



Automatic Detection of Epileptic Seizures with Recurrent and Convolutional Neural Networks

Salvador Carrión, Álvaro López-Chilet, Javier Martínez-Bernia,
Joan Coll-Alonso, Daniel Chorro-Juan, and Jon Ander Gómez^(✉)

Pattern Recognition and Human Language Technology Research Center, Universitat
Politècnica de València, Camí de Vera, 46022 València, Spain
{salcarpo,allochi,jamarbe2,jcolalo,dchojua,jon}@upv.es

Abstract. Computer-aided diagnosis based on intelligent systems is an effective strategy to improve the efficiency of healthcare systems while reducing their costs. In this work, the epilepsy detection task is approached in two different ways, recurrent and convolutional neural networks, within a patient-specific scheme. Additionally, a detector function and its effects on seizure detection performance are presented. Our results suggest that it is possible to detect seizures from scalp EEGs with acceptable results for some patients, and that the DeepHealth framework is a proper deep learning software for medical research.

Keywords: Deep learning · Neural networks · Epilepsy · Electroencephalogram

1 Introduction

The progress in technology and medicine has been improving humanity's quality of life over history. The 21st century is currently living a technological revolution, where Artificial Intelligence seems to be one of the most critical technologies for the future. This discipline is developing in giant steps, and the possibility of using AI techniques to help individuals in healthcare has motivated many researchers to develop systems that could improve the quality of life of many people. This work is focused on the Automatic Detection of Epileptic Seizures from scalp Electroencephalograms (EEG) and presents two deep learning based solutions to approach the problem.

Epilepsy is a chronic non-communicable disease of the brain that affects around 50 million people worldwide. According to the World Health Organization, this disease is characterised by recurrent seizures, which are episodes of involuntary movement that involve parts of the body or even the entire

This project has received funding from the European Commission - Horizon 2020 (H2020) under the DeepHealth Project (grant agreement no 825111), and the SELENE project (grant agreement no 871467).

© The Author(s) 2022

P. L. Mazzeo et al. (Eds.): ICIAP 2022 Workshops, LNCS 13373, pp. 522–532, 2022.

https://doi.org/10.1007/978-3-031-13321-3_46

body. Early detection and immediate warning of these seizures could significantly improve the quality of life of these people. Additionally, detecting these seizures with high reliability could help better understand this disease as well as open the door to improve existing treatments or computer-aided diagnosis to help neurologists.

The most common way of detecting seizures is through the analysis of the scalp Electroencephalogram (EEG), which is a non-invasive recording of the electrical activity of the brain. The main difficulty in this use case lies in the significant variability between different patients and also among the brain states of the same patient. In the EEG signals, it is possible to find different patterns corresponding to seizure periods of the same patient. Additionally, since seizures are rare events, algorithm designers must create solutions that work with limited seizure data. Nonetheless, with more research and advanced neural models, we are convinced that seizures could be detected automatically by intelligent systems with the required quality.

The main contributions of this work are the following:

- A recurrent approach to solve the seizure detection problem with raw EEG signals.
- A convolutional approach to solve the seizure detection problem with raw EEG signals.

This paper is organized as follows: Sect. 2 analyses state of the art for Epileptic Seizure Detection from EEG signals. Section 3 gives an introduction to the DeepHealth framework and its features. Section 4 describes all the parts of our experimentation setup. Section 5 shows the experiments carried out and their results. Section 6 presents our conclusions. Finally, Sect. 7 concludes the paper by giving our future research lines.

2 Related Work

Epilepsy detection using EEG signals has been approached with many different solutions, which generally rely on a first process of feature extraction and selection, and a machine learning classifier afterwards. Most of the approaches achieve high scores on detecting seizures, in terms of sensitivity and specificity. One of the first solutions that were successful was a self-organized map (SOM) neural network to detect seizures in a 24 long-term EEG recordings. This was presented in [4] in 1996 by Gabor et al., and they demonstrated the possibility of automatic seizure detection. In [6], discrete wavelet transform (DWT) was used in order to extract features from the EEG signals, and the method achieved a 76.0% of sensitivity on detecting seizures.

In [8], a novel algorithm for feature extraction called MinMaxHist is presented, with the purpose of describing the waveform characteristics of the spikes and sharp waves of the EEG signals in order to classify the events by using these kinds of features. The solution achieved 86.27% of accuracy in a patient-independent classifier. In [7], there was designed a feature vector with spectral

features using filter banks to measure the energy falling within the passband of each filter. They classified the feature vectors into ictal or interictal using a SVM and generated non-linear boundaries with a RBF kernel. Overall, 96% of the 173 test seizures were detected with an average detection latency of 4.6 s. Finally, in [3] they trained a CNN with raw signals in order to get the probability of each time window of being ictal. Afterwards, they proposed an onset-offset detector for determining the seizure onsets and offsets. The model correctly detected 90% of the seizures of the CHB-MIT Database. In [1], a deep convolutional autoencoder with a Bi-LSTM classifier achieved a 98.86% of accuracy in the CHB-MIT dataset, trained with raw signal. Additionally, an undersampling technique is applied to the Interictal class samples, in order to balance the data before feeding the network.

Regarding preprocessing techniques, in most of the works the signal is processed before using it. The common way is to extract spatial and spectral features in order to create a feature vector that is classified by a machine learning based system. In this work, the aim is to show different perspectives of approaching the task without applying any preprocessing to the data.

3 The DeepHealth Framework

The DeepHealth Framework is a flexible and scalable framework for running on HPC and Big Data environments based on two core technology libraries developed within the DeepHealth project: the European Distributed Deep Learning Library (EDDL) and the European Computer Vision Library (ECVL). These libraries will take full advantage of the current and coming development of HPC systems, and will provide a transparent use of heterogeneous hardware accelerators to optimise the training of predictive models, while considering performance and accuracy trade-offs.

Both libraries are being integrated into seven software platforms with the DeepHealth project. In a similar way, other commercial or research software platforms will have the possibility of integrating the two libraries, given that both are delivered as free and open-source software. Additionally, European large industry companies and SMEs will have the chance of using the DeepHealth Framework for training predictive models on Hybrid HPC + Big Data architectures and export the trained networks using the standard ONNX format [2], so, if their software applications already integrated another DL library able to read the ONNX format, all work is done.

The software used in this work involves the EDDL library, used to build and train the deep learning models, pyedflib to read EDF signals, and scikit-learn for some metrics calculation. The code is available in GitHub https://github.com/deephealthproject/UC13_pipeline.

4 Experimental Setup

4.1 Dataset

In the involved task of this work, the dataset used was the CHB-MIT Scalp EEG database, available on PhysioNet [5]. This dataset is composed of 24 sessions of scalp Electroencephalogram recordings from 23 pediatric subjects from the Children’s Hospital of Boston. Each patient contains between 9 and 42 continuous signal files of one hour long in most of the cases. The recordings are in EDF (European Data Format), and the sampling rate 256 Hz. The signals are composed of many channels, each one related to a pair of electrodes located on the scalp, according to the international 10–20 system standard. For this task, a total of 23 channels were used, which are the ones that repeat the most in the dataset, ignoring some extra channels in some cases and files with fewer channels. Each patient has a summary file that contains some annotations, such as the starting and ending times of the seizures. These annotations were used to label the signals into interictal and ictal segments, preparing the data for a binary classification task. An example of an EEG from subject *chb01* can be seen in Fig. 1.

The recordings of each subject were divided in three subsets, using a 65% of the recordings for training, a 15% for validation and the 20% left for testing. As seizures usually last a few seconds, the class distribution is highly unbalanced, with more than 900 h of interictal periods and just 3 h of ictal periods in the whole dataset.

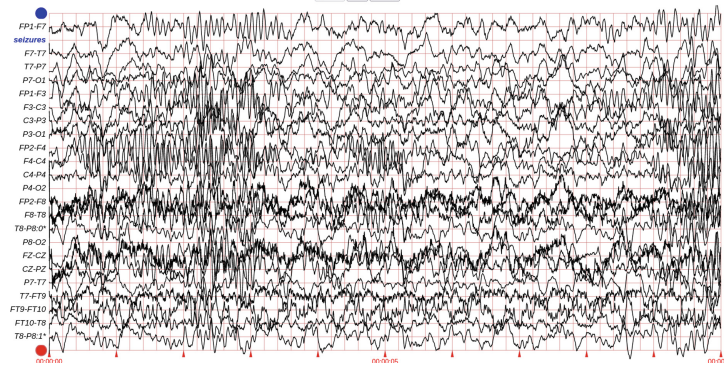


Fig. 1. This is a 10 s extract of an EEG recording from patient *chb01*. Snapshot taken from the LightWave visualizer of Physionet [5].

4.2 Training Details

Our solution for the epilepsy detection task involves a classifier based on neural networks and a post inference process that will act as a seizure detector. For

each one of the patients, a classifier will be trained only with data from the same patient, following a patient-specific approach. The first classification part will be approached by two different architectures: recurrent neural networks and convolutional neural networks.

In the recurrent approach, the input of the network are sequences of one-second-long periods extracted from raw signal. To generate each sequence, a one-second-long sliding window is shifted through the signal every 500 ms. In such a way that every single sample corresponds to a period of one second of the signal, which is overlapped at 50% with its surrounding previous and next samples.

Each time step, the network looks at a sequence of the last 10 s of the signal in order to predict the current state. In this approach, the channels that compose the signal are passed one by one by the same network, and the results are combined to make the prediction at each step. The model architecture has a single recurrent layer, followed by a fully-connected part. For the recurrent layer type, a Long Short-Term Memory (LSTM) and a Gated Recurrent Unit (GRU) were tested.

Alternatively, the convolutional approach is focused on classifying periods of 10 s extracted from the signal, by applying a first convolution along the temporal axis, similar to a Time Delay Neural Network (TDNN), followed by more convolutions with 3×3 kernels to extract features and reduce the dimensionality. The input for this second approach are periods of 10 s shifted each 0.25 s, and the signal channels are all processed together at once, giving an input size of 2560×23 , corresponding to the time dimension and the 23 channels that compose the signal. This approach applies Gaussian noise to the input, L2 regularization and Dropout, with the purpose of reducing overfitting caused by the class imbalance. In addition, in the training process of both approaches the generated mini-batches are class-balanced, applying undersampling to the interictal data and oversampling to the ictal data.

Both approaches perform a Z-score normalization to the data, but the input for the classifiers is raw signal, without any processing technique applied.

After the classification part, there is a post inference process with a function acting as a seizure detector. This function is based on a finite-state machine of two states: interictal and ictal. As the outputs of the network are ones or zeros (one for Ictal, zero for Interictal), the function analyses the predictions and makes transitions between the states, so it can raise alarms when a seizure onset is detected. It has some parameters that can be tuned to perform the detection: the size of the analysis window; α_{pos} , the minimum ratio of positive predictions needed inside the window to trigger a transition from Interictal to Ictal; α_{neg} , the maximum ratio of positive predictions needed inside the window to trigger a transition from Ictal to Interictal; and the detection threshold, which is the number of seconds that the model will be allowed to detect each seizure.

4.3 Evaluation Metrics

To evaluate our neural network models, the following metrics were monitored:

- **Accuracy**: Proportion of correct predictions among the total number of cases examined.
- **Balanced Accuracy**: Arithmetic mean of sensitivity and specificity.
- **F1-score (Macro)**: Harmonic mean of the precision and recall. Macro indicates the arithmetic mean of all the F1-scores per class.

For the detection of seizure events, the following metrics were evaluated:

- **Percentage of correctly detected seizures**: Recall associated to the detection of events.
- **Average latency**: Average time in seconds that the model needs to detect the seizures.
- **False Alarms per Hour**: Number of false alarms raised by a model per hour.

5 Experimentation

To begin with the experiments, a subset of eight patients was selected in order to find which one of the approaches performed better, as well as to find the best training configurations. This was done because of resources and time limits, as each configuration had to be trained and tested for each patient specifically. This subset was composed by patients *chb01*, *chb03*, *chb05*, *chb08*, *chb12*, *chb14*, *chb15* and *chb24*. After the best model and configuration were selected, it was tested with all of the subjects. At the end, different configurations for the post inference function parameters were analysed.

5.1 Results

First of all, both approaches (recurrent and convolutional) were trained and tested with the above mentioned subset of eight patients using SGD and Adam as optimizers and different initial learning rates, ranging from $1e-3$ to $1e-6$. The best results for these neural network classifiers are summarized in Table 1. As it can be seen, the accuracy of the models is quite high, but it is an overly optimistic metric because the dataset is unbalanced. Regarding the balanced accuracy, which is a good metric for unbalanced datasets, the best model is the recurrent one with the GRU layer trained with Adam, which achieved a 91.24%. Nevertheless, the convolutional model achieved the highest F1-Score, so it is also a possible candidate to be selected for testing with all of the patients.

In order to select between recurrent and convolutional models, the post-inference process was performed and the detection of seizures was evaluated with the best model of each approach. The results of the post-inference process for these models are summarized in Table 2. As it can be observed, the convolutional model detected 80% of the test seizures, with an average latency of 9.3s and 2.76 false alarms per hour. The recurrent model performed worse, especially in the

Table 1. Summary of the results obtained by the four configurations that best performed with the subset of eight patients. These results are the averaged values for the eight subjects.

Model	Optimizer	Accuracy (%)	F1-score	Balanced accuracy (%)
GRU	Adam (1e-6)	96.06	0.7215	91.24
GRU	SGD (1e-5)	88.24	0.5457	88.70
LSTM	Adam (1e-5)	99.10	0.7806	76.26
CONV	Adam (1e-6)	98.77	0.8084	85.31

Table 2. Results after the post-inference process of the test samples with the subset of eight patients. These results are the averaged values for the eight subjects.

Model	Optimizer	Detected seizures (%)	Avg latency (s)	False alarms per hour
GRU	Adam (1e-6)	76.67%	6.96	14.32
CONV	Adam (1e-6)	80.00%	9.30	2.76

number of false alarms per hour, with an average of 14.32. The average latency for detection was lower on this model, but there were many more false alarms.

After these first experiments with a subset of patients, the convolutional model was selected to be tested on the whole dataset. Table 3 shows the results achieved by the convolutional model for all the patients of the dataset after the post-inference process. In general, the model detected 56.67% of the seizures, with an average latency of 9.51 s and 2.15 false alarms on average. As it can be seen in Table 3, most of the patients only have one or two seizures in many hours in the test subset, so the percentage of detected seizures is very sensitive, with extreme observable values of 0% and 100%, which makes the average not very meaningful, even though it is weighted to the number of seizures of each patient.

There are 9 patients for whom the model could not detect any seizure in the test subset and 10 patients with a 100% of detected seizures. If we look at the 15 patients for whom the model could detect at least one seizure, the average rate of detected seizures is 85.93%, the average latency is 9.51 s and the average rate of false alarms per hour 2.31. Moreover, there are remarkable results for patients *chb02*, *chb05*, *chb09* and *chb10*, with a 100% of detected seizures and no false alarms. Finally, the best results in terms of accurate and fast detection were obtained with patients *chb03* and *chb10*: with 3.12 s of latency and 0.22 false alarms per hour in the case of patient *chb03*, and with 4.88 s of latency and no false alarms in the case of patient *chb10*. All the seizures of these two patients in the test subset were detected (2 seizures each).

Table 3. Results after the post-inference process with the selected model for all the subjects of the dataset.

Patient id	No. seizures	Detected seizures	Average latency (s)	False alarms per hour	Tested hours
<i>chb01</i>	2	100.00%	8.38	0.62	9.62
<i>chb02</i>	1	100.00%	6.25	0.00	8.98
<i>chb03</i>	2	100.00%	3.12	0.22	8.98
<i>chb04</i>	2	0.00%		6.86	35.16
<i>chb05</i>	1	100.00%	7.00	0.00	7.98
<i>chb06</i>	2	0.00%		3.46	15.03
<i>chb07</i>	1	0.00%		0.05	19.34
<i>chb08</i>	1	100.00%	17.00	1.40	4.99
<i>chb09</i>	1	100.00%	9.50	0.00	13.62
<i>chb10</i>	2	100.00%	4.88	0.00	13.98
<i>chb11</i>	1	100.00%	5.00	0.91	8.77
<i>chb12</i>	11	81.82%	12.72	3.18	5.98
<i>chb13</i>	4	50.00%	11.75	1.00	2.99
<i>chb14</i>	2	50.00%	9.75	9.45	6.98
<i>chb15</i>	7	57.14%	7.06	2.78	8.98
<i>chb16</i>	4	0.00%		0.00	4.99
<i>chb17</i>	1	0.00%		1.80	4.99
<i>chb18</i>	2	100.00%	11.12	4.57	8.98
<i>chb19</i>	1	0.00%		0.25	7.90
<i>chb20</i>	2	50.00%	14.00	3.84	7.54
<i>chb21</i>	1	0.00%		6.77	7.98
<i>chb22</i>	1	0.00%		0.29	6.98
<i>chb23</i>	4	0.00%		0.00	12.88
<i>chb24</i>	4	100.00%	6.31	0.32	6.28
Average		56.67%	9.51	2.15	10.00

Finally, after selecting the best model and testing it with all of the patients, an extra experiment was done in order to analyse the performance of the post-inference process function, by varying some of its parameters.

First, the length of the analysis window used in the post-inference function was in the range of 10 to 80 predictions. This length indicates how many predictions from the past the function has to use for predicting the current state. As in the case of the convolutional approach, predictions are performed every 0.25 s, so a window length of 80 will analyze the last 20 s to decide on changing of state (from interictal to ictal or vice-versa). The hyper-parameters α_{pos} and α_{neg} were set to 0.4, the same values used in all of the previous experiments. The results of this experiment can be observed in Fig. 2. It can be noticed that as the analysis window gets longer, the latency increases and the number of false alarms

is reduced. This is because the model can observe more predictions in order to predict if there is a seizure, so it is more confident on the predictions, but also slower because the latency is affected. The percentage of detected seizures is not much affected by these changes.

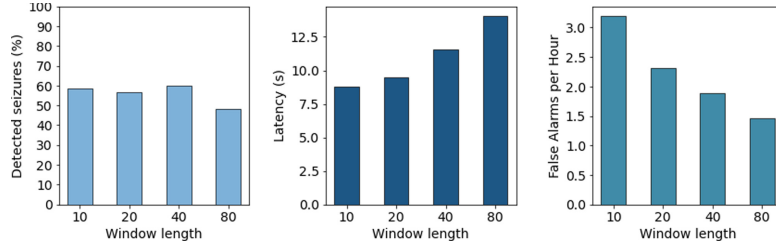


Fig. 2. Experimentation with different window lengths in the post-inference function.

The second experiment was devoted to see the effect of the hyper-parameters α_{pos} and α_{neg} on the performance. The following configurations were tested:

Configuration A: $\alpha_{pos} = \alpha_{neg} = 0.2$

Configuration B: $\alpha_{pos} = \alpha_{neg} = 0.4$

Configuration C: $\alpha_{pos} = \alpha_{neg} = 0.8$

Configuration D: $\alpha_{pos} = 0.8, \alpha_{neg} = 0.2$

Figure 3 shows the results for configurations A, B, C and D with a fixed window length of 20 predictions.

It can be observed that the configurations can vary the performance in terms of latency and false alarms. As higher are the values of α_{pos} and α_{neg} , the function needs more positive predictions inside the analysis window in order to raise an alarm, therefore the alarms are more delayed and the false alarms are

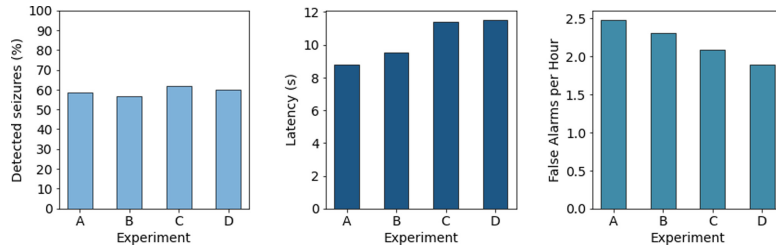


Fig. 3. Experimentation with configurations A, B, C, and D for the hyper-parameters α_{pos} and α_{neg} in the post-inference function.

reduced, because the model is more confident. Regarding the detection rate, it does not seem to be affected by these variations either.

6 Conclusions

This work was focused on Epileptic Seizure Detection and presented two approaches, one based on recurrent neural networks and the other based on convolutional neural networks. In both approaches raw data was used as input. Despite recurrent neural networks are designed to learn dependencies inherent to sequential data, the convolutional approach presented in this work achieved better results.

The task addressed in this work is challenging mainly for two reasons: (i) A strong imbalance of samples of each class in the dataset. The target class (ictal) has much fewer samples than the other class (inter-ictal). (ii) The high variability between patients. In fact, the results obtained and presented in previous sections are not consistent for all the patients; the models worked well for some patients but were not enough good for others.

A post-inference process was used in both approaches with a function acting as a seizure detector. The experimentation results show that it is possible to reach a trade-off between fast detection and a low number of false alarms by adjusting the parameters of the post-inference function.

Finally, the obtained results show that it is possible to achieve a good enough performance to detect seizures in the case of some patients, and suggest that, with more research and more data, the epilepsy detection task could be addressed with more reliable models; what could improve the quality of life of many people affected by Epilepsy. Additionally, the European Deep Learning Distributed Library (EDDL) showed to be a solid candidate to work with Deep Learning tasks.

7 Future Work

A patient-specific schema was used in this work. So that the model used for each patient was trained from scratch using only its data. This strategy led to run many experiments, and many more are still to be done, either following a patient-specific schema or a patient-independent schema. Nonetheless, our interests are in the following line of research:

- Test more models and configurations with all the subjects.
- Find a solvent feature extraction technique to improve results.
- Test our solution with other EEG datasets.
- Explore the seizure prediction task.

References

1. Ahmed, A., Magdy, B.: A deep learning approach for automatic seizure detection in children with epilepsy. *Front. Comput. Neurosci.* **15** (2021). <https://doi.org/10.3389/fncom.2021.650050>
2. Bai, J., Lu, F., Zhang, K., et al.: ONNX: Open Neural Network Exchange (2019). <https://github.com/onnx/onnx>
3. Boonyakitantont, P., Lek-uthai, A., Songsiri, J.: Automatic epileptic seizure onset-offset detection based on CNN in scalp EEG. In: *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1225–1229 (2020). <https://doi.org/10.1109/ICASSP40776.2020.9053143>
4. Gabor, A.J., Leach, R.R., Dowla, F.U.: Automated seizure detection using a self-organizing neural network. *Electroencephalogr. Clin. Neurophysiol.* **99**, 257–266 (1996). [https://doi.org/10.1016/0013-4694\(96\)96001-0](https://doi.org/10.1016/0013-4694(96)96001-0)
5. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), E215–E220 (2000). <https://doi.org/10.1161/01.cir.101.23.e215>
6. Saab, M.E., Gotman, J.: A system to detect the onset of epileptic seizures in scalp EEG. *Clin. Neurophysiol.* **116**, 427–442 (2005). <https://doi.org/10.1016/J.CLINPH.2004.08.004>
7. Shoeb, A., Gutttag, J.: Application of machine learning to epileptic seizure detection. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML 2010, Omnipress, Madison, WI, USA*, pp. 975–982 (2010)
8. Yang, S., et al.: Selection of features for patient-independent detection of seizure events using scalp EEG signals. *Comput. Biol. Med.* **119**, 103671 (2020). <https://doi.org/10.1016/J.COMPBIOMED.2020.103671>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

