hence short- and long-range functional connections among brain areas). In contrast, a spatial distance matrix ignores the long-range relationships across EEG channels. Therefore, using functional connectivity measures to construct a graph representation of multi-channel EEG data seems to be useful for better discriminating seizure and non-seizure events in neonatal brain activity. Furthermore, the results obtained after post-processing the outcome of the GCNN models demonstrated that the GCNN model based on MSC had a superior performance in detecting neonatal seizures than the model based on PLV. This finding indicates that functional connectivity measures like MSC, by quantifying both phase and amplitude coupling in pairs of EEG signals, provide more helpful information for detecting neonatal seizures than measures that only quantify phase synchrony, such as PLV.

Like other classification problems, models for neonatal seizure classification aim to achieve a high performance utilizing optimized feature extractors and classifiers. In recent years, few studies have focused on leveraging the potential of deep learning for the automated detection of neonatal seizures [16–18]. Our results confirmed that deep learning models based on a graph representation of the neonatal multi-channel EEG data can achieve high performance in the automated detection of neonatal seizures. In particular, our proposed GCNN model based on MSC functional connectivity maps and reinforced by the application of post-processing to the output of the trained model clearly outperformed conventional machine learning methods and the ones based on heuristic rules. It also achieved a performance superior or comparable to the previously published methods on Helsinki dataset (Table 5). As we can see in Table 5, the validation methods employed by Caliskan et al. and Tanveer et al. are based on splitting training and test data. This evaluation method is not patient-independent and there-

fore usually shows higher performances than completely patient-independent methods (LOSO).

A specific advantage of our GCNN model is that only temporal annotations on the neonatal EEG recordings are required for the training phase, unlike some best performing models and algorithms previously proposed that generally require also spatial information on the EEG channels where seizures were observed. This is a positive aspect of our GCNN model because the annotation of seizure events on long neonatal EEG recordings is a time-consuming process that can only be performed by highly trained experts, and it would become even more labor-intensive if also spatial information should be annotated. Therefore, the availability of a method that has high performance but requires only temporal annotations for the training phase is a big step ahead towards the clinical use of automated neonatal seizure detectors. Furthermore, the need for only temporal annotations is the premise for our proposed method to be more generalizable and able to detect neonatal seizures originating from different brain areas.

It should also be noted that our model is affected by some limitations. The first limitation is the small sample size which reduces the applicability of our best GCNN classifier. This is a common problem in deep learning models that generally require a large number of training data to guarantee a high accuracy but face the issue of limited availability of online databases [55–57]. Another drawback associated with our GCNN model is the relatively long computational time, mainly due to the extraction of the features needed to construct the EEG graph representations (0.936 s for each 32 s EEG epoch). Computational time is a crucial issue for models intended for neonatal EEG applications because EEG recordings usually have a duration of many hours. A reduction of the computational time could be achieved by using CNNs for feature extraction, hence increasing the overall performance of the model. In future versions of our GCNN model, we will consider the inclusion of other layers as temporal feature extractors that are stacked by graph convolutional layers. Finally, another drawback of our classifier is that it uses spectral-based GCNNs that are generally constructed using the Laplacian of the adjacency matrix, implying that the method is not robust with respect to variations of the number of EEG channels. This issue could be addressed by utilizing spatial-based GCNNs.

5. Conclusion

The seizure detection model proposed in this study is the first GCNN model for the classification of neonatal EEG. It achieved high performance and had the advantage of relying only on temporal annotations of seizures for the training phase. In contrast to seizure detectors based on CNNs, the proposed model combines temporal features extracted from individual EEG signals with short- and long-range spatial interdependencies among EEG channels, which are embedded in the graph representation of the multi-channel EEG data. Both functional and spatial connections among the EEG channels were explored, facilitating the evaluation of the efficiency of functional connectivity measures in providing more discriminative information for the detection of neonatal seizures in comparison to a representation simply based on spatial distances among the EEG channels.

Declaration of Competing Interest

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

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).

Appendix A

The features used to build the graph signals are calculated in time, frequency, and time-frequency domains. Time-domain features are based on the smoothed nonlinear energy operator (SNLEO) which is defined as follow:

$$\mathrm{SNLEO}\{S(n)\} = S(n-l)S(n-p) - S(n-q)S(n-s) \tag{A.1}$$

where S(n) indicates the EEG signal at sample n, and l, p, q and s represent time shifts. In this study, these free parameters were set to 1, 2, 0, and 3 respectively. A moving average filter with a length of 7 samples was used to smooth this operator. To calculate spike correlation, EEG signals are first segmented into epochs (for more details see Tapani et al., 2018), and then spike correlation is computed as follows:

$$\begin{aligned} SC_{i,j} &= \max_m R_{x_i x_j}(m) \\ R_{x_i x_j}(m;i,\ j) &= \frac{E\left[x_i(n)x_j(n+m)\right]}{\sigma_{x_i}\sigma_{x_j}} \end{aligned} \tag{A.2}$$

where x_i and x_j are the EEG epochs ($i = [2,\ \ldots,\ W-6]$, $j = [i+1,\ \ldots,\ i+5]$), W represents the number of spikes in each epoch, and σ is the standard deviation of the EEG epochs.

To extract frequency domain features the power spectral density (PSD) of the EEG signals are estimated using Welch's method. Short-time Fourier transform (STFT) is used to extract time-frequency domain features. Table A.1 gives a short description of the features used in this study.

Table A.1
Short descriptions of features in the proposed algorithm.

Feature	Description
Mean SNLEO	Mean of the SNLEO
STD SNLEO	Standard deviation of the SNLEO
Skewness	Skewness of the SNLEO
Regularity	Standard deviation of skewness of 2s sub-windows of the SNLEO
Spike Number	Number of spikes in an epoch
Spike Width	Median of width of spikes
Spike Gap	Median of inter-spike intervals
Spikiness	Mean of spike peaks over background
Mean of Spike Correlation	Mean of correlation between spikes
STD of Spike Correlation	Standard deviation of correlation between spikes
TDH1	Power in the first three harmonics divided by sum of the PSD
TDH2	Power in the fundamental divided by the sum of the PSD
TDH3	Logarithm of the power in the fundamental
Relative Delta Power	The ratio of the power in 0.5-4 Hz and the total power
Relative Theta Power	The ratio of the power in 4-8 Hz and the total power
Relative Alpha Power	The ratio of the power in 8-12 Hz and the total power
Relative Beta Power	The ratio of the power in 12-30 Hz and the total power
THD_{TF}	Sum of max STFT of each slice divided by sum of the STFT

Appendix B

Spectral-based GCNNs [26,43] rely on the Fourier transform and the normalized Laplacian matrix of the graph. The Normalized Laplacian matrix L of a graph G is defined as:

$$L = I_N - D^{-1/2} A D^{-1/2} \tag{B.1}$$

where I_N is the identity matrix, $D \in R^{N \times N}$ is the diagonal degree matrix, and $A \in R^{N \times N}$ is the adjacency matrix of graph G. Each element of the matrix D is the sum of each corresponding row in the adjacency matrix.

Now, let$x \in \mathbb{R}^N$ denote a graph signal defined at the nodes of graph G, in a way that x_i corresponds to the value of the features of node i. The Graph Fourier transform of x would be defined as:

$$\hat{x} = U^T x \tag{B.2}$$

where $U = [u_0,..., u^{N-1}] \in R^{N \times N}$ is the Fourier basis or eigenvectors of the Laplacian matrix L, in a way that $L = U \Lambda U^T$. The eigenvalues $\Lambda = diag([\lambda_0,..., \lambda_{N-1}]) \in R^{N \times N}$ are also regarded as the graph frequencies. Assuming g_θ as a filter with learnable parameter θ, the filtered signal y can be defined as:

$$y = g_\theta(L)x = g_\theta\left(U \Lambda U^T\right)x = U g_\theta(\Lambda) U^T x \tag{B.3}$$

where $g_\theta(\Lambda) = diag([g_\theta(\lambda_0),..., g_\theta(\lambda_{N-1})])$ represents the graph filter. Then, the k-order localized polynomial filter is defined as:

$$g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k T_k\left(\tilde{\Lambda}\right), \tag{B.4}$$

where $\tilde{\Lambda} = 2\Lambda/\lambda_{max} - I_N$, is the scaled eigenvalues, and λ_{max} is the maximum eigenvalue of L, and $\theta \in \mathbb{R}^K$ represents the coefficient vector of the Chebyshev polynomials. The Chebyshev polynomial $T_k(x)$ of order k can be recursively computed according to the following recurrence relation:

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x),$$
$$T_0(x) = 1, \; T_1(x) = x. \tag{B.5}$$

By introducing Chebyshev polynomials, the calculation of the eigenvectors of the Laplacian matrix is avoided, with a notable reduction in the computational cost. The graph convolution formulated in (B.4) can thus be rewritten as:

$$y = U g_\theta(\Lambda) U^T x = \sum_{k=0}^{K-1} \theta_k U T_k\left(\tilde{\Lambda}\right) U^T x = \sum_{k=0}^{K-1} \theta_k T_k\left(\tilde{L}\right) x \tag{B.6}$$

By limiting the layer-wise convolution operation to 1, we can still recover a rich class of convolutional filter functions by stacking multiple similar layers without being limited to the explicit parameterization given by the Chebyshev polynomials. By approximating $\lambda_{max} = 2$, and $\theta = \theta_0 = -\theta_1$ (to minimize the number of operations per layer) equation (B.6) is changed to:

$$y = \theta\left(I_N + D^{-1/2} A D^{-1/2}\right) x$$
$$y = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} x \; \theta \;\; (Normalized) \tag{B.7}$$

with a single learnable parameter θ where $\tilde{A} = A + I_N$ and $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. $\theta \in \mathbb{R}^{C \times F}$ is the matrix of convolutional filter parameters and $y \in \mathbb{R}^{N \times F}$ is the convolved feature matrix. The convolutional filter parameters can then be shared over the whole graph. The stacking of filters of this form effectively convolves the k^{th}-order neighborhood of a node, where k is the number of convolutional layers in the GCNN model [26] (Algorithm B.1).

Algorithm B.1
Training steps of the proposed GCNN classifier for neonatal seizure detection.

Input: EEG feature matrix, input graph representation, the true labels corresponding to each EEG segment, the learning rate and the number of epochs (***N***);
Output: The optimized model;
1: Initialization of model parameters;
2: Calculating the adjacency matrix;
for $n < N$ do:
3: Sample a mini-batch from the training dataset;
4: Calculating the normalized Laplacian matrix (equation B.1);
5: Calculating the output of graph convolutional layers (equation B.7);
6: Regularizing layer outputs using the ReLU operation;
7: Calculating the graph readout;
8: Calculating the output of the fully connected and sigmoid layers;
9: Calculating the cross-entropy loss function (equation 4);
10: Updating model parameters and Adam optimization;
end

References

[1] M. Levene, The clinical conundrum of neonatal seizures, Archives of Disease in Childhood - Fetal and Neonatal Edition 86 (2) (2002) F75–F77 Mar., doi:10.1136/fn.86.2.F75.
[2] J.M. Rennie, G. Chorley, G.B. Boylan, R. Pressler, Y. Nguyen, R. Hooper, Non-expert use of the cerebral function monitor for neonatal seizure detection, Archives of Disease in Childhood - Fetal and Neonatal Edition 89 (1) (2004) F37–F40 Jan., doi:10.1136/fn.89.1.F37.
[3] S.O. Wietstock, S.L. Bonifacio, J.E. Sullivan, K.B. Nash, H.C. Glass, Continuous Video Electroencephalographic (EEG) Monitoring for Electrographic Seizure Diagnosis in Neonates: A Single-Center Study, J Child Neurol 31 (3) (2016) 328–332 Mar., doi:10.1177/0883073815592224.
[4] P. Celka, P. Colditz, A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison, IEEE Transactions on Biomedical Engineering 49 (5) (2002) 455–462 May, doi:10.1109/10.995684.
[5] J. Gotman, D. Flanagan, J. Zhang, B. Rosenblatt, Automatic seizure detection in the newborn: methods and initial evaluation, Electroencephalography and Clinical Neurophysiology 103 (3) (1997) 356–362 Sep., doi:10.1016/S0013-4694(97)00003-9.
[6] A. Liu, J.S. Hahn, G.P. Heldt, R.W. Coen, Detection of neonatal seizures through computerized EEG analysis, Electroencephalography and Clinical Neurophysiology 82 (1) (1992) 30–37 Jan., doi:10.1016/0013-4694(92)90179-L.
[7] B. Olmi, L. Frassineti, A. Lanata, C. Manfredi, Automatic Detection of Epileptic Seizures in Neonatal Intensive Care Units Through EEG, ECG and Video Recordings: A Survey, IEEE Access 9 (2021) 138174–138191, doi:10.1109/ACCESS.2021.3118227.
[8] W. Deburchgraeve, et al., Automated neonatal seizure detection mimicking a human observer reading EEG, Clinical Neurophysiology 119 (11) (2008) 2447–2454 Nov., doi:10.1016/j.clinph.2008.07.281.
[9] A. Aarabi, F. Wallois, R. Grebe, Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis, Clinical Neurophysiology 117 (2) (2006) 328–340 Feb., doi:10.1016/j.clinph.2005.10.006.
[10] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, EEG-based neonatal seizure detection with Support Vector Machines, Clinical Neurophysiology 122 (3) (2011) 464–473 Mar., doi:10.1016/j.clinph.2010.06.034.
[11] E.M. Thomas, A. Temko, G. Lightbody, W.P. Marnane, G.B. Boylan, Gaussian mixture models for classification of neonatal seizures using EEG, Physiol. Meas. 31 (7) (2010) 1047–1064 Jun., doi:10.1088/0967-3334/31/7/013.
[12] K.T. Tapani, S. Vanhatalo, N.J. Stevenson, Time-Varying EEG Correlations Improve Automated Neonatal Seizure Detection, Int. J. Neur. Syst. 29 (04) (May 2019) 1850030, doi:10.1142/S0129065718500302.
[13] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural Eng. 16 (3) (Apr. 2019) 031001, doi:10.1088/1741-2552/ab0ab5.
[14] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: A review, Computer Methods and Programs in Biomedicine 161 (Jul. 2018) 1–13, doi:10.1016/j.cmpb.2018.04.005.
[15] R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: review, opportunities and challenges, Briefings in Bioinformatics 19 (6) (Nov. 2018) 1236–1246, doi:10.1093/bib/bbx044.
[16] A.H. Ansari, P.J. Cherian, A. Caicedo, G. Naulaers, M. De Vos, S. Van Huffel, Neonatal Seizure Detection Using Deep Convolutional Neural Networks, Int. J. Neur. Syst. 29 (04) (May 2019) 1850011, doi:10.1142/S0129065718500119.
[17] A. O'Shea, G. Lightbody, G. Boylan, A. Temko, Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture, Neural Networks 123 (Mar. 2020) 12–25, doi:10.1016/j.neunet.2019.11.023.
[18] D. Yu. Isaev, et al., Attention-Based Network for Weak Labels in Neonatal Seizure Detection, Proc Mach Learn Res 126 (Aug. 2020) 479–507.
[19] A. Caliskan, S. Rencuzogullari, Transfer learning to detect neonatal seizure from electroencephalography signals, Neural Comput & Applic 33 (18) (Sep. 2021) 12087–12101, doi:10.1007/s00521-021-05878-y.
[20] A. Daly, A. O'Shea, G. Lightbody, A. Temko, Towards Deeper Neural Networks for Neonatal Seizure Detection, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 2021, pp. 920–923, doi:10.1109/EMBC46164.2021.9629485. Nov..

[21] A. OShea, et al., Deep Learning for EEG Seizure Detection in Preterm Infants, Int. J. Neur. Syst. 31 (08) (Aug. 2021) 2150008, doi:10.1142/S0129065721500088.
[22] A.K. Abbas, G. Azemi, S. Ravanshadi, A. Omidvarnia, An EEG-based methodology for the estimation of functional brain connectivity networks: Application to the analysis of newborn EEG seizure, Biomedical Signal Processing and Control 63 (Jan. 2021) 102229, doi:10.1016/j.bspc.2020.102229.
[23] L. Stanković, M. Daković, E. Sejdić, Introduction to Graph Signal Processing, in: L. Stanković, E. Sejdić (Eds.), Vertex-Frequency Analysis of Graph Signals, Springer International Publishing, Cham, 2019, pp. 3–108, doi:10.1007/978-3-030-03574-7_1.
[24] D. Zeng, K. Huang, C. Xu, H. Shen, Z. Chen, Hierarchy Graph Convolution Network and Tree Classification for Epileptic Detection on Electroencephalography Signals, IEEE Transactions on Cognitive and Developmental Systems (2020) pp. 1–1, doi:10.1109/TCDS.2020.3012278.
[25] Z. Wang, Y. Tong, X. Heng, Phase-Locking Value Based Graph Convolutional Neural Networks for Emotion Recognition, IEEE Access 7 (2019) 93711–93722, doi:10.1109/ACCESS.2019.2927768.
[26] 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
[27] T. Zhang, X. Wang, X. Xu, C.L.P. Chen, GCB-Net: Graph Convolutional Broad Network and Its Application in Emotion Recognition, IEEE Transactions on Affective Computing (2019) pp. 1–1, doi:10.1109/TAFFC.2019.2937768.
[28] T. Song, W. Zheng, P. Song, Z. Cui, EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks, IEEE Transactions on Affective Computing 11 (3) (Jul. 2020) 532–541, doi:10.1109/TAFFC.2018.2817622.
[29] J. Li, S. Li, J. Pan, F. Wang, Cross-Subject EEG Emotion Recognition With Self-Organized Graph Neural Network, Frontiers in Neuroscience 15 (2021) 689, doi:10.3389/fnins.2021.611653.
[30] T. Song, et al., Variational Instance-Adaptive Graph for EEG Emotion Recognition, IEEE Transactions on Affective Computing (2021) pp. 1–1, doi:10.1109/TAFFC.2021.3064940.
[31] M. Wang, H. El-Fiqi, J. Hu, H.A. Abbass, Convolutional Neural Networks Using Dynamic Functional Connectivity for EEG-Based Person Identification in Diverse Human States, IEEE Transactions on Information Forensics and Security 14 (12) (Dec. 2019) 3259–3272, doi:10.1109/TIFS.2019.2916403.
[32] W. Zhang, F. Wang, S. Wu, Z. Xu, J. Ping, Y. Jiang, Partial directed coherence based graph convolutional neural networks for driving fatigue detection, Review of Scientific Instruments 91 (7) (Jul. 2020) 074713, doi:10.1063/5.0008434.
[33] Q. Lian, Y. Qi, G. Pan, Y. Wang, Learning graph in graph convolutional neural networks for robust seizure prediction, J. Neural Eng. 17 (3) (Jun. 2020) 035004, doi:10.1088/1741-2552/ab909d.
[34] Y. Zhao, et al., EEG-Based Seizure detection using linear graph convolution network with focal loss, Computer Methods and Programs in Biomedicine 208 (Sep. 2021) 106277, doi:10.1016/j.cmpb.2021.106277.
[35] Z. Jia, et al., GraphSleepNet: Adaptive Spatial-Temporal Graph Convolutional Networks for Sleep Stage Classification, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, Jul. 2020, pp. 1324–1330, doi:10.24963/ijcai.2020/184.
[36] R. Anirudh, J.J. Thiagarajan, Bootstrapping Graph Convolutional Neural Networks for Autism Spectrum Disorder Classification, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 3197–3201, doi:10.1109/ICASSP.2019.8683547.
[37] S. Zhang, D. Chen, Y. Tang, L. Zhang, Children ASD Evaluation Through Joint Analysis of EEG and Eye-Tracking Recordings With Graph Convolution Network, Front Hum Neurosci 15 (May 2021) 651349, doi:10.3389/fnhum.2021.651349.
[38] X. Zhao, F. Zhou, L. Ou-Yang, T. Wang, B. Lei, Graph Convolutional Network Analysis for Mild Cognitive Impairment Prediction, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 1598–1601, doi:10.1109/ISBI.2019.8759256. Apr.
[39] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future," *arXiv:2105.13137 [cs, q-bio]*, May 2021, Accessed: Sep. 03, 2021. [Online]. Available: http://arxiv.org/abs/2105.13137
[40] J.-P. Lachaux, E. Rodriguez, J. Martinerie, F.J. Varela, Measuring phase synchrony in brain signals, Human Brain Mapping 8 (4) (1999) 194–208, doi:10.1002/(SICI)1097-0193(1999)8:4⟨194::AID-HBM4⟩3.0.CO;2-C.
[41] G.C. Carter, Coherence and time delay estimation, Proceedings of the IEEE 75 (2) (Feb. 1987) 236–255, doi:10.1109/PROC.1987.13723.
[42] N.J. Stevenson, K. Tapani, L. Lauronen, S. Vanhatalo, A dataset of neonatal EEG recordings with seizure annotations, Sci Data 6 (1) (Mar. 2019) 190039, doi:10.1038/sdata.2019.39.
[43] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," in *Advances in Neural Information Processing Systems*, 2016, vol. 29. Accessed: Sep. 03, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2016/hash/04df4d434d481c5bb723be1b6df1ee65-Abstract.html
[44] L. Frassineti, A. Parente, C. Manfredi, Multiparametric EEG analysis of brain network dynamics during neonatal seizures, Journal of Neuroscience Methods 348 (Jan. 2021) 109003, doi:10.1016/j.jneumeth.2020.109003.
[45] D.R. Brillinger, Society for Industrial and Applied Mathematics, Time Series (2001), doi:10.1137/1.9780898719246.
[46] M.X. Cohen, Analyzing Neural Time Series Data: Theory and Practice, MIT Press, 2014.
[47] F.R.K. Chung, F.C. Graham, Spectral Graph Theory, American Mathematical Soc., 1997.
[48] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, Accessed: Sep. 03, 2021. [Online]. Available: http://arxiv.org/abs/1412.6980
[49] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. (2014).
[50] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *arXiv:1711.05101 [cs, math]*, Jan. 2019, Accessed: Sep. 03, 2021. [Online]. Available: http://arxiv.org/abs/1711.05101
[51] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," *arXiv:1811.03378 [cs]*, Nov. 2018, Accessed: Sep. 03, 2021. [Online]. Available: http://arxiv.org/abs/1811.03378
[52] M. Friedman, The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, Journal of the American Statistical Association 32 (200) (Dec. 1937) 675–701, doi:10.1080/01621459.1937.10503522.
[53] F. Wilcoxon, Individual Comparisons by Ranking Methods, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer, New York, NY, 1992, pp. 196–202, doi:10.1007/978-1-4612-4380-9_16.
[54] M.A. Tanveer, M.J. Khan, H. Sajid, N. Naseer, Convolutional neural networks ensemble model for neonatal seizure detection, Journal of Neuroscience Methods 358 (Jul. 2021) 109197, doi:10.1016/j.jneumeth.2021.109197.
[55] N. Rania, H. Douzi, L. Yves, and T. Sylvie, "Semantic Segmentation of Diabetic Foot Ulcer Images: Dealing with Small Dataset in DL Approaches," in *Image and Signal Processing*, Cham, 2020, pp. 162–169. doi: 10.1007/978-3-030-51935-3_17.
[56] W. Zhao, Research on the deep learning of the small sample data based on transfer learning, AIP Conference Proceedings 1864 (1) (Jul. 2017) 020018, doi:10.1063/1.4992835.
[57] S. Feng, H. Zhou, H. Dong, Using deep neural network with small dataset to predict material defects, Materials & Design 162 (Jan. 2019) 300–310, doi:10.1016/j.matdes.2018.11.060.