Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 22, Iss. 2, pp. 62-72, 2024
© Brazilian Computational Intelligence Society

COMPARISON BETWEEN METHODS OF CHOOSING EEG TRAINING DATA IN EPILEPTIC SEIZURES PREDICTION SYSTEMS

Luís Otávio Lopes Amorim, **Miguel Angelo de Abreu de Sousa**
Sara Dereste dos Santos, **Ricardo Pires**
Federal Institute of Education, Science and Technology of São Paulo, Brazil
amorim.luis@aluno.ifsp.edu.br, {angelo,sarad,ricardo_pires}@ifsp.edu.br

Sigride Thome-Souza
Institute of Psychiatry, University of São Paulo, Brazil
maria.sigride@hc.fm.usp.br

Abstract – Considering the large number of people that suffer with epilepsy, there is great interest in developing a system capable of predicting the occurrence of the seizures. Warning patients about the proximity of a seizure can help them to avoid dangerous situations. Many efforts in the development of this system have been placed using, basically, machine learning techniques. In most of these works, regardless of the technique employed, it is assumed that the system is able to learn when the patient's brain signals, obtained from electroencephalogram (EEG), change from interictal to pre-ictal state. Some published works do not explicitly clarify the method of choosing training and test data. This means that good results may be due to the use of an inadequate method. That is, the EEG segments may have been temporarily mixed up during training, which would help the network correctly predict the seizure. However, this would not work from a real point of view. The objective of this work is to investigate, through experiments, the effect of the way of choosing the training and test data on the accuracy of a seizure prediction system. Experimental results showed that the accuracy of the systems increases when training is performed with windows close to the seizure used for testing.

Keywords – Epilepsy, prediction, electroencephalogram, EEG, training data, machine learning.

1 INTRODUCTION

Epilepsy is one of the most common neurological diseases in the world, affecting more than 50 million people of all ages. It is diagnosed after a person has had at least two seizures that were not due to some other medical condition [1, 2].

The electrical activity of the brain is measured in the form of a so-called electroencephalogram (EEG), normally using more than 20 electrodes positioned on the surface of the skull [3]. That activity is due to the communication between brain cells through electrical impulses. The EEG then manifests as electric waves over time, one for each electrode. Each of these electrodes and its corresponding waveform is called a channel. The EEG is used in the main diagnostic tests for epilepsy, as well as being used in the diagnosis of other brain disorders.

The conventional EEG periods in epilepsy, fragmented in time windows, are depicted in Figure 1. The ictal period is that of a seizure. Pre-ictal and post-ictal are the periods immediately before and after a seizure, respectively. Interictal are periods far from seizures [4–6].

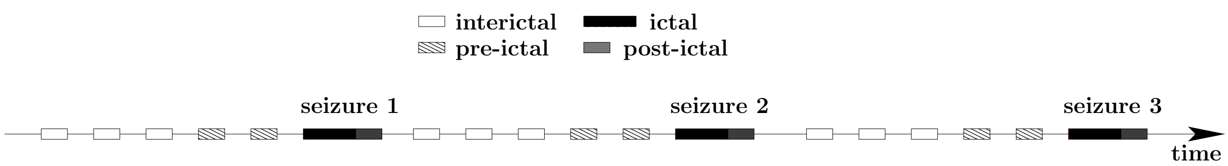

Figure 1: The EEG periods in epilepsy

The options for treating epilepsy are drugs or, in rarer cases, surgery. These drugs are not effective for a large portion of patients [7]. Therefore, there is great interest in the development of systems for predicting the occurrence of epileptic seizures, to be used by patients in their everyday lives. If such a system could warn its user about the imminence of an epileptic seizure (the transition from the interictal period to the pre-ictal period), with good reliability and well in advance, when warned, the user could put herself/himself in a safe situation, for example by ceasing to drive a car or walking down the street. For those for whom medicines are effective, these medicines could be used only after the user receives the warnings made by the system of prediction, which would be interesting for patients who have seizures rarely and to minimize possible side effects of such drugs.

A bibliographic study of the scientific production in systems for predicting epileptic seizures, detailed in Section 2, shows that many of those systems use some supervised machine learning technique, to differentiate the interictal period from the pre-ictal period.

© Brazilian Computational Intelligence Society

In Figure 2, one can see a hypothetical case in which EEG time windows were chosen for training and testing a system for predicting seizures. It can be seen that the training windows are never the same as the test windows. So this is apparently an appropriate choice for these windows.

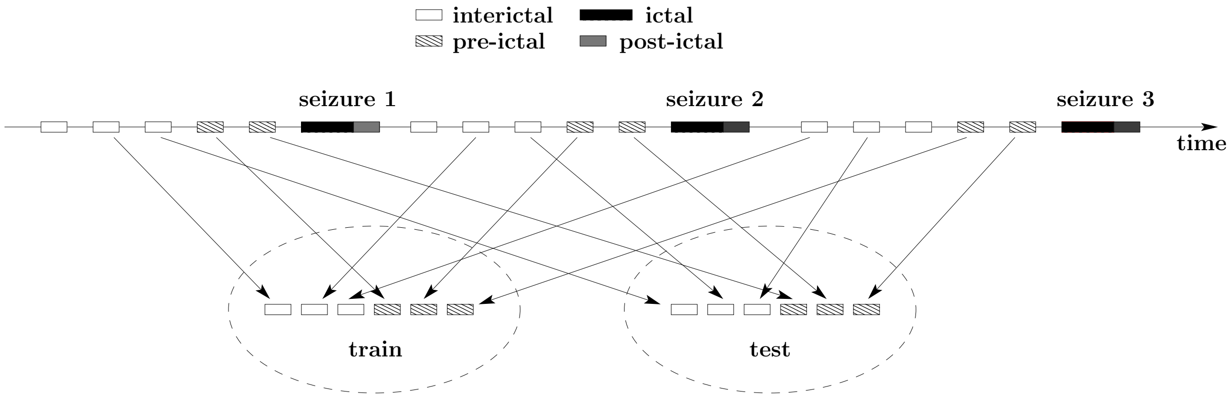

Figure 2: An apparently appropriate choice of train and test windows for training and testing a system for predicting epileptic seizures

But in Figure 3 one may see a potential problem with the choice used in Figure 2. In Figure 3, the dashed rectangle delimits a hypothetical period in which an extraneous physiological state, not related to epilepsy, is present. The rectangle contains two interictal windows, j_1 and j_2, near each other. In addition to possibly having interictal characteristics, j_1 and j_2 may have that extraneous physiological state characteristics. j_1 had been placed in the training set and labeled as belonging to the interictal class. In the test phase, the prediction system may detect the similarity between j_1 and j_2 due to that extraneous characteristics. So it probably will classify j_2 as belonging to the same class as j_1, attributing to j_2 the same interictal label attributed to j_1 in the train phase. In fact j_2 may be correctly classified not because it belongs to the interictal class, but because it has the same extraneous characteristics of j_1 and thus deserving the same label of j_1, whatever this label may be. In the test phase, this can wrongly increase the accuracy of a seizures prediction system.

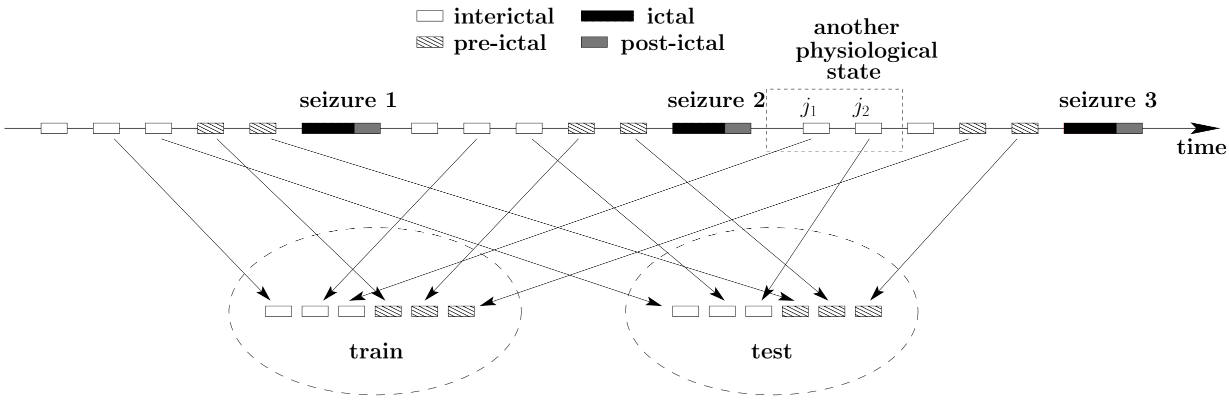

Figure 3: The problem with the choice of training and test windows in Fig. 2

Figure 4 represents what is a more appropriate way of choosing the train and test windows for prediction systems. One may see that the last available seizure in the EEG, with its preceding pre-ictal and interictal periods, are reserved solely for testing. Training windows are chosen only before other seizures. In this way, the problem shown in Figure 3 is avoided. Moreover, this is a more realistic way of choosing windows, with a view to obtaining a seizures predictor device. In this case, the device would be trained, in a laboratory, with EEG temporally close to some seizures of the user. Afterwards, in the day-to-day use of the device, there would be only periods to be classified in which no window would have been known during the system training.

From now on, the choice of training and test windows of Figure 2 will be called "method A" and that of Figure 4 will be called "method B".

Some papers published on systems for predicting epileptic seizures do not explicitly clarify if the choice of training and test windows has been made according to the method A or to the method B. This can raise the suspicion that they have obtained good results due to the use of the method A.

The objective of this work is to investigate, by means of experiments, the effect of the way of choosing the training and test windows on the accuracy of a seizures prediction system. To do this, an LSTM network is adopted for seizure prediction and only the way of choosing windows between experiments is varied, from method A to method B, in order to isolate its effect. It must be stressed that the objective here is not to create the best prediction system, but only to have the means to obtain conclusions on the effect on the accuracy of the way of choosing train and test windows. It may seem that the scope of this work is very

© Brazilian Computational Intelligence Society

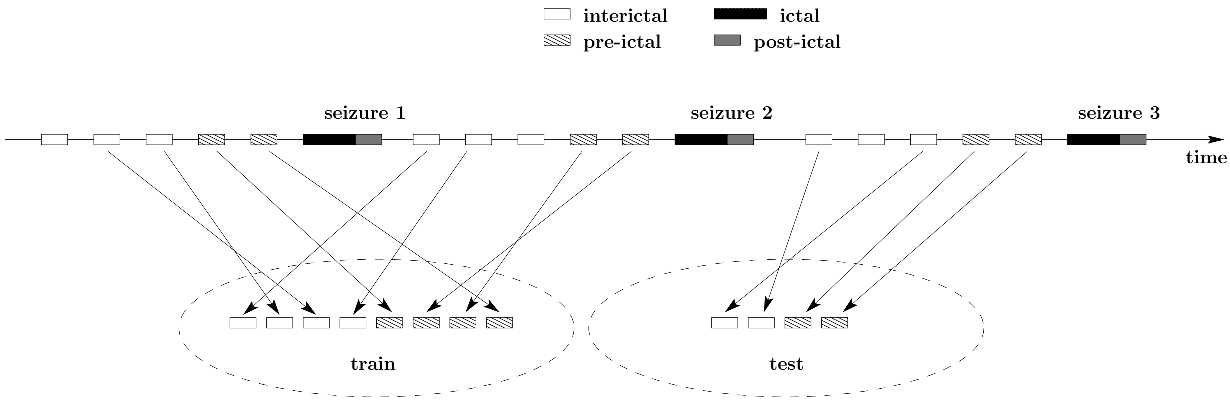

Figure 4: A more appropriate choice of training and test windows

limited. But, perhaps, the conclusions obtained here are important not only in the creation of systems for the prediction of epileptic seizures, but, in general, in the application of machine learning to EEG analysis for other purposes and in the analysis of physiological signals of other types, such as cardiac and muscular. In other words, the expected contribution of this work is to alert researchers in the area about this important issue, both in the preparation of their experiments and in the explicit indication, in future published articles, that their work used method B, considered here the more appropriate.

2 LITERATURE REVIEW

In the hope of building a system capable of alerting patients with epilepsy about the proximity of a new seizure, many works have been proposed. We limit ourselves here to evaluate systems that use EEG signals from the surface of the head (the so-called scalp EEG), for their convenience, rather than systems that use intracranial measurements. In addition, works whose focus is the detection of epileptic seizures such as [8] and [9] were also outside the scope of this work, since the objective here is to be able to predict the occurrence of a seizure in advance and not to detect an already initiated seizure. In these context, it is possible to cite cases where unsupervised learning has been applied to separate the classes by clustering the seizure and non-seizure events as presented by [10] and [11]. However, most works have used supervised machine learning techniques to accomplish the goal of distinguishing the interictal and pre-ictal periods [12–19], which is expected given that in the current context these techniques have been successfully used in a wide range of applications [20–25].

More recently, the use of Long-Short Term Memory (LSTM) networks in seizure prediction has drawn attention due to the superior results presented in some works [15, 26, 27]. An extraordinary case is the one presented in [28] in which they used LSTM architectures and obtained 100% accuracy for the 24 patients available in the Physionet CHB-MIT Scalp EEG Database described by [29, 30]. Various characteristics of the EEG segments were extracted and incorporated into the input vectors of the network, such as from cross-correlation, time domain, frequency domain and graph theory, which totaled vectors having 643 input characteristics for each 5 s segment of EEG from 18 EEG channels. The authors tested the performance of different LSTM architectures and concluded that the network with two LSTM layers alternating with two dropout layers was the best evaluated. The performance was also investigated by defining different durations of pre-ictal periods, ranging from 15 min to 120 min. In this case, 60 min duration was the one with the best results, with mean values of specificity and sensitivity of 100%. However, it is not possible to conclude which window separation method for training and testing, method A or method B, was adopted by the authors, since it was not explicitly addressed in the text. The same happens in [31]. The authors investigated the prediction of epileptic seizures during sleep and, for this, filtered the EEG signal at the predominant frequencies in this phase (delta and gamma). Only four patients from the Physionet CHB-MIT Scalp EEG Database were chosen for the study. The duration of the pre-ictal period of 30 min was adopted. WPT (Wavelet Packet Transform) techniques were combined, whose decomposition process can be applied at low and high frequencies, and 4 s sliding EEG windows were segmented, and the pre or interictal evaluation was performed for every two windows. A Bi-LSTM (Bidirectional LSTM) was used for classification, considering the cross-validation process. A moving average filter to smooth forecasts and decrease false forecasts was applied. The authors also compared the performance of the Bi-LSTM with an LSTM and with a Multilayer Perceptron (MLP). The former outperformed the other arrangements, with average results above 99% with a low rate of false alarms. In [32], it is proposed a deep learning based ensemble learning method to predict epileptic seizures. In that method, EEG signals were preprocessed using empirical mode decomposition followed by bandpass filter for noise removal. A three-layer customized convolutional neural network (CNN) was combined with handcrafted features to extract automated characteristics from preprocessed EEG signals. The feature set was used to train an ensemble classifier that combines the output of SVM, CNN and LSTM using model agnostic meta learning. An average sensitivity of 96.28% and specificity of 95.65% with an average anticipation time of 33 min on all subjects of CHB-MIT have been achieved, whereas, on American epilepsy society-Kaggle seizure prediction dataset, an average sensitivity of 94.2% and specificity of 95.8% have been achieved on all subjects. In [33], the authors combined CNN to extract the characteristics of the EEG windows after the pre-processing of the signals, which included filtering and Short Time Fourier Transform (STFT), with SVM to perform the classification in pre or interictal states. They achieved an

© Brazilian Computational Intelligence Society

average of sensitivity and specificity of 92.7% and 90.8% respectively. Previously, in [34] the authors applied Empirical Mode Decomposition (EMD) to extract the time and frequency domain features for training a prediction model. They have extracted the first four statistical moments and the three spectral moments of each 1 s window of the EEG, using all the patients from CHB-MIT. At the time, they achieved an average sensitivity of 92.23% and specificity of 93.38% and average prediction time of 23.6 minutes. In that group's more recent work [35], it is proposed an approach in which a regression-based method is applied as an alternative to notch filtering to increase the signal-to-noise ratio. A CNN extracts the features, followed by classification of interictal state and preictal state segments using LSTM to predict seizures. They achieved accuracy of 94.0%, sensitivity of 93.8%, and 91.2% specificity by using CHB-MIT dataset. For all these works it was not clearly informed which method, A or B, has been applied, as it also occurs in [36], where the authors proposed a model that combines a Dense Convolutional Network (DenseNet) and LSTM to predict epileptic seizures. The raw EEG data was segmented in 10 s windows with an overlapped window of 1 s. Different sizes of pre-ictal state have been studied: 5, 10 and 15 min duration. Also, the authors disregarded the 5 minutes prior to the seizures. That is, this period was not counted as pre-ictal. The mother function db4 of the discrete wavelet transform (DWT) was applied to the segments to convert them into a time-frequency type 2D image (EEG spectrogram). Then, the preprocessing data are used as the input data of DenseNet and the resulting feature map is used as the input data of LSTM. To evaluate the performance of the proposed method, experiments are conducted for each preictal length of 5, 10, and 15 min using the CHB-MIT dataset. As a result, they achieved a prediction accuracy of 93.28%, a sensitivity of 92.92%, a specificity of 93.65%, a false positive rate of 0.063 per hour, and an F1-score of 0.923 when the preictal length was 5 min, i.e. 10 min before the seizure.

Another alternative is based on the use of bi-LSTM network as presented in [6]. The authors used spectral power and mean spectrum amplitude features of delta, theta, alpha, beta, and gamma bands of 23-channel EEG spectrum for this task. Physionet CHB-MIT database was used and segments of 5-50 s of all available patients were extracted. They found that 30 s segments were sufficient for an accurate prediction. The results from bi-LSTM network were compared to those obtained by different methods, such as random forest, decision tree, k-nearest neighbor, support vector machine, and naive Bayes classifiers, surpassing all of them with average classification accuracy of 98.14%, average sensitivity of 98.51%, and average specificity of 97.78%. Once again, it is not informed how the pre and interictal windows were separated for the training and testing stages.

In [37], the use of CNN for all patients from CHB-MIT database is investigated. The authors proposed a method that reduced the number of channels used to only 6 (instead of 23) in order to achieve a more simple system that would be easily implemented. Despite the good metrics presented (97.83% and 92.36% in terms of average sensitivity and specificity, respectively) it was not clearly mentioned how the windows were divided in the training/testing process of the CNN applied in the work. In a previous work [38], similar experiments were followed, with good results, using Dense Convolutional Network (DenseNet) and all available database channels, but without presenting a clear definition of the method (A or B). The same situation also occurred in [39] where scalograms have been used to create a multi-channel vision transformer (MViT) algorithm. The EEG windows are converted by Continuous Wavelet Transform (CWT) and the images were then split into fixed-size non-overlapping patches, which were used as inputs to the MViT algorithm to automatically learn the distinctive EEG features needed for subtle seizure prediction, performed by an MLP. Different datasets from human and dogs were explored and very good results have been obtained for the CHB-MIT patients (almost 100% for the metrics).

A deep convolutional generative adversarial network (DCGAN) was proposed in [40] to generate synthetic EEG data to increase the prediction performance and to overcome good quality data scarcity issue. They also used transfer-learning (TL) for evaluating the performance of four well-known deep-learning (DL) models to predict epileptic seizure. The generated data were validated using one-class SVM and a new proposal namely convolutional epileptic seizure predictor (CESP). The CESP model presented higher accuracy ($\approx 5\%$) as compared to state-of-the-art augmentation techniques when trained on synthesized data and testing on real Epilepsyecosystem [41] and CHB-MIT datasets, respectively. Unlike other works, in this one, the authors point out that the tests were performed with temporal parts of the EEG that were not used in training (method B). The same occurs in the work of [42], where a combination of CNN-LSTM architectures resulted in an accuracy of approximately 98% with an average prediction time of 44.7 min. In this case, the authors affirm that multichannel images are first constructed by applying short-time Fourier transform (STFT) to EEG signals. After a preprocessing step, the neural network is trained on the STFTs in order to capture the spectral, spatial and temporal features within and between the EEG segments and classify them as pre-ictal or interictal stage which is performed based on a majority voting method.

In [43], the authors draw attention to the dynamic nature of EEG data, characterised by data changes with time, known as concept drifts. Then they investigate the effectiveness of automatic concept drift adaptation methods in seizure prediction. These proposed methods incorporate a retraining process after each seizure and use a combination of univariate linear features from time and frequency domains and SVM classifiers. The authors developed three seizure prediction pipelines to intrinsically adapt to concept drifts. The pipelines were iteratively trained, by accumulating data from previous seizures and tested on the following one. Thus it clear that the method B was used in that work.

In [44], the authors propose an algorithm using post-processing techniques to explore the existence of a set of chronological events of brain activity that precedes epileptic seizures. Thus it is assumed that the epileptic seizure is not the consequence a single event, but rather a succession of events. The methodology combines univariate linear features with a classifier based on Support Vector Machines (SVM) and two post-processing techniques to handle pre-seizure temporality employing knowledge from network theory. In the so-called Chronological Firing Power approach, they considered the preictal as a sequence of three brain activity events separated in time. In the Cumulative Firing Power approach, which provided the best results, they assumed the preictal period as a sequence of three overlapping events. In the data splitting phase of the procedure, the three first available

© Brazilian Computational Intelligence Society

seizures of each patient was used in training and the following seizures in testing, according to the method B.

Another recent published work in this area is [45]. It proposed a system integrated to the cloud that used a convolutional neural network. But it does not make clear whether its objective is to detect or predict epileptic seizures. At no point does it mention the duration considered for the pre-ictal period, nor how far in advance he would be able to predict seizures. It also does not mention details about the choice of training and testing data.

In [46], the pre-ictal period is divided into multiple temporal windows. The prediction method is based on deep residual shrinkage network (DRSN) and gated recurrent unit (GRU). The temporal dependency of the EEG in that windows is modeled by introducing GRU into a DRSN. Automatic feature extraction is achieved by applying soft threshold denoising and attention mechanism inside the neural network. The proposed method was tested on four patients from the CHB-MIT EEG dataset. The choice of training and validation data was made according to the method B. The sensitivity was 90.54%, the AUC value was 0.88 and the false prediction rate was 0.11/h.

Table 1 summarizes the analyses made in this section regarding whether there was an explicit statement that method B was used in the cited works.

Table 1: Summary of published papers on seizure prediction regarding whether they explicitly stated that they used method B in their work.

paper	Was it explicitly stated that method B was used?
Tsiouris et al [28]	no
Cheng et al [31]	no
Usman, Khalid, Bashir [32]	no
Usman, Khalid, Aslam [33]	no
Usman, S. M., Usman, M., Fong [34]	no
Aslam et al [35]	no
Ryu, Joe [36]	no
Singh, Malhotra [6]	no
Jana, Mukherjee [37]	no
Jana, Bhattacharyya, Das [38]	no
Hussein, Lee, Ward [39]	no
Rasheed et al [40]	yes
Shahbazi, Aghajan [42]	yes
Pontes et al [43]	yes
Batista et al [44]	yes
Kumar, Prakash [45]	no
Xu et al [46]	yes

3 METHODS

In the Introduction, it was stated that the objective of this work is to investigate the effect of the way of choosing the training and test windows on the accuracy of a seizures prediction system. In order to isolate that effect, several pairs of experiments were performed. Inside each pair, all parameters of the experiments were maintained constant, except that in the first experiment of the pair, EEG windows were chosen as in Figure 2 (method A) while in the second experiment of the pair, EEG windows were chosen as in Figure 4 (method B).

EEG signals of epilepsy patients used in this work were from the Physionet CHB-MIT Scalp EEG Database, described by [29, 30]. That database has EEG signals from more than 20 pediatric patients. The sampling frequency was 256 Hz and the resolution was 16-bits, using the International 10-20 electrodes position system. There are many variations among patients in the number of channels used and their numbering in the database. EEG signals from the first 11 patients were collected using the same first 18 channels, with constant numbering among them. But the last seizures of patient number 11 are very close one another, not leaving enough interictal and pre-ictal time between them. Therefore, the experiments performed here used only the first 18 channels from the first 10 patients in the database. The whole pre-ictal EEG periods of each patient were used in the experiments. For this purpose, they were subdivided in 5 s adjacent temporal windows. Then, the same number of 5 s EEG temporal windows were extracted from the interictal periods. For each patient, the total available interictal periods are greater than the total available pre-ictal periods. So, in order to have the same number of windows from the two classes, interictal 5 s windows were not adjacent, but were extracted from EEG data at regular time intervals, leaving gaps between them. Table 2 summarizes patients records characteristics. Looking directly at the database [29], it is possible to observe that the available EEG recording times, the number of seizures and the intervals between them vary greatly depending on the patient. Therefore, in the experiments, the number of EEG windows used also varied greatly depending on the patient. The experiments were performed on a Subject-Specific Model (SSM) basis, as is usual in this field.

© Brazilian Computational Intelligence Society

Table 2: Patients records characteristics. Gender: Male (M), Female (F).

patient	gender	number of seizures	total recording time (hh:mm:ss)
1	F	7	40:33:08
2	M	3	35:15:59
3	F	7	38:00:06
4	M	4	156:03:54
5	F	5	39:00:10
6	F	10	66:44:06
7	F	3	67:03:08
8	M	5	20:00:23
9	F	4	67:52:18
10	M	7	50:01:24
total		55	580:34:36

The methods used in the experiments here are inspired mainly by that presented in [28], also of the SSM type, due to the excellent accuracy results obtained there. As stated in Section 2, that work employed an LSTM as the classifier [47]. It receives vector sequences as its inputs and, in that case, classifies each sequence as belonging to the interictal or to the pre-ictal class.

Each experiment carried out in the present work used vectors assembled from a single commonly used feature type extracted from the EEG signals. The features types explored here were obtained from: Fast Fourier Transform (FFT), wavelet transform, cross-correlation between EEG channels and graph theory. Statistical features like those used in [28] had very poor performance here for all experiments configurations and their use have then been discarded. Details on the extraction for the four types of features used in the experiments are described below.

In the FFT features extraction, performed by using the numpy Python language package [48], FFT transform was applied for each 5 s EEG window and each EEG channel. Then six numbers were calculated for each channel by averaging the absolute values of the spectral lines in the following bands: delta ($\leq$ 3 Hz), theta (4 to 7 Hz), alpha (8 to 13 Hz), beta (14 to 30 Hz), gamma1 (30 to 55 Hz) and gamma2 (65 to 110 Hz). Thus for 18 channels and six numbers per channel, a vector composed of 108 numbers was assembled for each EEG window.

In the features extraction from wavelet transform, as in [28], the db4 was the mother wavelet. The discrete wavelet transform was applied here to each 5 s EEG window, separately to each channel. In each level of detail, a number was obtained as the average of the absolute values of the wavelet coefficients at that level. Seven such numbers were obtained for each of the 18 channels, resulting in an 126 components vector for each EEG window. This was performed using the Python package PyWavelets [49].

One of the features extracted from correlation calculations is the cross-correlation between EEG channels. For each pair of EEG channels, $a[t]$ and $b[t]$, where t represents the sample indices, the Pearson correlation coefficient is calculated to several relative displacements between these signals: $a[t]$ and $b[t-5]$, $a[t]$ and $b[t-4]$, ..., $a[t]$ and $b[t]$, ..., $a[t]$ and $b[t+4]$, $a[t]$ and $b[t+5]$, and the maximum among these coefficients is retained. Another feature extracted from correlation calculations is the decorrelation time, which is the smallest value of the integer k such that the Pearson correlation coefficient between $x[t]$ and $x[t+k]$ falls below 0.5. These calculations were performed using the numpy Python language package [48] resources.

For the extraction of features using Graph Theory, initially, a graph was built with a node representing each channel. The edges between each pair of nodes was weighted with the cross-correlation measure between these two EEG channels, as described above. Then several features were extracted from this graph: clustering coefficient, global efficiency, local efficiency, betweenness centrality, eccentricity, radius, diameter and characteristic path length [50]. For this, the Python package NetworkX was used [51].

As in [28], experiments were performed here for which sequences were assembled to the LSTM using 15, 30, 45 and 60 features vectors. Each such training sequence was assembled by randomly choosing vectors from their classes, without regarding their original ordering nor their adjacencies in the EEG signal. The only difference in the sequence assemblies between methods A and B was that for method A, the Figure 2 configuration was adopted using the extracted windows, whereas for method B, the Figure 4 configuration was adopted. In this way, different sequences were assembled between methods A and B, based on the same EEG temporal windows. The pre-ictal period duration was adopted as being 60 minutes, because this was the best value verified in [28]. The LSTM had here one layer, with 32 cells. In each experiment, it was trained using 40 epochs.

4 RESULTS

The obtained results are summarized in Table 3. The first column represents the four methods for extracting features from EEG used here. The second column specifies the four lengths of the vectors sequences used to train and test the LSTM. For each of these 16 combinations of features extraction methods with sequences lengths, experiments were performed for the first ten patients of the database, first using the method A and then using the method B of choosing train and test windows, as indicated in the third column. These totaled $4 \times 4 \times 10 \times 2 = 320$ experiments. The next four columns show the mean accuracy, sensitivity, specificity and F1-score and the standard error of the mean for the ten patients in the conditions specified in the first three columns. There is no clear trend in the influence of sequence length on the evaluated metrics regardless of the method or even the feature set. However, when comparing the methods, since the F1-score is a good metric for summarizing classification tests results, the last column shows for what of the two methods, A or B, the F1-score was higher, that is, for which the results were the best in each condition. Clearly, it can be seen that method A was superior in almost all cases.

Table 3: Results. Here, each accuracy, sensitivity, specificity and F1 score value is the average of results obtained in experiments performed for 10 patients.

features from	sequences length	method	accuracy	sensitivity	specificity	F1 score	method with the best average F1 score
correlation	15	A	0.52 ± 0.12	0.54 ± 0.16	0.52 ± 0.15	0.12 ± 0.04	B
	15	B	0.55 ± 0.09	0.49 ± 0.15	0.65 ± 0.13	0.23 ± 0.07	
	30	A	0.78 ± 0.10	0.64 ± 0.13	0.79 ± 0.12	0.45 ± 0.13	A
	30	B	0.71 ± 0.08	0.22 ± 0.11	0.91 ± 0.04	0.15 ± 0.08	
	45	A	0.87 ± 0.09	0.72 ± 0.14	0.89 ± 0.10	0.63 ± 0.15	A
	45	B	0.62 ± 0.10	0.35 ± 0.14	0.74 ± 0.13	0.13 ± 0.06	
	60	A	0.64 ± 0.14	0.77 ± 0.13	0.60 ± 0.16	0.48 ± 0.14	A
	60	B	0.72 ± 0.09	0.23 ± 0.13	0.88 ± 0.10	0.15 ± 0.10	
FFT	15	A	0.91 ± 0.03	0.75 ± 0.12	0.93 ± 0.03	0.61 ± 0.11	A
	15	B	0.70 ± 0.11	0.55 ± 0.13	0.63 ± 0.15	0.40 ± 0.12	
	30	A	0.79 ± 0.11	0.78 ± 0.13	0.77 ± 0.13	0.54 ± 0.13	A
	30	B	0.68 ± 0.11	0.36 ± 0.15	0.80 ± 0.13	0.24 ± 0.11	
	45	A	0.72 ± 0.11	0.43 ± 0.15	0.76 ± 0.13	0.24 ± 0.12	A
	45	B	0.73 ± 0.10	0.20 ± 0.13	0.90 ± 0.10	0.15 ± 0.09	
	60	A	0.63 ± 0.13	0.69 ± 0.15	0.61 ± 0.16	0.39 ± 0.14	A
	60	B	0.57 ± 0.11	0.25 ± 0.13	0.77 ± 0.13	0.04 ± 0.03	
graphs	15	A	0.79 ± 0.11	0.56 ± 0.15	0.79 ± 0.13	0.42 ± 0.14	A
	15	B	0.83 ± 0.06	0.14 ± 0.09	0.99 ± 0.01	0.17 ± 0.10	
	30	A	0.76 ± 0.11	0.50 ± 0.16	0.78 ± 0.13	0.26 ± 0.12	A
	30	B	0.65 ± 0.11	0.20 ± 0.11	0.78 ± 0.13	0.13 ± 0.08	
	45	A	0.74 ± 0.11	0.46 ± 0.16	0.77 ± 0.13	0.25 ± 0.11	A
	45	B	0.67 ± 0.10	0.30 ± 0.15	0.67 ± 0.15	0.13 ± 0.10	
	60	A	0.82 ± 0.08	0.41 ± 0.16	0.89 ± 0.10	0.25 ± 0.11	A
	60	B	0.68 ± 0.12	0.20 ± 0.13	0.80 ± 0.13	0.10 ± 0.10	
Wavelets	15	A	0.59 ± 0.13	0.78 ± 0.13	0.55 ± 0.16	0.40 ± 0.12	A
	15	B	0.54 ± 0.12	0.69 ± 0.14	0.39 ± 0.16	0.37 ± 0.10	
	30	A	0.47 ± 0.13	0.92 ± 0.08	0.43 ± 0.15	0.40 ± 0.12	A
	30	B	0.66 ± 0.10	0.49 ± 0.16	0.69 ± 0.15	0.26 ± 0.10	
	45	A	0.43 ± 0.14	0.90 ± 0.10	0.38 ± 0.16	0.36 ± 0.12	A
	45	B	0.80 ± 0.07	0.36 ± 0.15	0.78 ± 0.13	0.30 ± 0.12	
	60	A	0.39 ± 0.13	0.88 ± 0.10	0.29 ± 0.15	0.33 ± 0.12	A
	60	B	0.50 ± 0.11	0.60 ± 0.16	0.60 ± 0.14	0.24 ± 0.11	

The number of sequences that were used in each experiment varied from tens to thousands, depending on the available interictal and pre-ictal durations for each patient in the database, always looking to ensure that there was no imbalance between the two classes in the training process.

Another way to organize the results is by patient, as shown in the Table 4. It shows the same metrics as Table 3, but presented as averages, for each patient, of the results obtained in all 16 experiments (4 features $\times$ 4 sequences lenghts) for each method (A or B). For all 10 patients, it can be seen that method A provided the best F1-score on average.

Many published works assemble different features in the input vectors to the system, which certainly impacts its performance. However, if the adopted method includes EEG training windows as illustrated in method A, the results presented here indicate that the behavior of those systems will tend to be superior in relation to those performed in the most indicated way, as in method B.

5 CONCLUSION

In this paper, two ways of choosing training and test data are compared for epileptic seizures prediction systems. Experimental results showed that the biased method A tends to provide better results than the method B. But the method B is the more appropriate to be employed in a real application.

© Brazilian Computational Intelligence Society

Table 4: Results per patient. Here, each accuracy, sensitivity, specificity and F1 score value is the average of results obtained in all 16 experiments (4 features $\times$ 4 sequences lenghts) performed for each patient for each method (A or B).

patient	method	accuracy	sensitivity	specificity	F1 score	method with the best average F1 score
CHB01	A	0.84 ± 0.06	0.79 ± 0.10	0.85 ± 0.08	0.68 ± 0.10	A
	B	0.86 ± 0.02	0.25 ± 0.09	0.96 ± 0.03	0.25 ± 0.08	
CHB02	A	0.61 ± 0.10	0.74 ± 0.11	0.60 ± 0.12	0.38 ± 0.10	A
	B	0.56 ± 0.06	0.38 ± 0.12	0.62 ± 0.12	0.17 ± 0.05	
CHB03	A	0.54 ± 0.11	0.75 ± 0.11	0.50 ± 0.13	0.36 ± 0.10	A
	B	0.38 ± 0.07	0.26 ± 0.10	0.63 ± 0.13	0.26 ± 0.10	
CHB04	A	0.73 ± 0.10	0.69 ± 0.10	0.73 ± 0.10	0.35 ± 0.10	A
	B	0.72 ± 0.11	0.27 ± 0.11	0.72 ± 0.11	$\approx 0.00 \pm 0.00$	
CHB05	A	0.48 ± 0.08	0.60 ± 0.12	0.42 ± 0.12	0.23 ± 0.05	A
	B	0.57 ± 0.08	0.37 ± 0.12	0.61 ± 0.11	0.11 ± 0.03	
CHB06	A	0.71 ± 0.09	0.48 ± 0.12	0.74 ± 0.11	0.27 ± 0.09	A
	B	0.51 ± 0.06	0.36 ± 0.12	0.57 ± 0.12	0.16 ± 0.05	
CHB07	A	0.77 ± 0.09	0.54 ± 0.12	0.78 ± 0.10	0.26 ± 0.10	A
	B	0.81 ± 0.08	0.25 ± 0.10	0.83 ± 0.09	0.06 ± 0.03	
CHB08	A	0.81 ± 0.07	0.78 ± 0.08	0.83 ± 0.08	0.65 ± 0.09	A
	B	0.88 ± 0.04	0.49 ± 0.11	0.95 ± 0.03	0.45 ± 0.11	
CHB09	A	0.74 ± 0.10	0.68 ± 0.12	0.75 ± 0.11	0.42 ± 0.11	A
	B	0.80 ± 0.10	0.66 ± 0.12	0.80 ± 0.10	0.36 ± 0.10	
CHB10	A	0.54 ± 0.10	0.67 ± 0.12	0.52 ± 0.12	0.24 ± 0.07	A
	B	0.54 ± 0.04	0.24 ± 0.11	0.81 ± 0.10	0.16 ± 0.07	

The main conclusion here is that papers on seizures prediction methods should make it very clear whether the training and test EEG time windows were chosen as in method A or in method B. Otherwise, there will be suspicion as to whether good results were due to using the less appropriate method A.

In the literature review, few works were found that explicitly states that they used the method B of choosing training EEG windows. The present work results experimentally justifies this choice, demonstrating that this issue can no longer be neglected. The contribution of this work is to alert researchers in the area about this important issue, both in the preparation of their experiments and in the explicit indication, in future published articles, that their work used method B, considered here the more appropriate. Otherwise, good experimental results obtained in laboratory, but due to the use of the inappropriate method A, will not be reproducible in a real life application.

A limitation of method B is that it requires that there be sufficient time in the interictal period and preictal period for testing the system after the previous seizures used for training, that is, there must be at least one last seizure in the patient's EEG, sufficiently removed from previous crises.

The conclusions obtained here may be important not only in the creation of systems for the prediction of epileptic seizures, but also in the application of machine learning to EEG analysis for other purposes and in the analysis of physiological signals of other types, such as cardiac and muscular.

Although this was not the objective of this work, from Table 3, restricting the comparison to the experiments using method B, it is seen that, on average, the form of feature extraction that provided the best F1-score values was the wavelets, followed, in descending order, by FFT, correlation and graphs. Regarding sequence lengths, using method B, the best average F1-score was obtained for length 15, followed by 30, 45 and 60, that is, the F1-score fell with the increase in sequence length. But, as this was not the objective of this work, more general conclusions about the best ways of extracting features, best sequence lengths, best EEG window durations, best classification algorithms and best values for other parameters should be sought in the literature or investigated through experiments specifically prepared for these purposes.

In Figure 2, one may see another potential problem in the method presented in [28]: the sequences assembled there to the LSTM do not maintain the original ordering or adjacencies between EEG windows. As a future work, a comparison might be made between the results of this assembling method and those of a method in which the original ordering and adjacencies were respected.

REFERENCES

[1] World Health Organization. "Epilepsy". https://www.who.int/news-room/fact-sheets/detail/epilepsy, 2022.

[2] Epilepsy Foundation. "Epilepsy Foundation". https://www.epilepsy.com, 2022.

[3] Mayo Foundation for Medical Education and Research. "EEG (electroencephalogram)". https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875, 2022.

[4] M. Pinto, A. Leal, F. Lopes, A. Dourado, P. Martins, C. A. Teixeira *et al.*. “A personalized and evolutionary algorithm for interpretable EEG epilepsy seizure prediction”. *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[5] M. Mula and F. Monaco. “Ictal and peri-ictal psychopathology”. *Behavioural neurology*, vol. 24, no. 1, pp. 21–25, 2011.

[6] K. Singh and J. Malhotra. “Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features”. *Complex & Intelligent Systems*, pp. 1–14, 2022.

[7] K. Gadhoumi, J.-M. Lina, F. Mormann and J. Gotman. “Seizure prediction for therapeutic devices: A review”. *Journal of neuroscience methods*, vol. 260, pp. 270–282, 2016.

[8] M. Savadkoohi, T. Oladunni and L. Thompson. “A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal”. *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1328–1341, 2020.

[9] P. Kunekar, M. K. Gupta and P. Gaur. “Detection of epileptic seizure in EEG signals using machine learning and deep learning techniques”. *Journal of Engineering and Applied Science*, vol. 71, no. 1, pp. 21, 2024.

[10] L. A. S. Kitano, M. A. A. Sousa, S. D. Santos, R. Pires, S. Thome-Souza and A. B. Campo. “Epileptic seizure prediction from EEG signals using unsupervised learning and a polling-based decision process”. In *International Conference on Artificial Neural Networks*, pp. 117–126. Springer, 2018.

[11] J. Birjandtalab, M. B. Pouyan and M. Nourani. “Unsupervised EEG analysis for automated epileptic seizure detection”. In *First International Workshop on Pattern Recognition*, edited by X. Jiang, G. Chen, G. Capi and C. Ishll, volume 10011, p. 100110M. International Society for Optics and Photonics, SPIE, 2016.

[12] A.-M. Tăuţan, M. Dogariu and B. Ionescu. “Detection of epileptic seizures using unsupervised learning techniques for feature extraction”. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2377–2381. IEEE, 2019.

[13] B. Abbasi and D. M. Goldenholz. “Machine learning applications in epilepsy”. *Epilepsia*, vol. 60, no. 10, pp. 2037–2047, 2019.

[14] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O’Brien and A. Razi. “Machine learning for predicting epileptic seizures using EEG signals: A review”. *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 139–155, 2020.

[15] S. Selim, E. Elhinamy, H. Othman, W. Abouelsaadat and M. A.-M. Salem. “A review of machine learning approaches for epileptic seizure prediction”. In *2019 14th International Conference on Computer Engineering and Systems (ICCES)*, pp. 239–244. IEEE, 2019.

[16] Y. Si. “Machine learning applications for electroencephalograph signals in epilepsy: a quick review”. *Acta Epileptologica*, vol. 2, no. 1, pp. 1–7, 2020.

[17] M. S. Nafea and Z. H. Ismail. “Supervised Machine Learning and Deep Learning Techniques for Epileptic Seizure Recognition Using EEG Signals - A Systematic Literature Review”. *Bioengineering*, vol. 9, no. 12, 2022.

[18] M. Natu, M. Bachute, S. Gite, K. Kotecha and A. Vidyarthi. “Review on epileptic seizure prediction: machine learning and deep learning approaches”. *Computational and Mathematical Methods in Medicine*, vol. 2022, 2022.

[19] M. I. B. Ahmed, S. Alotaibi, S. Dash, M. Nabil, A. O. AlTurki *et al.*. “A Review on Machine Learning Approaches in Identification of Pediatric Epilepsy”. *SN Computer Science*, vol. 3, no. 6, pp. 1–15, 2022.

[20] J. Waring, C. Lindvall and R. Umeton. “Automated machine learning: Review of the state-of-the-art and opportunities for healthcare”. *Artificial intelligence in medicine*, vol. 104, pp. 101822, 2020.

[21] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang and L. Yang. “Physics-informed machine learning”. *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[22] E. H. Houssein, M. M. Emam, A. A. Ali and P. N. Suganthan. “Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review”. *Expert Systems with Applications*, vol. 167, pp. 114161, 2021.

[23] N. Stephenson, E. Shane, J. Chase, J. Rowland, D. Ries, N. Justice, J. Zhang, L. Chan and R. Cao. “Survey of machine learning techniques in drug discovery”. *Current drug metabolism*, vol. 20, no. 3, pp. 185–193, 2019.

[24] J. G. Greener, S. M. Kandathil, L. Moffat and D. T. Jones. “A guide to machine learning for biologists”. *Nature Reviews Molecular Cell Biology*, vol. 23, no. 1, pp. 40–55, 2022.

[25] K. G. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis. “Machine learning in agriculture: A review”. *Sensors*, vol. 18, no. 8, pp. 2674, 2018.

[26] B. Maimaiti, H. Meng, Y. Lv, J. Qiu, Z. Zhu, Y. Xie, Y. Li, W. Zhao, J. Liu, M. Li *et al.*. "An Overview of EEG-based Machine Learning Methods in Seizure Prediction and Opportunities for Neurologists in this Field". *Neuroscience*, 2021.

[27] H. Lekshmy, D. Panickar and S. Harikumar. "Comparative analysis of multiple machine learning algorithms for epileptic seizure prediction". In *Journal of Physics: Conference Series*, volume 2161, p. 012055. IOP Publishing, 2022.

[28] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris and D. I. Fotiadis. "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals". *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 2018.

[29] A. H. Shoeb. "Application of machine learning to epileptic seizure onset detection and treatment", 2009. PHD Thesis, Massachusetts Institute of Technology.

[30] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng and H. E. Stanley. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[31] C. Cheng, B. You, Y. Liu and Y. Dai. "Patient-specific method of sleep electroencephalography using wavelet packet transform and Bi-LSTM for epileptic seizure prediction". *Biomedical Signal Processing and Control*, vol. 70, pp. 102963, 2021.

[32] S. M. Usman, S. Khalid and S. Bashir. "A deep learning based ensemble learning method for epileptic seizure prediction". *Computers in Biology and Medicine*, vol. 136, pp. 104710, 2021.

[33] S. Usman, S. Khalid and M. H. Aslam. "Epileptic Seizures Prediction Using Deep Learning Techniques". *IEEE Access*, vol. PP, pp. 1–1, 02 2020.

[34] S. M. Usman, M. Usman and S. Fong. "Epileptic seizures prediction using machine learning methods". *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[35] M. H. Aslam, S. M. Usman, S. Khalid, A. Anwar, R. Alroobaea, S. Hussain, J. Almotiri, S. S. Ullah and A. Yasin. "Classification of EEG Signals for Prediction of Epileptic Seizures". *Applied Sciences*, vol. 12, no. 14, 2022.

[36] S. Ryu and I. Joe. "A Hybrid DenseNet-LSTM Model for Epileptic Seizure Prediction". *Applied Sciences*, vol. 11, no. 16, 2021.

[37] R. Jana and I. Mukherjee. "Deep learning based efficient epileptic seizure prediction with EEG channel optimization". *Biomedical Signal Processing and Control*, vol. 68, pp. 102767, 2021.

[38] R. Jana, S. Bhattacharyya and S. Das. "Epileptic seizure prediction from EEG signals using DenseNet". In *2019 IEEE symposium series on computational intelligence (SSCI)*, pp. 604–609. IEEE, 2019.

[39] R. Hussein, S. Lee and R. Ward. "Multi-Channel Vision Transformer for Epileptic Seizure Prediction". *Biomedicines*, vol. 10, no. 7, 2022.

[40] K. Rasheed, J. Qadir, T. J. O'Brien, L. Kuhlmann and A. Razi. "A generative model to synthesize eeg data for epileptic seizure prediction". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2322–2332, 2021.

[41] L. Kuhlmann, P. Karoly, D. R. Freestone, B. H. Brinkmann, A. Temko, A. Barachant, F. Li, G. Titericz Jr, B. W. Lang, D. Lavery *et al.*. "Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG". *Brain*, vol. 141, no. 9, pp. 2619–2630, 2018.

[42] M. Shahbazi and H. Aghajan. "A generalizable model for seizure prediction based on deep learning using CNN-LSTM architecture". In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 469–473. IEEE, 2018.

[43] E. D. Pontes, M. Pinto, F. Lopes and C. Teixeira. "Concept-drifts adaptation for machine learning EEG epilepsy seizure prediction". *Scientific Reports*, vol. 14, no. 1, pp. 8204, 2024.

[44] J. Batista, M. F. Pinto, M. Tavares, F. Lopes, A. Oliveira and C. Teixeira. "EEG epilepsy seizure prediction: the post-processing stage as a chronology". *Scientific Reports*, vol. 14, no. 1, pp. 407, 2024.

[45] S. Kumar and S. Prakash. "Prediction of Epileptic Seizures based on EEG Signal using CNN Model". In *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, pp. 84–89. IEEE, 2023.

[46] X. Xu, Y. Zhang, R. Zhang and T. Xu. "Patient-specific method for predicting epileptic seizures based on DRSN-GRU". *Biomedical Signal Processing and Control*, vol. 81, pp. 104449, 2023.

[47] F. A. Gers, J. Schmidhuber and F. Cummins. "Learning to forget: Continual prediction with LSTM". *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[48] NumPy. "NumPy - The fundamental package for scientific computing with Python". https://numpy.org/, 2022.

[49] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt and A. O'Leary. "PyWavelets: A Python package for wavelet analysis". *Journal of Open Source Software*, vol. 4, no. 36, pp. 1237, 2019.

[50] M. Rubinov and O. Sporns. "Complex network measures of brain connectivity: Uses and interpretations". *Neuroimage*, vol. 52, pp. 1059–1069, 2010.

[51] NetworkX, developers. "NetworkX, Network Analysis in Python". https://networkx.org/, 2022.