

An Effective Deep Neural Network Architecture for Cross-Subject Epileptic Seizure Detection in EEG Data

Imene JEMAL ^{a,c,1}, Amar MITICHE ^a, Lina ABOU-ABBAS ^{b,c}, Khadidja HENNI ^{b,c} and Neila MEZGHANI ^{b,c}

^a*INRS-EMT, Université du Québec, Montréal, Canada*

^b*Centre de Recherche du CHUM, Montréal, Canada*

^c*Centre de Recherche LICEF, Université TÉLUQ, Montréal, Canada*

Abstract. For several decades, the detection of epileptic seizures has been an active research topic. The performance of current patient-specific algorithms is satisfactory. However, due to significant variability of EEG data between patients, cross-subject seizure characterization and detection remains a challenging task. The purpose of this study is to propose and investigate a modified convolutional neural network (CNN) architecture based on separable depth-wise convolution for effective automatic cross-subject seizure detection. The architecture is conceived with a reduced number of trainable parameters to reduce the model complexity and storage requirements to easily deploy it in connected devices for real-time seizure detection. The performance of the proposed method is evaluated on two public datasets collected in the Children's Hospital Boston and the University of Bonn respectively. The method achieves the highest sensitivity-false positive rate/h of 91.93%–0.005, 100%–0.057 for the CHB-MIT and Ubonn datasets respectively.

Keywords. Epilepsy, Seizure detection, Deep learning, Convolutional Neural Network, EEG

1. Introduction

Epilepsy is a neuro-degenerative disorder manifested by recurrent unprovoked seizures. It is the second most frequent neurological disease [1]. Most often, EEG records are the basis for a diagnosis. The visual inspection of hours of EEG data is impractical because it is time consuming and requires interpretations by experts. Therefore, several studies have been conducted to develop computer-aided diagnostic systems that can automatically detect epileptic seizures [2, 3]. Several EEG-based epilepsy detection models have been proposed. However, epileptic patterns are highly variable across seizures and across patients, which makes real-time application of these models in clinical settings quite a challenging task. Models go along two veins: General cross-subject modeling which apply to patients at large, and patient-specific modeling which applies to patients individ-

¹Corresponding Author: Imene Jemal, INRS-EMT, Université du Québec, Montréal, Canada; E-mail: imene.jemal@inrs.ca

ually. Patient-specific modeling is generally impracticable because it requires recording sufficient seizure onsets for each patient, separately from others. Cross-subject modeling does not have to treat each patient separately, but it faces the major problem of adapting the detection algorithm to unseen data of new patients, mainly due to the significant cross-subject EEG pattern variability [4].

Recent studies of deep learning (DL) [5] for epilepsy detection, which automatically encodes EEG features characteristic of epilepsy, have been much more potent than traditional feature selection and classification methods [6]. One of the first deep learning studies on epilepsy detection [7] used a convolutional neural network (CNN) for feature extraction in an image-based representation of EEG signals, followed by Long Short Memory units (LSTM) for classification. The method was evaluated on the CHB-MIT dataset for subject-specific (sensitivity = 95-100%) and cross-subject (sensitivity = 85%) models. Along this vein, [8] used CNNs to distinguish interictal epileptiform discharges from normal activity. The method achieved the higher accuracy of 87.51% on the CHB-MIT dataset. The study in [9] compared time domain and frequency domain EEG representations in CNN feature coding for subject-specific seizure detection. On the CHB-MIT and the Freiburg datasets, frequency domain modeling performed significantly better (97.5% vs 95.4% accuracy). The investigation of [10] proposed a pyramidal one-dimensional convolution neural network architecture, achieving higher detection sensitivity, specificity, and accuracy of 89%, 99% and 98.2% respectively. The experiments, however, were carried out on a relatively small amount of data from 5 patients in the Ubonn database. Although it did not address inter-ictal and ictal period classification, the study in [11] used a relatively large EEG records dataset of 300 patients from the Temple University Hospital EEG database to compare CNN with conventional classification (support vector machine (SVM) and random forest (RF)) for distinguishing healthy from epileptic patients in cross-subject EEG data. Performance was better in terms of area under the precision-recall curve (AUPR) with the tiny visual geometry group CNN architecture (AUPR = 0.9242), than SVM (AUPR = 0.8651) and RF (AUPR = 0.8578).

Convolutional neural networks have recently received a lot of attention in the field of seizure detection. CNNs [5] can learn effective nonlinear local features of increasing complexity as processing progresses from the input layer to the output layer. CNNs were first described by [12] as neural networks composed of a sequence of convolution and pooling layers. The original CNN was subsequently upgraded to have a larger architecture, called AlexNet, which was followed by even more complex structures keeping the original basic ideas. The investigation of [13] introduced the Inception-V1 architecture (GoogLeNet) with processing steps that express correlation between channels followed by spatial pattern learning. The architecture allowed richer pattern feature learning using less network parameters. Similar in concept, the Xception architecture [14] starts with depth-wise convolution applied on channels to be followed by a point-wise convolution to combine the coded features. It has the particularity of not using non-linearity between layers. The architecture showed better performances than Inception-V3 in classification tasks on the ImageNet dataset and a larger image classification dataset comprising 350 million images and 17,000 classes.

The purpose of this study is to investigate epilepsy detection in EEG data by a novel CNN architecture based on separable depth-wise layers. Unlike others, the initial layer of this CNN performs a convolution to learn a representation of the raw signal in terms of frequency components. This is in agreement with feature extraction by filter bank sig-

nal decomposition [15]. The architecture also includes separable depth-wise layers: this necessitates significantly fewer network parameters than the standard 2D convolution layer, which has the effect of: 1) lowering model complexity and subsequent execution, 2) reducing storage requirements to allow execution on connected devices and, 3) allowing model training on either small or large datasets. This architecture, pertinent to cross-subject modeling, is investigated here to distinguish between ictal and inter-ictal periods in EEG data. The cross-subject modeling can significantly increase the applicability of the algorithm because it allows processing of data from unseen subjects, unlike models that are patient-specific.

This CNN architecture was evaluated using the publicly available CHB-MIT and Ubonn databases. As described in detail subsequently, it reached high performance, with 91.82% (5 patients in CHB-MIT) and 99.60% (23 patients in Ubonn).

The remainder of this paper is organized as follows. Section 2 presents the method in detail. The experimental setup is presented in Section 3. Finally, Section 4 contains the results and a discussion.

2. Methods

In this section, we first explain the difference between standard convolution and depth-wise separable convolution operations. Afterward, a detailed description of the deployed architecture is presented.

2.1. Standard vs Separable depth-wise convolution

A standard convolution layer simply applies a convolution operation between the input and learnable weighted filters to obtain a new representation of the data called a feature-map (See figure 1a). The purpose of the learned filters is to simultaneously capture spatial correlation and cross-channel correlation.

A separable convolution layer splits the convolution kernel into two smaller kernels, which has the effect of reducing the number of parameters. A classic example is the decomposition of the Sobel edge detector kernel into two smaller kernels as shown in the following Equation 1:

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \quad (1)$$

The depth-wise separable convolution [16] is similar to the separable convolution. However, the convolution operation is decomposed differently into two steps : (1) a depth-wise convolution to learn local patterns for each channel and (2) a point-wise convolution to find linear combination of the extracted feature capturing the correlations between the channels (See figure 1b).

- Depth-wise convolution : This type of layer is so named because it takes into consideration the depth dimension (the number of channels) where the convolution of each channel with a different kernel is performed separately as shown in Figure 1b. This step allows learning filters for each channel.

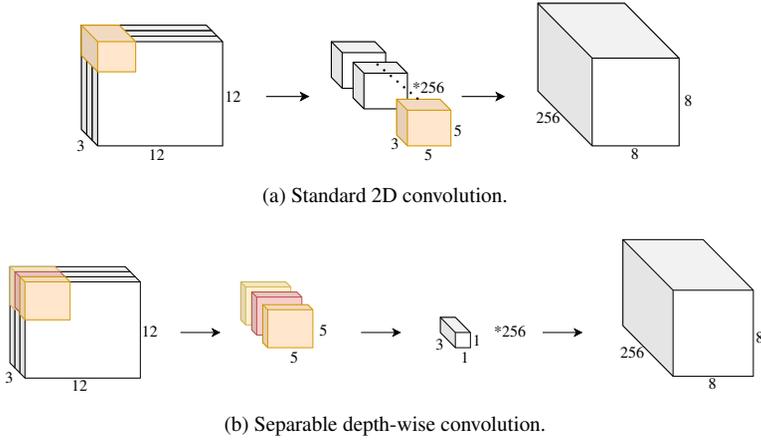


Figure 1. Illustration of different convolution layer types.

- **Point-wise convolution :** This convolution uses a 1×1 kernel with depth equal to the number of channels to iterate through each point to learn a linear combination of the feature-maps (output from the depth convolution). This step allows capturing the correlation across the channels.

Figure 1 shows an example of standard and separable depth-wise convolution. Normal convolution transforms the input 256 times using Kernels of size $5 * 3 * 3$ which lead to a total number of parameters of $5 * 3 * 3 * 256$. On the contrary, separable depth-wise convolution applies a single transformation (kernel of size $5 * 3 * 3$) and simply elongate it to 256 channels using 256 kernels of size $3 * 1 * 1$. The number of parameters is reduced to $5 * 3 * 3 + 3 * 1 * 1 * 256$.

2.2. Architecture design

In this section, we introduce a convolution neural network architecture inspired by [15, 17]. The architecture uses a reduced number of parameters, allowing it to be trained with very limited data as well as with larger data sets. Full details of the network architecture are summarized in the table 1.

- The network starts with a 2D convolution to learn F1 frequency filters. Indeed, this block is inspired by the concept of the filter bank, which is a set of band-pass filters that separate the input signal into several components, each corresponding to a single frequency sub-band of the original signal. This technique is usually performed before the feature extraction step [15].
- Subsequently, we alternate between separable depth-wise convolution layers and pooling layers. As mentioned above, the separable depth-wise convolution begin with the depth-wise convolution to learn specific filters for each channel (signal component in a specific frequency sub-band). This is followed by a point-wise convolution combining the learned features across channels. Pooling layers are used to reduce the dimensions. We applied batch normalization before the non-linear activation to stabilize the training. In order to regularize the model, we added a dropout layer.

- Finally, the features are flattened by the fully connected layer and fitted to a soft-max classification.

Table 1. The detailed network architecture, where C = number of channels, T, T', T'' = signal duration, F1, F2, F3 = number of convolution kernels to learn, N = number of classes, respectively.

Layer	#Filters	Size	#Output	Activation
Input			(1,C,T)	
Conv2D	F1	(1,128)	(F1,C,T)	Linear
BatchNorm2d			(F1,C,T)	
Reshape			(F1*C,1,T)	
DepthwiseConv	F2	(1,32)	(32*F2*F1*C,1,T')	Linear
PointwiseConv	2	(1,1)	(2,1,T')	Linear
BatchNorm2D			(2,1,T')	
Activation			(2,1,T')	Relu
AveragePool2D		(1,8)	(2,1,T''//8)	
Dropout			(2,1,T''//8)	
DepthwiseConv	F3	(1,16)	(16*F3*2,1,T'')	Linear
PointwiseConv	2	(1,1)	(2,1,T'')	Linear
BatchNorm2D			(2,1,T'')	
Activation			(2,1,T'')	Relu
AveragePool2D		(1,4)	(2,1,T''//4)	
Dropout			(2,1,T''//4)	
Linear			(2*T''//4)	
Dense			N=2	Softmax

3. Experimental setup

3.1. EEG data

We evaluated the proposed architecture on two publicly available datasets of EEG data (see Table 2).

3.1.1. CHB-MIT database

The database collected at the Boston Children's Hospital [18] contains scalp EEG data of 23 patients. The EEG data were recorded through 19 electrodes distributed over the scalp according to the international standard 10/20. The signals were amplified and sampled with a frequency of 256 Hz. During 940 hours of EEG recording, 198 epileptic seizures were recorded. The time of onset of a seizure and its duration has been identified by experts. Signals were filtered to remove artifacts and noise. We divided the records into 5-seconds-long non-overlapping windows. The selected window length yielded 601,689 and 2,157 inter-ictal and pre-ictal samples respectively.

3.1.2. UBonn university database

The data were collected from 5 monitored patients at Bonn University [19]. It consists of five sets (denoted A-E) each containing 100 single-channel EEG segments of 23.6 seconds. Sets A and B contain surface EEG recordings from healthy people. Sets C and D were recorded from epileptic patients in seizure-free intervals. Set E is the only set that contains activity recorded during seizures. The data was sampled at a rate of 173.61 Hz. The segments were selected after a visual inspection for artifacts like muscle activity or eye blinking. For classifying seizure-free and seizure EEG segments, set A-D are labeled as normal EEG records and set E is reserved for seizure events.

Table 2. Public databases for seizure detection

Dataset	CHB-MIT	UBonn
Number of subjects	23	5
Number of seizure	198	100*
Total duration(hour)	940	3.24
Recording type	Scalp	Scalp and Intracranial
Number of channels	17	1**
Sampling frequency(Hz)	256	173.73

*100 seizures file each of 23.36 s duration.

**Multi-channel data was converted to a single channel.

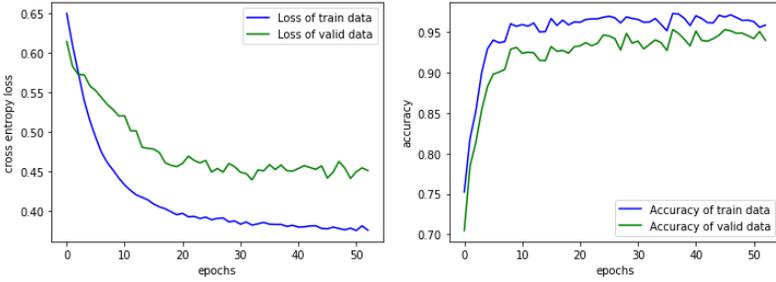
3.2. Model training

The CNN (Table 1) was trained using each dataset (Table 2) separately. Classifiers were implemented using Pytorch [20], while data pre-processing was performed using MNE-Python package [21]. The three-way holdout method was employed to optimise the performances in the hyper-parameter tuning step. In fact, we divided the data into three stratified (having same classes proportions) sets: the training set, the validation set, and the test set, each containing data from different patients. To tackle the problem of imbalanced data set, we used a cost-sensitive cross-entropy loss. The Adam optimizer was proposed as a gradient-based method with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of 0.002. We used the early stopping criteria to prevent over-fitting where training runs up to 500 epochs, or until the validation loss does not decrease anymore for at least 20 epochs.

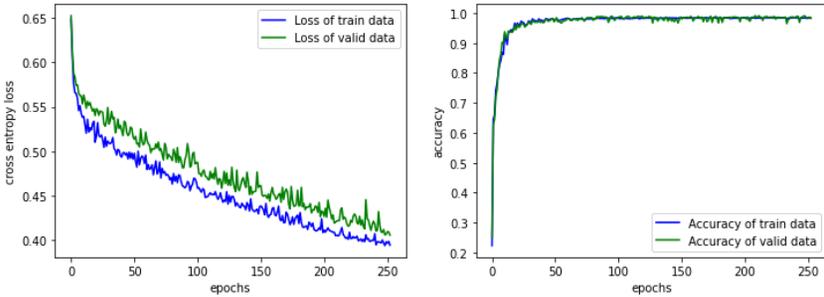
4. Results and Discussion

We evaluated the performance of the proposed cross-subject method on the CHB-MIT and UBonn university public databases. Figure 2 shows the learning curves: the cross-entropy loss and the model accuracy curves, for training and validation data, as a function of epochs. The training and validation loss decreases to a point of stability, with a small gap between the loss values indicating that the models are well fitted.

Table 3 shows the accuracy, sensitivity, F1-score, and the false positive rate per hour obtained by evaluating the neural network with unseen data from new patients. The results reveal that the proposed architecture performs very well. Indeed, for the CHB-



(a) Training and validation Loss and accuracy of the model trained on CHB-MIT dataset.



(b) Training and validation Loss and accuracy of the model trained on Ubonn.

Figure 2. Learning curves for all three datasets.

MIT data set, our method achieved a maximum accuracy rate of 91.82% with a sensitivity of 91.93% and a false alarm rate of 0.005/hour. With the Ubonn database, results were boosted to an accuracy of 99.60%, a sensitivity of 100%, and a false alarm rate of 0.057/hour.

Table 3. Model's performances for different datasets.

Database	Accuracy	Sensitivity	F-measure	FRP
CHM-MIT	91.82	91.93	95.73	0.005
UBonn	99.60	100	99.75	0.057

The results clearly show that CNNs are capable to extracting discriminative features in EEG data to allow cross-subject classification of inter-ictal and ictal data intervals. The hyper-parameters related to the network structure (number of layers, size of convolution filters), the parameters of the activation and regularization functions, as well as the training parameters (learning rate and batch size, and optimization algorithm parameters), have been carefully chosen with the three-fold out method by observing the train and validation learning curves. Although not done in this study, performance can potentially benefit from further extensive fine-tuning of the hyperparameters.

A comparison of our method with other CNN-based solutions is given in Table 4. For a fair comparison, we focused on networks evaluated on the same data sets. However, unlike our cross-subject solution, all models are patient-specific except for [10]. Regarding the CHB-MIT dataset, our method performs significantly better than the patient-specific

CNN-based model in [8]. Although the method has slightly lower accuracy than [9], it has the significant advantage of the ability to generalize across patients. In addition, the method had a better performance, although slightly, on the Ubonn university data than [10].

Table 4. Comparison of Benchmarking of recent seizure detection CNN-based studies and our work.

study	Dataset	Model	Method	Acc
[8]	CHM-MIT	patient-specific	CNN on raw EEG signals	87.51
[9]	CHM-MIT	patient-specific	CNN on frequency domain signals	95.4
This work	CHM-MIT	cross-subject	CNN with separable depth-wise convolutions	91.82
[10]	Ubonn	cross-subject	pyramidal one-dimensional CNN	98.2
This work	Ubonn	cross-subject	CNN with separable depth-wise convolutions	99.60

Thanks to the specific choice of the type of layers in our study, the proposed architecture has no more than 2,700 parameters in total. the minimal number of used parameters promotes the deployment of this model in connected devices for real-time seizure detection. In comparison, other neural networks for epileptic seizure detection [10] [11], used 8,326 and 16,401 parameters.

Overall, the classification results show that this study CNN architecture, which uses depth-wise convolution layers, performs well for epileptic seizure detection, using both a small and a fairly large database. Future work will focus on evaluating the architecture on even larger datasets, as well as applying it to other EEG-based classification tasks, such as epilepsy prediction, rather than detection, and seizure types categorization.

References

- [1] World Health Organization. Epilepsy; World Health Organization: Geneva, Switzerland, 2019. [Online]. Available : <https://www.who.int/news-room/fact-sheets/detail/epilepsy>.
- [2] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 975–982.
- [3] L. Abou-Abbas, I. Jemal, K. Henni, A. Mitiche, and N. Mezghani, "Focal and generalized seizures distinction by rebalancing class data and random forest classification," in International Conference on Bioengineering and Biomedical Signal and Image Processing, Springer, 2021, pp. 63–70.
- [4] I. Jemal, A. Mitiche, and N. Mezghani, "A study of eeg feature complexity in epileptic seizure prediction," Applied Sciences, vol. 11, no. 4, p. 1579, 2021.
- [5] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning. MIT press Cambridge, 2016, vol. 1, no. 2.
- [6] Y. LeCun, Y. Bengio et al., "Convolutional networks for images, speech, and time series," The handbook of brain theory and neural networks, vol. 3361, no. 10, p. 1995, 1995.
- [7] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in Machine learning for healthcare conference. PMLR, 2016, pp. 178–190.
- [8] A. Antoniadis, L. Spyrou, C. C. Took, and S. Sanei, "Deep learning for epileptic intracranial eeg data," in 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2016, pp. 1–6.

- [9] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic seizure detection based on eeg signals and cnn," *Frontiers in neuroinformatics*, vol. 12, p. 95, 2018.
- [10] I. Ullah, M. Hussain, H. Aboalsamhet al., "An automated system for epilepsy detection using eeg brain signals based on deep learning approach," *Expert Systems with Applications*, vol. 107, pp. 61–71, 2018.
- [11] T. Uyttenhove, A. Maes, T. Van Steenkiste, D. Deschrijver, and T. Dhaene, "Interpretable epilepsy detection in routine, interictal eeg data using deep learning," in *Machine Learning for Health*. PMLR, 2020, pp. 355–366.
- [12] Y. LeCun, L. D. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard et al., "Learning algorithms for classification: A comparison on handwritten digit recognition," *Neural networks: the statistical mechanics perspective*, vol. 261, no. 276, p. 2, 1995.
- [13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich et al., "Going deeper with convolutions," *arXiv preprint arXiv:1409.4842*, vol. 1409, 2014.
- [14] F. Chollet, "Xception: deep learning with depthwise separable convolutions (2016)," *arXiv preprint arXiv:1610.02357*, 2016.
- [15] R. T. Schirrmester, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggenberger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [16] L. Sifre and S. Mallat, "Rigid-motion scattering for texture classification," *arXiv preprint arXiv:1403.1687*, 2014.
- [17] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [18] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
- [19] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.
- [20] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.
- [21] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen et al., "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, vol. 7, p. 267, 2013.