UNDER REVIEW

RESCIENCE C

Replication

[Re] Deep Convolution Neural Network and Autoencoders-Based Unsupervised Feature Learning of EEG Signals

Bruno Aristimunha[1], Diogo Eduardo Lima Alves[1], Walter Hugo Lopez Pinaya[1, 2], Raphael Y. de Camargo[1]

[1]Center for Mathematics, Computation and Cognition (CMCC), Federal Univesity of ABC (UFABC), Rua Arcturus, 03. Jardim Antares, São Bernardo do Campo, CEP 09606-070, SP, Brazil. – [2]Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

Edited by
(Editor)

Reviewed by
(Reviewer 1)
(Reviewer 2)

Received
–

Published
–

DOI
–

This paper presents our efforts to reproduce and improve the results achieved by the authors of the original article. We follow the steps and models described in their article and the same public data sets of EEG Signals. Epilepsy affects more than 65 million people globally, and EEG Signals are critical to analyze and recognize epilepsy. Although the efforts in the last years, it is still challenging to extract useful information from these signals and select useful features in a diagnostic application. We construct a deep convolution network and autoencoders-based model (AE-CDNN) in order to perform unsupervised feature learning. We use the AE-CDNN to extract the features of the available data sets, and then we use some common classifiers to classify the features. The results obtained demonstrate that the proposed AE-CDNN outperforms the traditional feature extraction based classification techniques by achieving better accuracy of classification.

1 Introduction

Epilepsy is a chronic neurological disorder, and it is becoming one of the most common neurological diseases in the world [1]. Approximately 1% of the world's population is affected by epilepsy representing more than 65 million people affected [2, 3]. This disorder is characterized by the occurrence of spontaneous convulsions due to the abnormal synchronous firing of the cortical neurons [4]. This physical reaction can generate many problems for patients, including physical harm caused by the loss of consciousness, shame and discrimination [5].

Frequent seizures are dangerous conditions because, at the moment of disruption of the body can occur falls, fractures, burns, car accidents, and other serious physical injuries [6]. Epilepsy can be defined as a permanent predisposition in the brain to cause epileptic seizures [4].

To be diagnosed with epilepsy, the patient must have at least two seizures that are caused by comorbidities known in the medical literature, such as: extremely low blood sugar [7]. Even when correctly diagnosed and treated, the epileptic patient still suffers side effects and sporadic seizures. The epileptic seizures can cause even irreversible damage to the brain, and then we can visualize the importance of analyzing epilepsy to improve the life quality and the medical treatments for these patients [4].

To confirm the diagnostic, epileptologists should generally inspect the long-term electroencephalograms (EEG) of the scalp visually. EEG is a measure of the voltage fluctuation generated by the ion current of neurons in the brain, which reflects the activity of

Correspondence should be addressed to Bruno Aristimunha (b.aristimunha@gmail.com)
The authors have declared that no competing interests exists.
Code is available at https://github.com/bruAristimunha/ReScience-submission.

the brain's bio-electricity and may contain many physiological and disease information [8].
After the discovery that during a patient's seizure the brain activity changes, the EEG has become the most common epilepsy diagnostic tool. Many studies have been made, and the general problem consists in acquiring methods to classify the patients' EEG signals efficiently [9].
However, this costly task still presents several challenges for automatic crisis detection, among them: The scarce number of public data sets; The lack of standardization in seizure classification methodologies; The lack of standardization of data preprocessing; The cost of a specialist to label time intervals; The unbalance of the time series given the rare occurrence of the event; The difficulty of reproducing the works in the literature [10].
With this problem in hand, this paper reproduces the results obtained in [11], with public data labeled and preprocessed. In addition, we get new results by combining the proposal classifiers into a classifier by set voting, and we add new metrics.
The remainder of this paper is organized as follows: Section 2 presents a few works related to the classification of epileptic seizures in EGG. Section 3 introduce the methodological proposal employed, and their differences with the work of [11]. Section 4 lists the experimental validation process using epilepsy datasets. Section 5 presents the corresponding results and analyzes our approach. Finally, conclusions were summarized in Section 6.

2 Related Work

Several papers use automated methods for detecting seizures. We can extract several discriminative characteristics of the signals, among them, we mention the autocorrelation, probability of synchronization, functional connectivity network properties, EEG morphology and the reconstructed powers of the time series.
The oscillatory characteristics present in the time series of patients with epilepsy were extensively studied by temporal frequency analysis for classification [12, 13, 14, 15]. Using this technique, we mention the Discrete Wavelet Transform (DWT), which, despite requiring hand-designed parameters, it is the most used [16].
There is no standard rule for manually label seizures in databases, which makes it difficult to compare the results of these methods. Besides the few papers that use the same sets, few are looking for the same task.
In [17], it is used High Order Spectra (HOS) and spectrum-based energy resources for the automated detection of epilepsy. The proposed method yields good results when using Gaussian Mixture (GMM) for classification (93.11% and 88.78%). In [18], we have that the authors extracted the entropy of the permutation of the signals, and employed these in an SVM for classification. The result of this methodology, acquired 93.55% for our first database, in task A vs E.
Some researchers have applied Deep Belief Networks (DBNs) to the detection of seizures [19]. In the line of deep learning algorithms, Convolutional Neural Networks (CNNs) attract growing interest in the literature. In [20], they propose a CNN that learns based on the spectral information of each channel and a LSTM network with a single layer to classify the channels of the objects.
[21] propose an approach unusual when decoding each window of possible interval as an *EEG word* from the *EEG* dictionary. They explore temporal knowledge by learning context information from EEG fragments (Context-EEG). The authors obtained a 22.93% error rate in the control classification vs epileptic crisis using the second dataset present in [11].
[22] obtained 100% sensitivity with a simpler methodology than our. For each channel, it builds a autoencoder and through the error of reconstruction is classified. The dataset

was self and with no access available. [16] propose a pyramidal model of one dimension for convolution (P-1D-CNN). The method obtains 99.1 ± 0.9 in our first dataset.

3 Methodology Proposal

In this section, we describe implementation details, as the core is the reproducible aspect of our reference article. We introduce the idea and implementation of autoencoder/feature learning and our version of the model in [11], explaining the differences we have made to the original model.
In our study, we keep the autoencoder and feature learning as proposed. However, in the classification, in addition to the individual classifiers, we also employ a large ensemble learning classifier, which decides by majority vote the object class.

3.1 Implementation Details

We decided to reproduce the implementation described in the article using Keras [23] and backend in TensorFlow [24]. Our repository includes a list of all the required libraries employed in acquiring the datasets and running the model (the original and the proposed one). According to the methodology proposed in [25], we store all the checkpoints for the trained models, for reproduction purposes. Besides that, the training logs can be visualized using TensorBoard tool.
Given the lack of information about implementation in the original paper, some assumptions or cuts are made:

- The number of epoch in the AutoEnconder is assumed to be 5000;
- The number of samples per batch size is assumed to be 256;
- A column of the first database is removed, there is disagreement in the literature on the total instances, 4097 or 4096. In the specific database we use there is 4097. The removed attribute is at the endpoint of each object;
- In the second dataset, we use the channel reported by the author to train the AutoEncoder;
- The loss function presented in equation 12 of the [11] was implemented and we also compared the result obtained with *MAPE*;
- The value of the seeds selected in all classifiers, data splitting and elsewhere was 42;
- The train-validation ratio was $80\% - 20\%$ to AutoEnconder, in classifiers we use cross-validation with 5 or 10-fold;
- Given a sizing problem, we resized the values using the MinMax method, before the classification process.
- The classifier presented in the final subsection (NN2) was not reproduced for lack of information;

The experiments were performed using a CPU with Intel Core i7-5930K with 3.50 GHz and two GPUs: Nvidia Quadro K5200 and GeForce GTX 970. Some experiments were also run using Nvidia Titan X.

3.2 AutoEncoders

The autoencoder implemented is a specific case of neural network structure. It is formed by three layers, an input layer, an output layer and a hidden layer. The training is done to set the weights of the hidden layer to force the input layer and output layer to be as close to each other as possible. Our features are extracted from the hidden layer, which reduces the dimension of data.
Therefore we have a encoding process and a decoding process, and we obtain the hidden layer h by applying the encoding function:

$$h = encoder(x) = g(W * x + b), \tag{1}$$

where W is the weight matrix between the input layer and hidden layer. In the decoding function the hidden layer h is the input and $y = decoder(h)$ as output, the function is defined as follows:

$$y = decoder(x) = g(W' * x + b'), \tag{2}$$

where W' is the weight matrix between the hidden layer and output layer. Since we want the input and output to be as close as possible, we have the object function for the model training process:

$$\min \sum |y^{(i)} - x^{(i)}|, \tag{3}$$

where $y^{(i)}$ is the output signal and $x^{(i)}$ is the input signal.

3.3 Feature Learning Model

In this subsection, we will omit equations and minor details (for complete information, see [26, 22]. Since we have the dimension reduced by autoencoder we focus on the next challenge: how to obtain effective features from EEG signals. The AE-CDNN implemented follows the steps:

- Encoder: sample input, convolution layer, down-sampling layer, reshape operation, full connection layer, and the feature coding.
- Decoder: feature coding as input, full connection layer, reshape operation, deconvolution layer, up-sampling layer and the reconstruction samples.

Basically, the convolution layer acts as our feature extractor. It performs many successive convolution calculations of the input data and the expectation is to maintain the main components of the input data. The pooling layer is a down-sampling method which reduces data dimension. It uses windows to slide and extract the feature maps. These intervals do not overlap each other, and with then we obtain the pooled feature maps. The feature sizes tested were $m \in \{2, 4, 8, 16, 32, 64, 128, 256\}$[1].
The convolution and pooling operations can be iterated multiple times. Reshape operation uses the pooled feature maps to construct an one-dimension vector and a full-connection layer to transform this one-dimension vector.
Considering x as the input and y as the output, now we need to re-transform the one-dimension vector which will generate the y output, recall we want to minimize the difference between x and y and we have the following equation to calculate loss Mean Absolute Error:

$$\text{Loss MAE} = \frac{1}{N} \sum_{i=1}^{N} |x^{(i)} - y^{(i)}|.$$

[1] Size $m = 256$ has not been tested in [11].

In addition, given the possible interpretations in the original text, we have also used/implemented two loss functions, namely Mean Absolute Percentage Error - MAPE and Mean Absolute Average Error - MAAE [2], that are contained below:

$$\text{Loss MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|x^{(i)} - y^{(i)}|}{x^{(i)}}.$$

The difference between the loss functions is only in the fact that one takes in the denominator the value per $x^{(i)}$ and the other takes the average $\bar{x}^{(i)}$.

$$\text{Loss MAAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|x^{(i)} - y^{(i)}|}{\bar{x}^{(i)}}.$$

3.4 Classification

Since we have extracted the features with reduced dimension, we use supervised learning models on these features in order to classify the EEG signals. We evaluate each classifier and then we compare the results obtained with each one. The classical classifiers used are: K-Nearest Neighbors (K-NN), Support-Vector Machine - Linear Kernel and Radial Basis Kernel (SVM1, SVM2), Decision Tree (DT), Random Forest (RT), Multilayer Neural Network (MLP), Adaptive Boosting Algorithm (ADB) and Gaussian Naive Bayesian (GNB).
The proposed modification combines these classifiers and creates a single classifier that decides by voting. In short, the classifiers were combined by ensemble learning, and the result of the classification became the classification most voted by the classifiers.

4 Experimental Methodology

In this paper, as in our reference paper [11], we use unsupervised learning method in EEG signals in order to obtain useful features. This process is needed because the original data is high-dimensional. By using the auto-encoder, we can extract features with reduced dimension.

4.1 Bonn University EEG database

We can use different approaches to detect epileptic crisis. Then, to acquire a comparative measure, we verify our outputs using the method described in 3 and the original one showed in [11]. This database is public and was published by [27]. The study groups were the control, inter-ictal and ictal distributed into five sets (denotated A-E). Each containing 100 records of 23.6 seconds duration and frequency of 173.6 Hz on a single channel, with 12-bit resolution. Each data segment has 4097 samples. These recordings underwent a pre-processing in which the signals had a band filter between 0.53 to 40 Hz. There was also the removal of artifacts such as muscle movements or flicker movements.
Using labels A, B, C, D and E for the subsets, we have that A and B contain records of 5 healthy volunteers. Set A corresponds to open-eye activity and subset B to closed-eye activity. The subsets C and D have signals during the absence (interictal epileptiform activity) of 5 epileptic patients. And E records the signals during epileptic patients' seizure (ictal intervals). According to [28], this dataset is a compilation of recordings under different conditions.

[2] The formula presented in the original article by [11] differs from the MAPE formula, despite having similar intuitions. Thus, we chose to implement this loss equation, and we have not found its use elsewhere.

4.2 Children's Hospital of Boston EEG database

The second database, also public, contains the EEG signals from a Childrenѝ Hospital of Boston [26]. It was recorded by measuring the brain's electrical activity to obtain EEG signals by connecting multiple electrodes to the patients' scalp. The data incorporates the EEG signals of 23 children with refractory epilepsy.
This database, built in partnership with the Massachusetts Institute of Technology (MIT), has 5 men and 18 women between 3 and 22 years. The frequency range was 256 Hz with 16 resolution bits. Most patients contain 23 channels and some with 24 channels. In contrast to the first set of data, we have multiple channels here, then we need to select channels. The selection followed the methodology used in [29], which analyzes the variance of each patient, and after that, chooses the channel of greater variance to represent that individual. The channel reported by the authors was FT9-FT10.
In the data of the first ten patients, 200 windows of the same size of the control set were chosen from the epileptic patients we choose 200, with size of 4096, in the same way of the control group.

4.3 Performance Measures

According to [30], most of the state-of-the-art systems for epilepsy use the metrics defined below. The adaptation of these metrics for evaluating our system contributes to fair comparison with state-of-the-art systems. The definitions of these metrics are given in Table 1.

Acurracy	Precision	Specificity	Sensitivity	F-Measure
$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TN}{TN+FP}$	$\frac{TP}{FN+TP}$	$\frac{2 \cdot Precision \cdot Sensitivity}{Sensitivity+Precision}$

Table 1. Use of Metrics and Definition in our paper. Only the Acurracy was considered in [11].

where False Negatives - FN is the number of epileptic cases, which are predicted as control, True Positives - TP is the number of epileptic cases, which are predicted as epileptic, True Negative - TN is the number of control case that is predicted as control and False Positives - FP is the number of control cases that are identified as epileptic by the system. In addition, there was also the AUC-ROC (Area Under The Curve - Receiver Operating Characteristic) defined as the cumulative distribution function of the true positive rate vs the false-negative rate denoted by a threshold.

5 Results and Discussion

In this section, we analyze three analytical approaches. In the first subsection we analyze the variance present in the channels. The second contains the reproduction of all possible tables and figures, with a discussion of the reasons for the differences. In the third we present an extension of the results, evaluating other classification metrics, proposing a new classifier and varying the classifier parameters.

5.1 Checking the Variance

According to the original authors, the choice of the channel in the second dataset observed the variance present in the channels. For that, they followed the methodology: "1) calculate the variance of each channel in each sample, and select the channel with the maximum variance for each sample; 2) count these channels."
Thereby, we model the three interpretations of what is the sample defined by the author. In the first, we analyzed each recording file of the dataset as a sample, having an average length of 921600 referring to the recorded $3600s$. In these files, we compute and list the

electrode with more variance and discard the rest. We accumulate and count for all files. The results obtained can be seen in the Figure below:

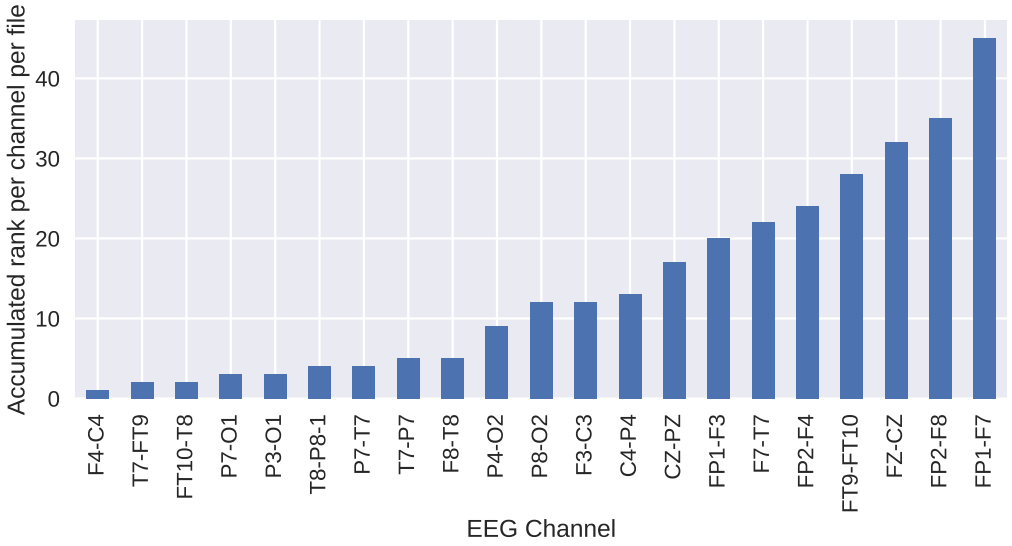

Figure 1. Accumulated variance per sample, considering a sample as each recording file.

In the second interpretation, we understand that each sample is accumulated per person with all his recordings. So the variance was calculated in parallel in the files and combined for each person. For each person, we count the occurrence of the channel with more variance. As shown in Figure 2.

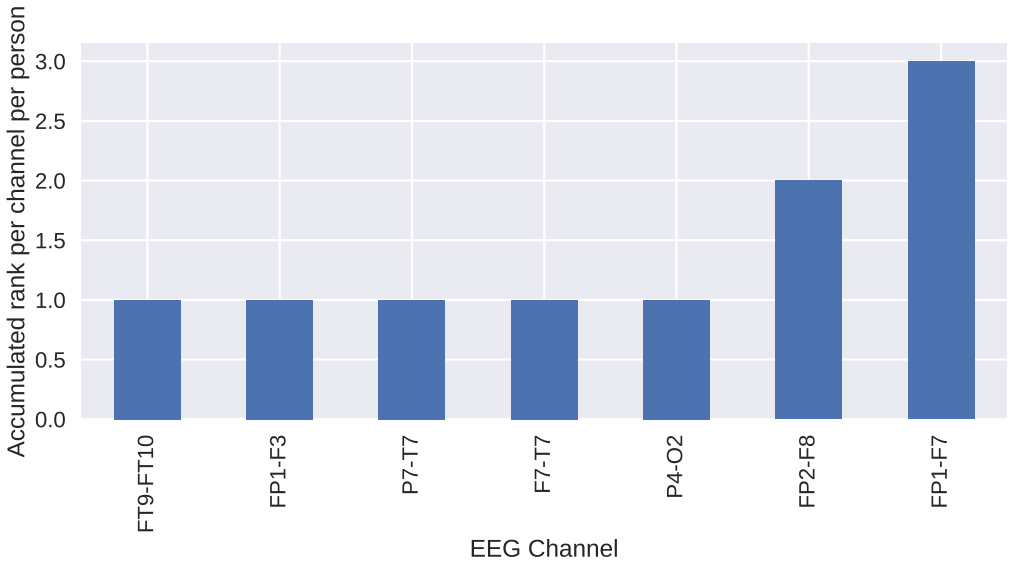

Figure 2. Accumulated variance per sample, considering a sample as being all the recordings of each person.

Finally, as a final interpretation, we calculate the cumulative variance across all people and all records, thus, we did not perform a sampling process. In other words, we put all the files together and calculate the variance as if it were a single record. For this, we compute the variance, number of points and average per channel in each file and accumulate through the cumulative variance calculation algorithm. The result of this

analysis approach can be seen in Figure 3.

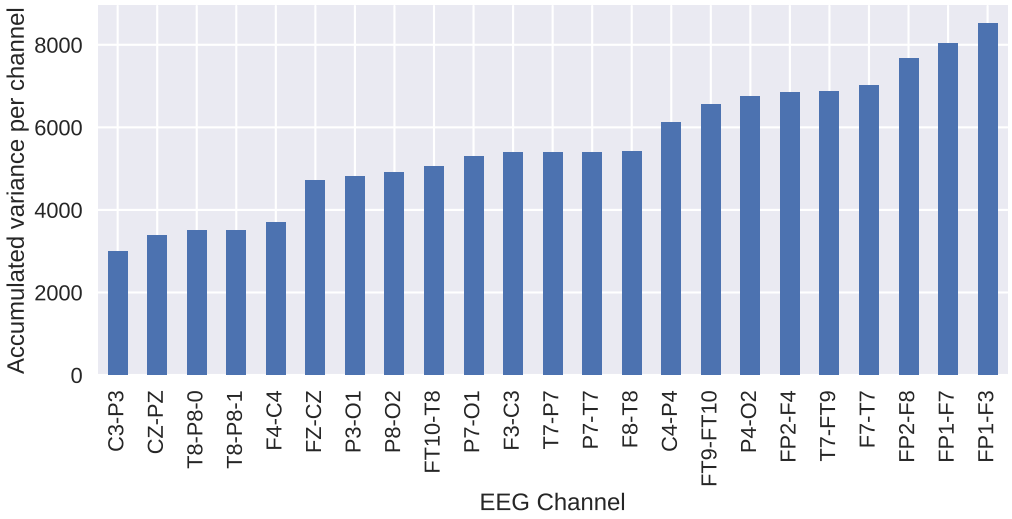

Figure 3. Total accumulated variance in the first ten people of the dataset, as granular as possible.

The results obtained in the first, second and third interpretations were not consistent with those reported by the author. There is still another possible scenario, however, not reproducible, where the authors randomly sampled the dataset and evaluated the variance.
However, given the associated uncertainty, we decided to repeat the choice of the channel $FT9 - FT10$, although this is not the one with the most variance in the modeled scenarios.

5.2 Reproduction of the values reported by the original author

The results obtained in our reproduction experiment, for the first dataset, are presented in Accuracy Tables 2 and 3 and the differences between results can be seen in Figures 4 and 5. We employed the same methodology as the one used in the original paper, performing a 5-fold cross-validation for each classifier, and we show the mean values. For each table reproduced, we also present the original result and the difference between them.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	average
2	0.480	0.600	0.600	0.600	0.600	0.600	0.600	0.400	0.56000
4	0.708	0.600	0.600	0.688	0.682	0.600	0.738	0.560	0.64700
8	0.764	0.600	0.600	0.728	0.756	0.600	0.710	0.556	0.66425
16	0.786	0.600	0.600	0.762	0.778	0.586	0.766	0.588	0.68325
32	0.794	0.600	0.598	0.722	0.812	0.660	0.806	0.638	0.70375
64	0.814	0.602	0.696	0.722	0.786	0.718	0.796	0.650	0.72300
128	0.780	0.716	0.754	0.700	0.786	0.792	0.782	0.616	0.74075

Table 2. Accuracy values obtained by the same methodology - Boon Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	average
2	0.738	0.600	0.60	0.694	0.694	0.600	0.730	0.450	0.63825
4	0.738	0.600	0.60	0.728	0.724	0.600	0.736	0.520	0.65575
8	0.712	0.600	0.60	0.694	0.658	0.600	0.742	0.562	0.64600
16	0.818	0.600	0.60	0.792	0.790	0.568	0.792	0.654	0.70175
32	0.796	0.600	0.60	0.724	0.786	0.602	0.804	0.644	0.69450
64	0.812	0.600	0.60	0.730	0.804	0.688	0.786	0.664	0.71050
128	0.790	0.616	0.72	0.726	0.758	0.750	0.772	0.620	0.71900

Table 3. Accuracy values obtained by the same methodology - Boon Dataset with MAAE.

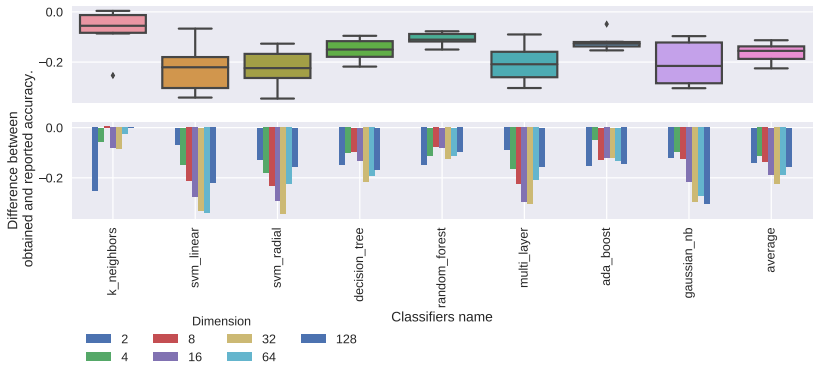

Figure 4. Classification Accuracy Results of AE-CDNN-MAE for Dataset 1 [11], Reproduced Original and Difference.

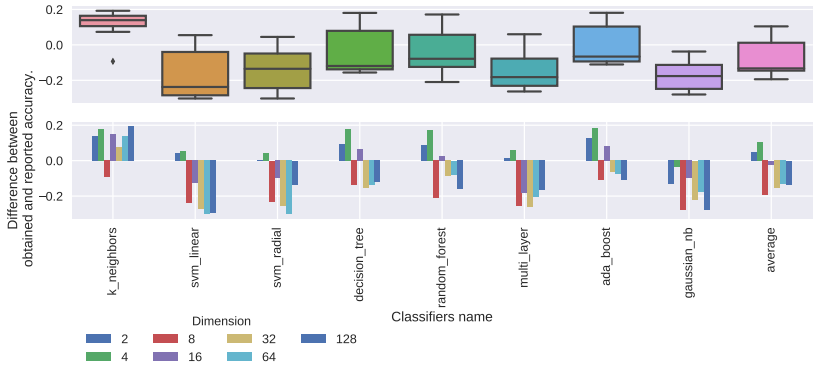

Figure 5. Classification Accuracy Results of AE-CDNN-MAAE for Dataset 1 [11], Reproduced Original and Difference.

We can perceive some differences when compared to the original results. In Table 2, we acquired the best average with a dimension equal to 64, while the original document acquired the best average when the dimension is equal to 128. The original document acquired higher accuracy values in most cases, even though when the dimension is equal to 2 or 4, our accuracy values are higher. The best precision in our article and in the original article was obtained by the random forest algorithm.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	average
2	0.7175	0.5025	0.5575	0.7400	0.6725	0.4525	0.7525	0.7700	0.645625
4	0.7700	0.5675	0.6875	0.7950	0.7525	0.7150	0.7925	0.6975	0.722187
8	0.8050	0.6050	0.7575	0.8100	0.8300	0.7650	0.8200	0.7800	0.771563
16	0.7950	0.7725	0.8600	0.7900	0.8450	0.8250	0.8200	0.8625	0.821250
32	0.7425	0.8300	0.8550	0.8325	0.8325	0.8500	0.8375	0.8625	0.830313
64	0.7650	0.8200	0.8400	0.7900	0.8375	0.8450	0.8375	0.8700	0.825625
128	0.6475	0.8200	0.8425	0.7275	0.8000	0.8275	0.7950	0.7975	0.782187

Table 4. Accuracy values obtained by the same methodology - CHBMIT Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	average
2	0.7150	0.5025	0.5800	0.7125	0.7000	0.5425	0.7000	0.7575	0.651250
4	0.8150	0.5975	0.7525	0.8150	0.8175	0.7425	0.8050	0.8050	0.768750
8	0.8300	0.5975	0.7025	0.7600	0.8100	0.7275	0.8025	0.7725	0.750312
16	0.8350	0.7475	0.8375	0.8125	0.8500	0.8225	0.8175	0.8625	0.823125
32	0.7775	0.7675	0.8300	0.7700	0.8150	0.8150	0.8150	0.8500	0.805000
64	0.7875	0.7600	0.8475	0.7900	0.8325	0.8200	0.8300	0.8650	0.816562
128	0.6500	0.8100	0.8275	0.7600	0.8100	0.8075	0.7775	0.8300	0.784062

Table 5. Accuracy values obtained by the same methodology - CHBMIT Dataset with MAAE.

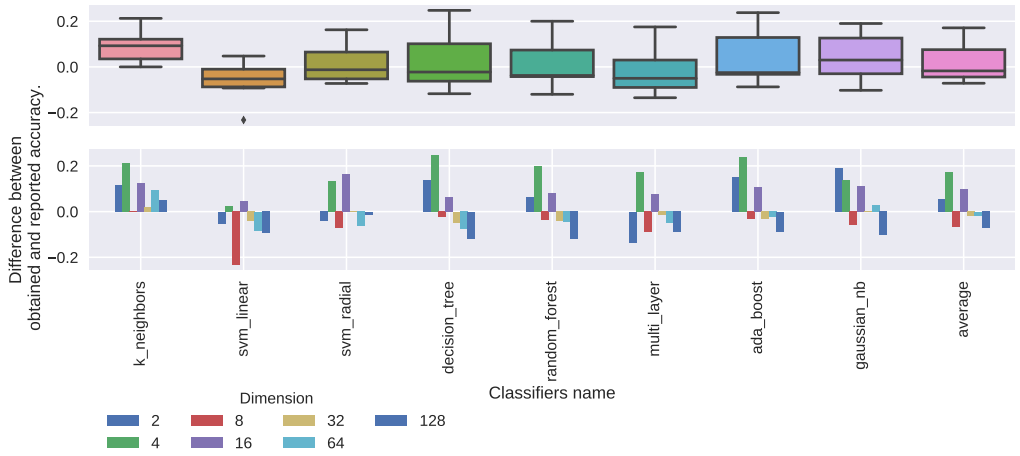

Figure 6. Classification Accuracy Results of AE-CDNN-MAE for Dataset 2 [11], Reproduced Original and Difference.

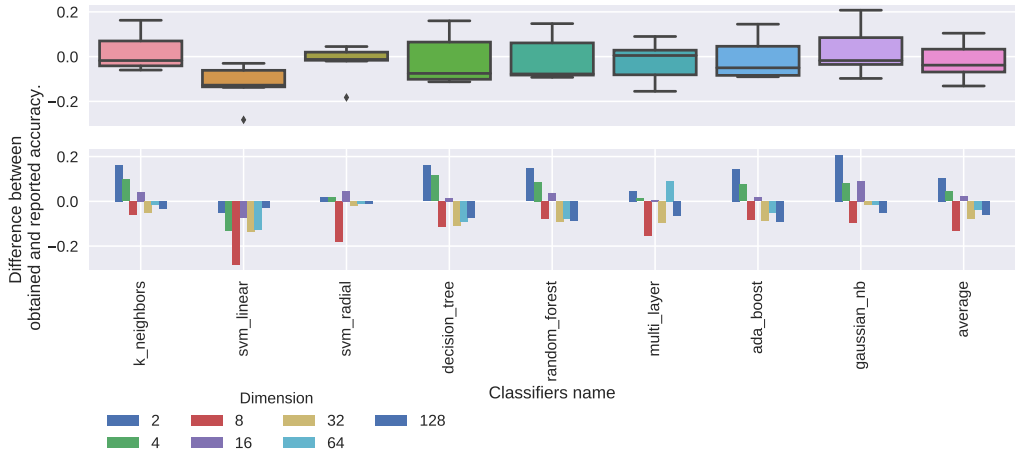

Figure 7. Classification Accuracy Results of AE-CDNN-MAAE for Dataset 2 [11], Reproduced Original and Difference.

Considering Dataset 2 and Tables 4, 5 we acquired similar results when compared with the results obtained by the original authors [11]. In general, the original paper acquired a maximum accuracy greater than those obtained by our reproduction implementation, but the average and unique values per dimension are close in most cases considering both functions AE-CDNN-MAE and AE-CDNN-MAAE. However, for Dataset 1 the accuracy values obtained in this paper are significantly lower than the ones obtained by the original paper considering both AE-CDNN-MAE and AE-CDNN-MAAE, as shown in Figure 8.

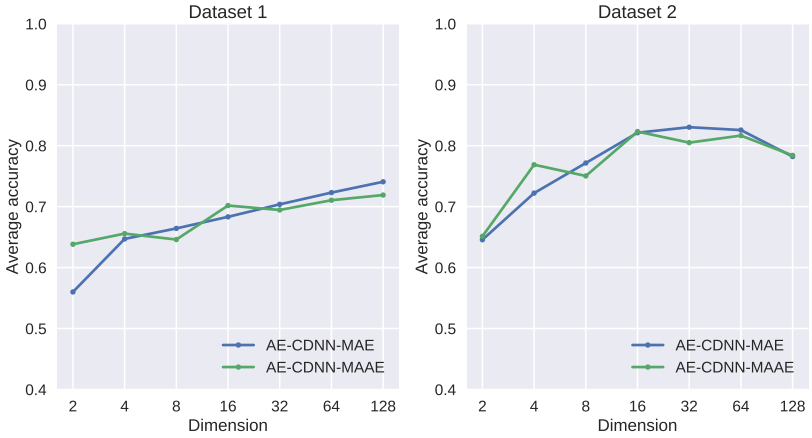

Figure 8. Average Accuracy Results of AE-CDNN-MAE and AE-CDNN-MAAE, with different dimension values in the two dataset.

When analyzing the behavior of the different loss functions, we see that the MAAE function does not always obtain superior results than the MAE function. The same is observed in the original article, as well as a similar behavior, but the average accuracy obtained are significantly higher than those obtained by our reproduction. Similarly, when we analyze the loss function MAAE and MAPE, in Figure 9, we have that the behavior of both is not very divergent, being MAAE generating a higher accuracy in the first dataset. In the second dataset, MAPE has a more stable behavior and generates greater accuracy.

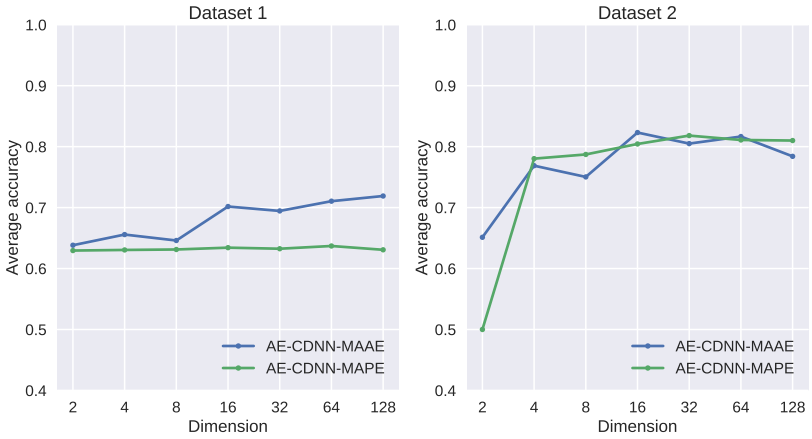

Figure 9. Average Accuracy Results of AE-CDNN-MAAE and AE-CDNN-MAPE in the two dataset.

When observing the values obtained in the classification, in the k-fold, we have that the accuracy values follow the proportion of the data, indicating the non-learning of the classification methods. We observed below the result for the accuracy inspection, for the cross validation, for $m = 2$. Analyzing the accuracy obtained by classifiers in Tables 6 and 7 we observe the values obtained by AE-CDNN-MAAE and AE-CDNN-MAE are close, however the function AE-CDNN-MAAE acquired smoothly better results and with less variation, in general.

In the original paper we observe similar differences between the two functions, the results for AE-CDNN-MAAE are smoothly better for most classifiers, but considering **gaussian_nb**, for example, the function AE-CDNN-MAAE acquired much better results comparing with AE-CDNN-MAE. Although the results in original paper also have few

5-fold	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb
1	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.4
2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.4
3	0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.4
4	0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.4
5	0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.4

Table 6. Accuracy in Classification, with loss MAE, on each fold cross-validation, for Dataset 1.

5-fold	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb
1	0.71	0.6	0.6	0.67	0.64	0.6	0.69	0.43
2	0.71	0.6	0.6	0.70	0.68	0.6	0.74	0.41
3	0.81	0.6	0.6	0.70	0.77	0.6	0.75	0.43
4	0.75	0.6	0.6	0.72	0.65	0.6	0.73	0.44
5	0.71	0.6	0.6	0.68	0.73	0.6	0.74	0.54

Table 7. Accuracy in Classification, with loss MAAE, on each fold cross-validation, for Dataset 1.

variations for the classifiers **svm_linear**, **svm_radial** and **multi_layer** we had no variance in these classifers for function AE-CDNN-MAAE.
In the second dataset, in Tables below, when analyzing by fold we have that results are worse than those reported by the authors. However, the results is consistent with the hypothesis during the process, there was no feature learning. Also given the balance of this second dataset, we have that all methods do not present a better result than the random chance.

5-fold	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb
1	0.7750	0.5000	0.7250	0.8250	0.7625	0.3250	0.8125	0.850
2	0.7125	0.5000	0.5375	0.7375	0.6875	0.4500	0.7750	0.775
3	0.7125	0.5125	0.5250	0.7125	0.6250	0.3875	0.7250	0.750
4	0.7000	0.5000	0.5000	0.7250	0.6500	0.4750	0.7500	0.775
5	0.6875	0.5000	0.5000	0.7000	0.6375	0.6250	0.7000	0.700

Table 8. Accuracy in Classification, with loss MAE, on each fold cross-validation, for Dataset 2.

5-fold	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb
1	0.7875	0.5125	0.7250	0.7750	0.7750	0.6250	0.7375	0.8750
2	0.6875	0.5000	0.5375	0.7250	0.7125	0.4875	0.6875	0.7375
3	0.7375	0.5000	0.5750	0.7000	0.7000	0.4125	0.7125	0.7125
4	0.6875	0.5000	0.5250	0.7000	0.6375	0.6375	0.6625	0.7500
5	0.6750	0.5000	0.5375	0.6625	0.6750	0.5500	0.7000	0.7125

Table 9. Accuracy in Classification, with loss MAAE, on each fold cross-validation, for Dataset 2.

When analyzing the reduced values by class, specifically with $m = 4$ we have 3 of the 4 attributes are 0, in the best scenario, indicating that there was no learning in Auto Encoder to distinguish the behavior by class. This bad representation of latent space occurs regardless of the loss function.

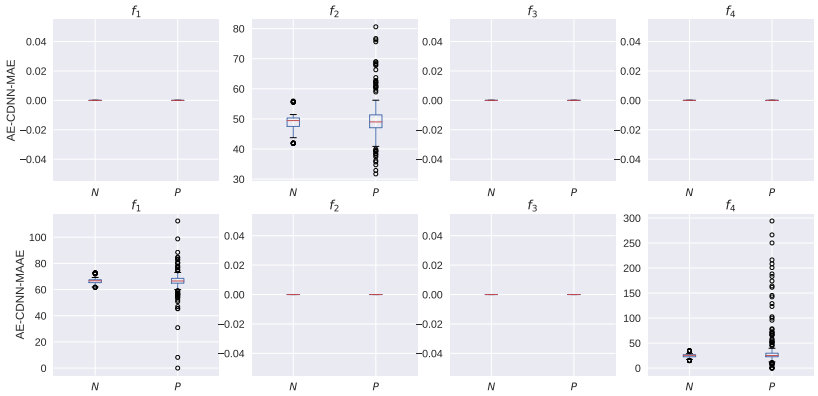

Figure 10. Feature Distribution of AE-CDNN-MAE and AE-CDNN-MAAE, with $m = 4$, in the first dataset.

When we analyze the behavior of the loss functions at the epoch, in Figure 11, in the first dataset, we have that these do not have a parallel with those reported by the original author. In addition, numerically in the second function MAAE the values also do not present an adequate dimension with that originally reported. Consequently, we also have an indication that in the MAAE function there was not an adequate generalization in the validation.

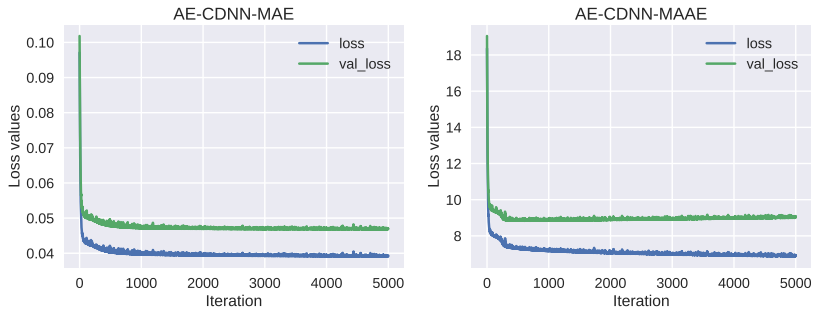

Figure 11. Change of loss function of AE-CDNN-MAE and AE-CDNN-MAAE, in the first dataset, with $m = 4$.

Even assuming that the author used the MAPE loss function, we still do not obtain an adequate result in loss at the epoch, as show the Figure 12.

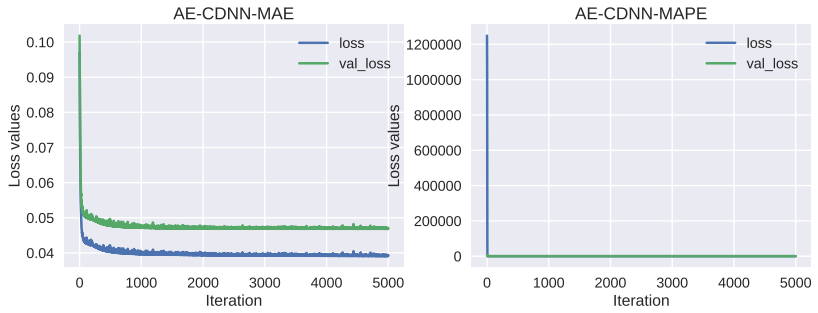

Figure 12. Change of loss function of AE-CDNN-MAE and AE-CDNN-MAPE, in the first dataset, with $m = 4$.

These differences also occur in the establishment in the baseline methods, indicating

that there is some cut in the training set that was not included in this modeling, given the lack of information in the article. In Figure 13 we observe similar average accuracy between AE-CDNN-MAE, AE-CDNN-MAAE, PCA and SRP for both datasets.

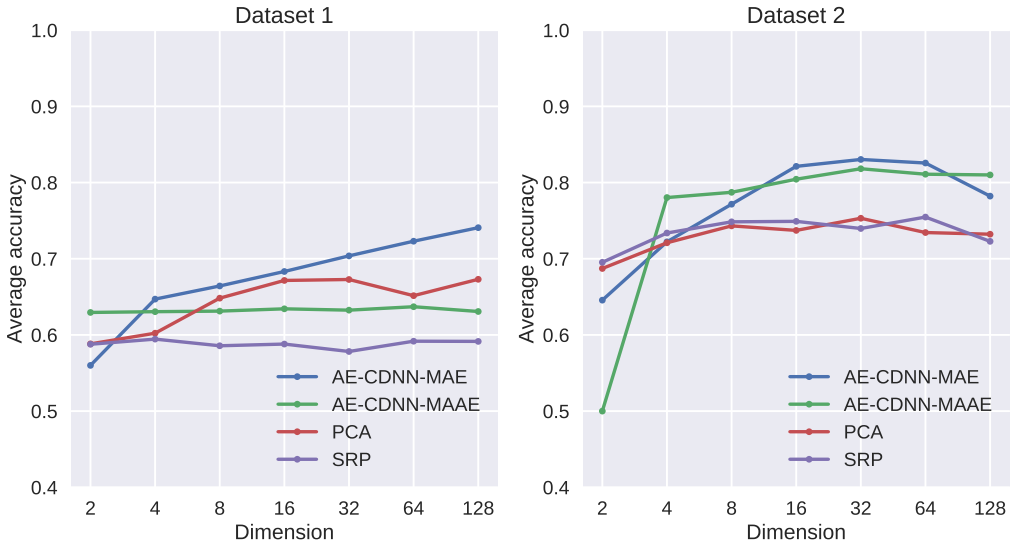

Figure 13. Comparison of accuracy for different loss functions (AE-CDNN-MAE, AE-CDNN-MAAE), and also with baseline (PCA, SRP).

The same is observed in the original article, as well as a similar behavior, but the average accuracy obtained for Dataset 1 are significantly higher than those obtained by our reproduction, as shown in the Table 10:

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	average
16	0.790	0.6	0.600	0.746	0.786	0.586	0.760	0.584	0.6815
32	0.802	0.6	0.602	0.722	0.828	0.712	0.792	0.638	0.7120

Table 10. Accuracy in Classification, in the first dataset with $CV = 10$.

When we analyze the result assuming a 10-fold, we have an increase in the accuracy values for the first dataset, however, still below that reported by the author.

5.3 Extension of the values reported by the original author

In Precision Tables 12, 12, 13 and 14, we realize that one of the most precise and specific method was the Gaussian Naive Bayesian; however, when analyzing the behavior in the Sensitivity metric, we do not have a satisfactory result. This indicates that the method pinpoints true negatives rather than true positives. If treated from a medical field, this result is worrying. The cases that the method indicates are true positives; however, this method misses many cases.

We also analyze that we cannot consider Support Vector Machine (Linear and Radial) or Multi-Layer results with the lowest m. The result in specificity indicates that the method behaves unwanted, possibly indicating all values as true positives. This rule burdens the medical system because further detection of the seizure requires further investigation for a complete diagnosis of the disease.

By our method, we note that the three metrics indicate that a progression in the number of features generates an improvement in seizure detection. Similar behavior is observed in Multi-Layer, only for Accuracy and Sensitivity, in this case, a beneficial behavior for the application. No trend was observed in the other methods. The average of the methods does not exceed our Ensemble method in almost any scenario.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.240000	0.600000	0.600000	0.600000	0.600000	0.600000	0.600000	0.000000	0.600000	0.493333
4	0.761207	0.600000	0.600000	0.723284	0.754811	0.600000	0.793074	0.733051	0.733492	0.699880
8	0.852667	0.600000	0.600000	0.803283	0.821594	0.600000	0.776664	0.851611	0.765242	0.741229
16	0.865398	0.600000	0.600000	0.795307	0.841202	0.594226	0.804707	0.990909	0.810583	0.766926
32	0.867721	0.600000	0.608802	0.861702	0.871569	0.698277	0.856168	0.975400	0.890353	0.803333
64	0.878247	0.601212	0.905826	0.807926	0.856696	0.824680	0.839116	0.932654	0.906038	0.839155
128	0.854412	0.838267	0.901891	0.792755	0.838718	0.853514	0.822159	0.933862	0.928558	0.862682
256	0.791030	0.793130	0.822870	0.770020	0.782793	0.851849	0.818501	0.935455	0.905121	0.830086

Table 11. Precision values obtained by the same methodology - Boon Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.789263	0.600000	0.600000	0.743167	0.756539	0.600000	0.789438	0.936842	0.742624	0.728653
4	0.790654	0.600000	0.600000	0.767785	0.788363	0.600000	0.777860	0.799008	0.769693	0.721485
8	0.764962	0.600000	0.600000	0.741821	0.726164	0.600000	0.794038	0.730995	0.728294	0.698475
16	0.877778	0.600000	0.600000	0.890245	0.854214	0.591486	0.844507	0.977719	0.866431	0.789153
32	0.882355	0.600000	0.600000	0.836468	0.854270	0.604293	0.860440	0.969540	0.849098	0.784052
64	0.901967	0.600000	0.600000	0.864191	0.876858	0.736459	0.831789	0.971097	0.904552	0.809657
128	0.885015	0.636666	0.903454	0.835740	0.822943	0.825939	0.816705	0.916337	0.924939	0.840860
256	0.850399	0.750629	0.905316	0.791950	0.822421	0.826755	0.834939	0.913030	0.919723	0.846129

Table 12. Precision values obtained by the same methodology - Boon Dataset with MAAE.

In the first dataset, when we analyze the accuracy we have that the naive bayes Gaussian classifier presents a drop of (40%, 23%, for first and second loss respectively) if compared to the accuracy. The average difference, in precision minus accuracy, is 6%, indicating that the precision metric achieves slightly higher results on average in the samples. The k_neighbors classifier is the classifier, in the first loss function, that has the least average difference in results, while the svm_linear method shows the same result for the second loss set. At the other end, we have the largest variation in both gaussian_nb data sets.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.735390	0.200000	0.600000	0.800320	0.689779	0.529713	0.826404	0.913222	0.843136	0.681996
4	0.785431	1.000000	0.952640	0.799541	0.776843	0.948763	0.820187	0.954135	0.942184	0.886636
8	0.838084	1.000000	0.967593	0.850238	0.868500	0.970370	0.871131	0.956970	0.966061	0.920994
16	0.892353	0.983304	0.941364	0.827188	0.854650	0.904503	0.828368	0.896774	0.927254	0.895084
32	0.901458	0.907126	0.892490	0.857564	0.849451	0.882085	0.863030	0.881194	0.893566	0.880885
64	0.931415	0.935186	0.877539	0.827526	0.848791	0.887200	0.875478	0.961510	0.943699	0.898705
128	0.885399	0.859207	0.861136	0.724203	0.806363	0.835438	0.813402	0.903969	0.873366	0.840276
256	0.765029	0.854887	0.849727	0.826371	0.821598	0.862396	0.799330	0.909664	0.897742	0.842972

Table 13. Precision values obtained by the same methodology - CHBMIT Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.735720	0.200000	1.000000	0.800840	0.724835	0.745189	0.788793	0.881378	0.861187	0.748660
4	0.836491	1.000000	0.953905	0.856105	0.854853	0.943686	0.837736	0.945054	0.941781	0.907735
8	0.858794	1.000000	0.964706	0.782093	0.854144	0.942510	0.826481	0.929300	0.939031	0.899673
16	0.890443	0.964021	0.970243	0.842738	0.865953	0.939027	0.824735	0.920818	0.938834	0.906312
32	0.867838	0.931770	0.933396	0.818189	0.819460	0.893602	0.839291	0.900392	0.902032	0.878441
64	0.968042	0.915112	0.934774	0.833354	0.839657	0.891356	0.858777	0.898607	0.919075	0.895417
128	0.933640	0.865131	0.849652	0.780718	0.818932	0.822684	0.806209	0.917503	0.872201	0.851852
256	0.919000	0.877973	0.864954	0.778029	0.805650	0.868525	0.790839	0.920900	0.885787	0.856851

Table 14. Precision values obtained by the same methodology - CHBMIT Dataset with MAAE.

Meanwhile, in the second set of data, generated by the two loss functions, the difference between precision and accuracy is greater in the smallest dimensions, while the values are more stable, and close to the accuracy values in the largest dimensions. Such stability behavior is also observed in the absolute values.

Specificity and Sensitivity — When we analyze the specificity, we have that the SVM method, with different kernels, cannot obtain a separation of the hyperspaces of the attributes to distinguish the non-schizoid events. In this way, we have that the classifier cannot distinguish when the person is without epileptic attack. From a medical point of view, there are not so many implications for this, since the weighting of importance is inclined to detect true positives. The SVM sensitivity for these cases, in high dimensions (above 32) presents reasonable values, approximately 70% in the worst scenarios. In general, the panorama of the accumulated sensitivity indicates that the worst classifiers, regardless of the number of dimensions, are the Gaussian naive bayes, and the K-neighbors

for high dimensions. The ensemble classifier has average cumulative sensitivity (71% in the worst case scenario), with the exception of lower case scenarios 2.
The performance of the Gaussian classifier may be related to the fact that the inputs are highly dependent on each other, thus violating the method's premise of independence. In the case of the k-neighbors classifier, given the presence of the values 0 in various dimensions, as shown previously, which can affect the distance assumptions necessary for the method.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.600	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.177778
4	0.645	0.000	0.000	0.540	0.660	0.000	0.695	0.770	0.545	0.428333
8	0.810	0.000	0.000	0.710	0.745	0.000	0.685	0.915	0.620	0.498333
16	0.820	0.000	0.000	0.665	0.775	0.000	0.705	0.995	0.695	0.517222
32	0.820	0.000	0.080	0.835	0.820	0.505	0.795	0.985	0.855	0.632778
64	0.830	0.005	0.915	0.745	0.800	0.785	0.765	0.955	0.890	0.743333
128	0.800	0.805	0.890	0.730	0.770	0.795	0.735	0.960	0.920	0.822778
256	0.670	0.700	0.730	0.685	0.680	0.785	0.720	0.960	0.875	0.756111

Table 15. Specificity values obtained by the same methodology - boon Dataset with mae.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.690	0.000	0.000	0.605	0.655	0.000	0.690	0.940	0.575	0.461667
4	0.690	0.000	0.000	0.635	0.700	0.000	0.665	0.900	0.610	0.466667
8	0.650	0.000	0.000	0.605	0.610	0.000	0.695	0.765	0.540	0.429444
16	0.830	0.000	0.000	0.840	0.795	0.065	0.775	0.985	0.800	0.565556
32	0.845	0.000	0.000	0.785	0.795	0.040	0.800	0.980	0.775	0.557778
64	0.870	0.000	0.000	0.835	0.835	0.585	0.750	0.980	0.875	0.636667
128	0.845	0.190	0.905	0.805	0.755	0.765	0.730	0.945	0.915	0.761667
256	0.780	0.565	0.905	0.750	0.740	0.760	0.755	0.940	0.910	0.789444

Table 16. Specificity values obtained by the same methodology - boon Dataset with maae.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.750	1.000	1.000	0.840	0.715	0.190	0.865	0.945	0.895	0.800000
4	0.795	1.000	0.980	0.800	0.790	0.975	0.830	0.980	0.970	0.902222
8	0.845	1.000	0.985	0.865	0.875	0.985	0.880	0.975	0.980	0.932222
16	0.915	0.990	0.950	0.840	0.860	0.925	0.835	0.905	0.940	0.906667
32	0.935	0.925	0.895	0.865	0.855	0.890	0.870	0.885	0.900	0.891111
64	0.950	0.955	0.890	0.845	0.845	0.900	0.885	0.970	0.955	0.910556
128	0.970	0.875	0.865	0.715	0.810	0.840	0.825	0.935	0.890	0.858333
256	0.980	0.875	0.860	0.850	0.830	0.875	0.795	0.960	0.920	0.882778

Table 17. Specificity values obtained by the same methodology - chbmit Dataset with mae.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.755	1.000	1.000	0.855	0.740	0.585	0.845	0.920	0.920	0.846667
4	0.840	1.000	0.975	0.865	0.855	0.970	0.840	0.960	0.960	0.918333
8	0.860	1.000	0.985	0.800	0.870	0.970	0.835	0.950	0.960	0.914444
16	0.910	0.985	0.980	0.850	0.870	0.960	0.830	0.930	0.955	0.918889
32	0.915	0.970	0.955	0.850	0.825	0.920	0.850	0.915	0.930	0.903333
64	0.975	0.950	0.940	0.855	0.840	0.915	0.870	0.905	0.935	0.909444
128	0.970	0.885	0.860	0.800	0.825	0.830	0.815	0.940	0.890	0.868333
256	0.970	0.895	0.880	0.795	0.810	0.875	0.785	0.945	0.905	0.873333

Table 18. Specificity values obtained by the same methodology - chbmit Dataset with maae.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.400000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000	0.822222
4	0.750000	1.000000	1.000000	0.786667	0.696667	1.000000	0.766667	0.420000	0.830000	0.805556
8	0.733333	1.000000	1.000000	0.740000	0.763333	1.000000	0.726667	0.316667	0.816667	0.788519
16	0.763333	1.000000	1.000000	0.826667	0.780000	0.976667	0.806667	0.316667	0.860000	0.814444
32	0.776667	1.000000	0.943333	0.646667	0.806667	0.763333	0.813333	0.406667	0.786667	0.771481
64	0.803333	1.000000	0.550000	0.706667	0.776667	0.673333	0.816667	0.446667	0.710000	0.720370
128	0.766667	0.656667	0.663333	0.680000	0.796667	0.790000	0.813333	0.386667	0.686667	0.693333
256	0.790000	0.760000	0.816667	0.683333	0.763333	0.820000	0.840000	0.363333	0.773333	0.734444

Table 19. Sensitivity values obtained by the same methodology - Boon Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.770000	1.000000	1.000000	0.753333	0.720000	1.000000	0.756667	0.123333	0.806667	0.770000
4	0.770000	1.000000	1.000000	0.790000	0.740000	1.000000	0.783333	0.266667	0.863333	0.801481
8	0.753333	1.000000	1.000000	0.753333	0.690000	1.000000	0.773333	0.426667	0.806667	0.800370
16	0.810000	1.000000	1.000000	0.760000	0.786667	0.903333	0.803333	0.433333	0.850000	0.816296
32	0.763333	1.000000	1.000000	0.683333	0.780000	0.976667	0.806667	0.420000	0.830000	0.806667
64	0.773333	1.000000	1.000000	0.660000	0.783333	0.756667	0.810000	0.453333	0.773333	0.778889
128	0.753333	0.900000	0.596667	0.673333	0.760000	0.740000	0.800000	0.403333	0.716667	0.704815
256	0.820000	0.743333	0.580000	0.596667	0.800000	0.766667	0.806667	0.413333	0.676667	0.689259

Table 20. Sensitivity values obtained by the same methodology - Boon Dataset with MAAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.685	0.005	0.115	0.640	0.630	0.715	0.640	0.595	0.570	0.510556
4	0.745	0.135	0.395	0.790	0.715	0.455	0.755	0.415	0.475	0.542222
8	0.765	0.210	0.530	0.755	0.785	0.545	0.760	0.585	0.600	0.615000
16	0.675	0.555	0.770	0.740	0.830	0.725	0.805	0.820	0.770	0.743333
32	0.550	0.735	0.815	0.800	0.810	0.810	0.805	0.840	0.810	0.775000
64	0.580	0.685	0.790	0.735	0.830	0.790	0.790	0.770	0.775	0.749444
128	0.325	0.765	0.820	0.740	0.790	0.815	0.765	0.660	0.750	0.714444
256	0.320	0.730	0.815	0.715	0.790	0.780	0.820	0.405	0.720	0.677222

Table 21. Sensitivity values obtained by the same methodology - CHBMIT Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.675	0.005	0.160	0.570	0.660	0.500	0.555	0.595	0.500	0.468889
4	0.790	0.195	0.530	0.765	0.780	0.515	0.770	0.650	0.630	0.625000
8	0.800	0.195	0.420	0.720	0.750	0.485	0.770	0.595	0.560	0.588333
16	0.760	0.510	0.695	0.775	0.830	0.685	0.805	0.795	0.730	0.731667
32	0.640	0.565	0.705	0.690	0.805	0.710	0.780	0.785	0.700	0.708889
64	0.600	0.570	0.755	0.725	0.825	0.725	0.790	0.825	0.745	0.728889
128	0.330	0.735	0.795	0.720	0.795	0.785	0.740	0.720	0.760	0.708889
256	0.360	0.755	0.805	0.710	0.785	0.825	0.820	0.740	0.780	0.731111

Table 22. Sensitivity values obtained by the same methodology - CHBMIT Dataset with MAAE.

F-measure and ROC-AUC — When analyzing the behavior of the methods for the reason of the metrics of $F - measure$ and $ROC - AUC$ we have that in the first data set, the SVM and multi_layer methods present the best results. At the other end, we have the appearance with the naive bayes and k neighbor methods.
We analyze the behavior of the F-measure. The relationship between sensitivity and precision is captured by this measure, as the gaussian_nb, svm_linear, k_neighbors methods did not obtain good results, which is coherent with the accuracy result. The measures generally show close results with each other.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.300000	0.750000	0.750000	0.750000	0.750000	0.750000	0.750000	0.000000	0.750000	0.616667
4	0.754107	0.750000	0.750000	0.752132	0.724419	0.750000	0.778141	0.533321	0.778392	0.730057
8	0.787506	0.750000	0.750000	0.764210	0.788092	0.750000	0.748952	0.458592	0.788159	0.731723
16	0.810292	0.750000	0.750000	0.803315	0.807892	0.738853	0.805459	0.477344	0.833610	0.752974
32	0.817637	0.750000	0.736370	0.736678	0.836645	0.726379	0.834150	0.571023	0.833943	0.760314
64	0.838025	0.750943	0.681680	0.751834	0.813110	0.738576	0.827031	0.601992	0.793788	0.755220
128	0.806969	0.730350	0.762578	0.730307	0.816380	0.819962	0.817041	0.544558	0.788074	0.757358
256	0.786350	0.775245	0.816871	0.719902	0.772494	0.835462	0.828010	0.519867	0.832356	0.765173

Table 23. F-measure values obtained by the same methodology - Boon Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.779122	0.750000	0.750000	0.746804	0.736919	0.750000	0.768445	0.175775	0.771759	0.692092
4	0.778634	0.750000	0.750000	0.773125	0.761991	0.750000	0.779028	0.398417	0.812578	0.728197
8	0.757647	0.750000	0.750000	0.744687	0.706879	0.750000	0.781574	0.538175	0.763225	0.726910
16	0.842155	0.750000	0.750000	0.811490	0.816997	0.714559	0.822670	0.597872	0.856864	0.773623
32	0.818092	0.750000	0.750000	0.742939	0.813528	0.746526	0.831861	0.582919	0.838457	0.763814
64	0.830623	0.750000	0.750000	0.744480	0.827180	0.741628	0.819404	0.616891	0.831923	0.768014
128	0.811293	0.734768	0.715060	0.744310	0.788886	0.778855	0.807601	0.557483	0.805273	0.749281
256	0.834139	0.722685	0.700275	0.677415	0.810441	0.794943	0.819765	0.566966	0.776670	0.744811

Table 24. F-measure values obtained by the same methodology - Boon Dataset with MAAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.707670	0.009756	0.171093	0.711041	0.656321	0.551665	0.718891	0.718498	0.676504	0.546827
4	0.763670	0.235636	0.543509	0.794177	0.743698	0.603795	0.783922	0.573805	0.621949	0.629351
8	0.797322	0.342395	0.680108	0.797685	0.822303	0.695209	0.809249	0.723907	0.738173	0.711817
16	0.753615	0.702287	0.839335	0.778501	0.839085	0.802547	0.812893	0.855125	0.836583	0.802219
32	0.668265	0.808371	0.842259	0.825213	0.824406	0.841946	0.827499	0.853190	0.842797	0.814883
64	0.677854	0.787736	0.823181	0.777484	0.833683	0.834291	0.824623	0.849371	0.846159	0.806043
128	0.461901	0.807834	0.831994	0.729346	0.794994	0.824191	0.784681	0.754340	0.800844	0.754458
256	0.450184	0.785377	0.824660	0.762775	0.802143	0.818116	0.808121	0.553352	0.794555	0.733254

Table 25. F-measure values obtained by the same methodology - CHBMIT Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.698258	0.009756	0.251173	0.663391	0.684510	0.463725	0.648123	0.705439	0.631161	0.528393
4	0.811006	0.320984	0.680609	0.805932	0.811680	0.664390	0.799736	0.767600	0.753578	0.712835
8	0.826095	0.321100	0.584106	0.749253	0.797223	0.637806	0.796049	0.722898	0.699157	0.681521
16	0.818304	0.661183	0.803174	0.805980	0.844897	0.788908	0.814428	0.852531	0.819065	0.800941
32	0.724628	0.697587	0.790896	0.746017	0.811508	0.789309	0.807596	0.836740	0.782786	0.776341
64	0.715778	0.698471	0.819348	0.774569	0.824374	0.796561	0.816922	0.855324	0.815164	0.790723
128	0.463376	0.793962	0.815910	0.747477	0.804942	0.801733	0.765678	0.797050	0.808231	0.755373
256	0.492048	0.808966	0.829264	0.734030	0.784070	0.845676	0.802964	0.812416	0.824258	0.770410

Table 26. F-measure values obtained by the same methodology - CHBMIT Dataset with MAAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000
4	0.697500	0.500000	0.500000	0.663333	0.678333	0.500000	0.730833	0.595000	0.687500	0.616944
8	0.771667	0.500000	0.500000	0.725000	0.754167	0.500000	0.705833	0.615833	0.718333	0.643426
16	0.791667	0.500000	0.500000	0.745833	0.777500	0.488333	0.755833	0.655833	0.777500	0.665833
32	0.798333	0.500000	0.511667	0.740833	0.813333	0.634167	0.804167	0.695833	0.820833	0.702130
64	0.816667	0.502500	0.732500	0.725833	0.788333	0.729167	0.790833	0.700833	0.800000	0.731852
128	0.783333	0.730833	0.776667	0.705000	0.783333	0.792500	0.774167	0.673333	0.803333	0.758056
256	0.730000	0.730000	0.773333	0.684167	0.721667	0.802500	0.780000	0.661667	0.824167	0.745278

Table 27. Roc-auc values obtained by the same methodology - Boon Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.730000	0.500000	0.500000	0.679167	0.687500	0.500000	0.723333	0.531667	0.690833	0.615833
4	0.730000	0.500000	0.500000	0.712500	0.720000	0.500000	0.724167	0.583333	0.736667	0.634074
8	0.701667	0.500000	0.500000	0.679167	0.650000	0.500000	0.734167	0.595833	0.673333	0.614907
16	0.820000	0.500000	0.500000	0.800000	0.790833	0.484167	0.789167	0.709167	0.825000	0.690926
32	0.804167	0.500000	0.500000	0.734167	0.787500	0.508333	0.803333	0.700000	0.802500	0.682222
64	0.821667	0.500000	0.500000	0.747500	0.809167	0.670833	0.780000	0.716667	0.824167	0.707778
128	0.799167	0.545000	0.750833	0.739167	0.757500	0.752500	0.765000	0.674167	0.815833	0.733241
256	0.800000	0.654167	0.742500	0.673333	0.770000	0.763333	0.780833	0.676667	0.793333	0.739352

Table 28. Roc-auc values obtained by the same methodology - Boon Dataset with MAAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.7175	0.5025	0.5575	0.7400	0.6725	0.4525	0.7525	0.7700	0.7325	0.655278
4	0.7700	0.5675	0.6875	0.7950	0.7525	0.7150	0.7925	0.6975	0.7225	0.722222
8	0.8050	0.6050	0.7575	0.8100	0.8300	0.7650	0.8200	0.7800	0.7900	0.773611
16	0.7950	0.7725	0.8600	0.7900	0.8450	0.8250	0.8200	0.8625	0.8550	0.825000
32	0.7425	0.8300	0.8550	0.8325	0.8325	0.8500	0.8375	0.8625	0.8550	0.833056
64	0.7650	0.8200	0.8400	0.7900	0.8375	0.8450	0.8375	0.8700	0.8650	0.830000
128	0.6475	0.8200	0.8425	0.7275	0.8000	0.8275	0.7950	0.7975	0.8200	0.786389
256	0.6500	0.8025	0.8375	0.7825	0.8100	0.8275	0.8075	0.6825	0.8200	0.780000

Table 29. Roc-auc values obtained by the same methodology - CHBMIT Dataset with MAE.

Dimension	k_neighbors	svm_linear	svm_radial	decision_tree	random_forest	multi_layer	ada_boost	gaussian_nb	ensemble	average
2	0.7150	0.5025	0.5800	0.7125	0.7000	0.5425	0.7000	0.7575	0.7100	0.657778
4	0.8150	0.5975	0.7525	0.8150	0.8175	0.7425	0.8050	0.8050	0.7950	0.771667
8	0.8300	0.5975	0.7025	0.7600	0.8100	0.7275	0.8025	0.7725	0.7600	0.751389
16	0.8350	0.7475	0.8375	0.8125	0.8500	0.8225	0.8175	0.8625	0.8425	0.825278
32	0.7775	0.7675	0.8300	0.7700	0.8150	0.8150	0.8150	0.8500	0.8150	0.806111
64	0.7875	0.7600	0.8475	0.7900	0.8325	0.8200	0.8300	0.8650	0.8400	0.819167
128	0.6500	0.8100	0.8275	0.7600	0.8100	0.8075	0.7775	0.8300	0.8250	0.788611
256	0.6650	0.8250	0.8425	0.7525	0.7975	0.8500	0.8025	0.8425	0.8425	0.802222

Table 30. Roc-auc values obtained by the same methodology - CHBMIT Dataset with MAAE.

6 Conclusion

In this article, we re-implemented the approach proposed in [11] and propose the use of a different classifier. This classification approach, based on deep learning for detecting epileptic seizures using EGG had not been explored previously. We adopted a Auto-Enconder that allowed us to construct a smaller representation space. Among the variety of metrics, using the ensemble method results in better ROC-AUC results.

The original authors left some gaps that made it impossible to fully reproduce their obtained results. For example, the lack of information about the neural classifier used in the last sub-section, about the sampling process of the first and second data sets, and the number of times or batch size. Consequently, the results obtained can be considered at most a replication.
As a second contribution, the developed codes can be easily ported to other tasks. Moreover, it could be used to evaluate other variants of the neural network architecture, techniques for classifying the signals, data augmentation, among other possibilities.

References

1. E. H. Reynolds. "The ILAE/IBE/WHO Epilepsy Global Campaign History." In: **Epilepsia** 43.s6 (2002), pp. 9–11. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1528-1157.43.s.6.5.x.
2. E. B. et. al. "Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016." In: **The Lancet Neurology** 18.4 (2019), pp. 357–375.
3. A. K. Ngugi, C. Bottomley, I. Kleinschmidt, J. W. Sander, and C. R. Newton. "Estimation of the burden of active and life-time epilepsy: A meta-analytic approach." In: **Epilepsia** 51.5 (2010), pp. 883–890. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1528-1167.2009.02481.x.
4. C. E. Stafstrom and L. Carmant. "Seizures and epilepsy: an overview for neuroscientists." In: **Cold Spring Harbor perspectives in medicine** 5.6 (2015), a022426.
5. S. V. Thomas and A. Nair. "Confronting the stigma of epilepsy." In: **Annals of Indian Academy of Neurology** 14.3 (2011), p. 158.
6. M. MOLLAOĞLU and E. BOLAYIR. "Injuries in patients with epilepsy and some factors associated with injury." In: **Nöro Psikiyatri Arşivi** 50.3 (2013), p. 269.
7. P. E. Schauwecker. "The effects of glycemic control on seizures and seizure-induced excitotoxic cell death." In: **BMC neuroscience** 13.1 (2012), p. 94.
8. E. Niedermeyer and F. L. da Silva. **Electroencephalography: basic principles, clinical applications, and related fields**. Lippincott Williams & Wilkins, 2005.
9. A. Puce and M. S. Hämäläinen. "A review of issues related to data acquisition and analysis in EEG/MEG studies." In: **Brain sciences** 7.6 (2017), p. 58.
10. A. Craik, Y. He, and J. L. Contreras-Vidal. "Deep learning for electroencephalogram (EEG) classification tasks: a review." In: **Journal of neural engineering** 16.3 (2019), p. 031001.
11. T. Wen and Z. Zhang. "Deep Convolution Neural Network and Autoencoders-based Unsupervised Feature Learning of EEG Signals." In: **IEEE Access** PP (May 2018), pp. 1–1.
12. M. Saab and J. Gotman. "A system to detect the onset of epileptic seizures in scalp EEG." In: **Clinical Neurophysiology** 116.2 (2005), pp. 427–442.
13. L. Kuhlmann, A. N. Burkitt, M. J. Cook, K. Fuller, D. B. Grayden, L. Seiderer, and I. M. Mareels. "Seizure detection using seizure probability estimation: Comparison of features used to detect seizures." In: **Annals of biomedical engineering** 37.10 (2009), pp. 2129–2145.
14. A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag. "Patient-specific seizure onset detection." In: **Epilepsy & Behavior** 5.4 (2004), pp. 483–498.
15. A. Shoeb, A. Kharbouch, J. Soegaard, S. Schachter, and J. Guttag. "A machine-learning algorithm for detecting seizure termination in scalp EEG." In: **Epilepsy & Behavior** 22 (2011), S36–S43.
16. I. Ullah, M. Hussain, H. Aboalsamh, et al. "An automated system for epilepsy detection using EEG brain signals based on deep learning approach." In: **Expert Systems with Applications** 107 (2018), pp. 61–71.
17. K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim. "Application of higher order spectra to identify epileptic EEG." In: **Journal of medical systems** 35.6 (2011), pp. 1563–1571.
18. N. Nicolaou and J. Georgiou. "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines." In: **Expert Systems with Applications** 39.1 (2012), pp. 202–209.
19. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli. "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals." In: **Computers in biology and medicine** 100 (2018), pp. 270–278.
20. R. Hussein, H. Palangi, R. Ward, and Z. J. Wang. "Epileptic seizure detection: A deep learning approach." In: **arXiv preprint arXiv:1803.09848** (2018).
21. G. Xun, X. Jia, and A. Zhang. "Detecting epileptic seizures with electroencephalogram via a context-learning model." In: **BMC medical informatics and decision making** 16.2 (2016), p. 70.

22. A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi. "Autoencoding of long-term scalp electroencephalogram to detect epileptic seizure for diagnosis support system." In: **Computers in biology and medicine** (2019).
23. F. Chollet et al. "Keras: The python deep learning library." In: **Astrophysics Source Code Library** (2018).
24. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. "Tensorflow: A system for large-scale machine learning." In: (2016), pp. 265–283.
25. A. D. la Fuente and R. Aduviri. "[Re] Variational Sparse Coding." Python. In: **ReScience C** 5.2 (May 2019), #2.
26. A. Shoeb and J. Guttag. "Application of Machine Learning to Epileptic Seizure Detection." In: ICML'10 (2010), pp. 975–982.
27. R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger. "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state." In: **Physical review. E, Statistical, nonlinear, and soft matter physics** 64 (Jan. 2002), p. 061907.
28. C. Kamath. "Analysis of EEG dynamics in epileptic patients and healthy subjects using Hilbert transform scatter plots." In: **Open Access Library Journal** 2.1 (2015), p. 1.
29. A. H. Shoeb. "Application of machine learning to epileptic seizure onset detection and treatment." PhD thesis. Massachusetts Institute of Technology, 2009.
30. Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert. "Deep learning-based electroencephalography analysis: a systematic review." In: **Journal of neural engineering** (2019).