Proceedings of Machine Learning Research 1:1–16, 2018 COPA 2018

Detecting seizures in EEG recordings using Conformal Prediction

Charalambos Eliades PAMBOSELIADES@HOTMAIL.COM
Harris Papadopoulos H.PAPADOPOULOS@FREDERICK.AC.CY
Dept. of Computer Science and Engineering
Frederick University
7 Y. Frederickou St., Palouriotisa, Nicosia 1036, Cyprus

Editor: Alex Gammerman, Vladimir Vovk, Zhiyuan Luo, Evgueni Smirnov and Ralf Peeters

Abstract

This study examines the use of the Conformal Prediction (CP) framework for the provision of confidence information in the detection of seizures in electroencephalograph (EEG) recordings. The detection of seizures is an important task since EEG recordings of seizures are of primary interest in the evaluation of epileptic patients. However, manual review of long-term EEG recordings for detecting and analyzing seizures that may have occurred is a time-consuming process. Therefore a technique for automatic detection of seizures in such recordings is highly beneficial since it can be used to significantly reduce the amount of data in need of manual review. Additionally, due to the infrequent and unpredictable occurrence of seizures, having high sensitivity is crucial for seizure detection systems. This is the main motivation for this study, since CP can be used for controlling the error rate of predictions and therefore guaranteeing an upper bound on the frequency of false negatives.
Keywords: EEG, Seizure, Confidence, Credibility, Prediction Regions

1. Introduction

Epileptic seizures reflect the clinical signs of the synchronized discharge of brain neurons. The effects of this situation can be characterized by disturbances of mental function and/or movements of body (Lehnertz et al., 2003). This neurological disorder occurs in approximately $0.6-0.8\%$ of the entire population. Two-thirds of the patients can successfully control their seizures onset with the use of anti epileptic drugs, and another $8-10\%$ could benefit from resective surgery. Unfortunately there are cases where the patients have to live their lives with uncontrolled seizure attacks (approximately 25% of the patients) (Mormann et al., 2007).

Traditionally, suspected seizures are evaluated using a routine electroencephalogram EEG, which is a 20-minute recording of the patient's brain waves. However in those 20 minutes a seizure onset might not occur and the identification of an epileptic subject might be difficult. For this reason it is desirable to continuously record EEG using portable recording devices. Apart from the fact that such devices can continually record for several days they have the additional advantage that the patient is not away from his/her environment and routine (Waterhouse, 2003a,b). Clinical neurophysiologists can then analyze the EEG recordings and identify any seizures. Analyzing several days of EEG recordings however is a time consuming task. Consequently, a technique that can automatically detect the sections of EEG recordings which correspond to seizures is highly desirable. With the use of such a technique neurophysiologists can focus their analysis only on the detected sections, which will significantly reduce the time needed for this task. An important requirement for EEG detection techniques is to have high sensitivity, since failing to identify some of the typically infrequently occurring seizures will mean that important data existing in the recording will be discarded.

In this study we propose the use of Conformal Prediction (CP) in order to provide a way of controlling the frequency with which an EEG detection technique will make an error. More specifically, CP produces prediction sets that are guaranteed to contain the true class with a prespecified frequency (confidence level). By providing neurophysiologists with all the sections in the EEG recording that contain the seizure class in their prediction set for a very high confidence value, say 99.5%, we can guarantee that a seizure occurrence will be discarded with at most 0.5% frequency. Note that providing a small number of false positive detections is not a problem as these can be relatively quickly discarded by the neurophysiologist.

The proposed approach consists of the following steps. Discrete wavelet transformation (DWT) is firstly applied to each EEG clip to extract the brain waves. Then the fast Fourier transformation (FFT) is calculated and seven features are extracted. The optimum combination of waves that maximizes accuracy is then identified by examining the performance of all possible combinations on the training data. The resulting features are then used to train a Tree Bagging Conformal Predictor, which can be applied to a test instance for providing a prediction set that will satisfy the required confidence level.

The rest of the paper starts with an overview of related work on the detection of seizures in EEG recordings in Section 2. Section 3 gives a brief description of the general CP framework. Section 4 outlines the methodology we follow for data preprocessing. Section 5 defines the developed Non-conformity Measures (NCM) and completes the description of the proposed CP approach. Section 6 presents the experimental setting and performance measures used in our evaluation and reports the obtained experimental results. Finally, Section 7 gives our conclusions and plans for future work.

2. Related Work

EEG is a technique used to capture and record brain activity by measuring voltage fluctuations on different scalp areas according to the international electrode placement system. EEG captures the amplitude and the frequency of the electrical signal. The amplitude ranges from $5\mu V$ to $200\mu V$ and the frequency ranges from 0-100Hz (American-EEG-Society, 1994).

EEG recording can be performed on a specific subject during awake or sleep. While a subject is in the first sleeping state alpha wave activity decreases, beta wave activity increases and theta wave activity slightly increases. At the second sleeping state as sleep deepens, high-voltage theta or delta wave activity appears. Stage 3 and 4 sleep is characterized by progressively higher amplitude and low-frequency delta wave activity (American-

EEG-Society, 1994). While a subject is awake a wide range of waves can be observed. We provide a brief description of the different brain waves below.

The Brain waves are indicated by Greek letters: Delta 0 to 4 Hz, Theta 4 to 8 Hz, Alpha 8 to 12 Hz, Beta 12 to 32 Hz and Gamma 32 to 64 Hz.

Delta: Delta waves have a frequency range below 4Hz. As mentioned above Delta waves are usually observed during sleep state. We can also observe Delta Waves during infancy, or in serious organic brain diseases. Irregular delta wave activity with a frontal emphasis is related to destructive or compressive lesions involving the diencephalon and upper midbrain, to deep frontal lesions, and to acute metabolite and electrolyte disturbances. Animals are known to have more activity in this range (Thakor and Sherman, 2013).

Theta: Theta waves have a frequency range of 4-8Hz. In a child's brain they occur mainly in parietal and temporal regions. In healthy and alert adults, such theta wave activity is generally inconspicuous or absent, but it does appear during periods of disappointment, frustration, stress and certain stages of sleep as mentioned above (Gómez-Gil et al., 2014). It should be noted that theta activity appears after a generalized seizure, in patients with metabolic disorders, white matter encephalopathy, or extensive lesions of the upper brainstem (Thakor and Sherman, 2013).

Alpha: Alpha waves have a frequency range of 8-12Hz. These waves can be recorded from the occipital region (and sometimes from parietal and frontal regions as well) during consciousness, and is weakened by visual and other sensory stimulus. Alpha waves can be observed in a relaxed person and when eyes are closed. As mentioned above Alpha waves tend to disappear in sleeping subjects (Thakor and Sherman, 2013) .

Beta: Beta waves have a frequency range of 12-32Hz. Beta waves can be recorded from the frontal and parietal lobes. Beta waves can be observed during intense mental activity and tension (American-EEG-Society, 1994).

Gamma: Gamma waves have a frequency range of 32-64Hz. Gamma waves are associated with brain processing. Studies using intra cranial electrodes showed that Gamma activity is related with states of high attention, conscious perception, information processing and motor activity.

Unfortunately while recording EEG on the gamma band there are three sources of artifacts: the 1st one is the power source noise 50/60Hz depending on the country, the 2nd one is EMG (Electromiograph) from the scalp and neck muscles and the 3rd one is electrical potentials produced by eye muscle contraction at the start of saccades (saccades are quick, simultaneous movements of both eyes in the same direction). This results to the need of developing effective filters to handle these artifacts Nottage (2009).

In the past decades many researchers investigated if it is possible to extract information from EEG recordings which could lead to the development of an accurate and robust model for automatic seizure detection. In the following we present a brief description for the data and the models used on the literature.

According to Gómez-Gil et al. (2014) off line identification is composed of the following steps:

1. Preprocessing of the input signal: EEG segment is filtered in a way that it will not contain any artifacts or unwanted frequencies. A comparison between different filters like: Chebyshev II, Elliptic, Equiripple and Least Squares takes place.

Table 1: A brief overview of related work together with the corresponding accuracy.

Author(s)	Feature extraction method	Accuracy(%)
Ayala et al. (2011)	Interelectrode mean	95.9
Mporas et al. (2014)	Time and frequency domain feature	90
Kumar et al. (2014)	DWT-based ApEn	60-100 (depending on subject)
Gómez-Gil et al. (2014)	Maximal overlap discrete wavelet transform over delta and alpha bands of segments of 23.6 s, previously filtered using a 10-order Butterworth low pass filter	90
Gómez-Gil et al. (2014)	Chebyshev II filtering, DWT with Haar wavelet over segments of 1 s	99.26

2. Feature extraction: Using signal technics, stats and math each segment is transformed or converted to the appropriate value/s such that the classifier could extract the appropriate information for robust classification. In the particular paper Discrete wavelet transformation (DWT) and maximal overlap discrete wavelet transform (MODWT) are used as feature extractors obtaining sub-bands Alpha and Delta. Next for each sub-band the mean, absolute mean and variance amplitude is calculated. The feature vector has a dimensionality of 6.

3. Recognition: The feature vectors are the inputs for the sigmoid Feed Forward Neural Network (FF-NN) that will calculate classification probability. Gómez-Gil et al. (2014) have tested different Neural Networks (NNs) with different numbers of nodes in the hidden layer. The training algorithm used is Levenberg Marquardt at a learning rate of 0.5, with a maximum of 1,000 epochs using 3 fold cross validation.

This study concluded that the best results are obtained using segments of one second denoised using a Chebyshev II filter, features are obtained using a DWT with a Haar wavelet and fitted a FF-NN with 18 hidden Neurons. In this case an accuracy of 99.26% is obtained, with a sensitivity of 98.93% and a specificity of 99.59%. For the 23.6 second segments the best results are obtained when Least Squares Filter combined with a DWT-Db2 wavelet and fitted a FF-NN with 6 hidden neurons. In this case an accuracy of 93.23% is obtained, with a sensitivity of 93.87 % and a specificity of 90.07%. In both cases EEG was recorded on healthy subjects with open eyes (Extra-cranial) and Intra-cranial EEG recording of seizure stages, the EEG data has been provided by the University of Bonn.

Ayala et al. (2011) use a similar approach of detecting seizures on EEG recordings using a FF-NN based on the gamma waves (36-44Hz). This study has shown that the power amplitude contains the information needed to classify seizure from nonseizure with an accuracy of 95.90%, a sensitivity of 92.59%, and a specificity of 96.84%. A FF-NN with topology 2-5-1 and sigmoidal activation functions at second and third layer was used. Data was recorded during presurgical monitoring at the Miami Children Hospital at a 500Hz sampling frequency involving 14 patients aged from 3 to 17 years old. The features that fitted the FF-NN were: A) the time duration, defined as the time duration of each given point(S) of the interelectrode mean exceeding the set statistical threshold, which was evaluated as avg(S)+std(S), B) the maximum value of S. Interelectrode mean is the average of all electrodes power spectrum, which contributes to handle EEG recordings from patients with different number of electrodes.

Kumar et al. (2014) have proposed an approach which is based on wavelet decomposition of the EEG into the sub-bands A1 (0–43.4Hz), A2 (0–21.7 Hz), A3 (0–10.85 Hz), A4 (0–5.43 Hz),A5 (0–2.70 Hz), D1 (43.4–86.8 Hz), D2 (21.7–43.4 Hz), D3(10.85–21.7 Hz), D4 (5.42–10.85 Hz), and D5 (2.70–5.43Hz) using DWT. After sub-band extraction ApEn (Approximate Entropy) for all sub-bands (D1–D5) and (A1–A5) is calculated and used as NN feature vectors. The analysis concluded that during seizure activity, EEG had lower ApEn (Approximate Entropy) values compared to normal EEG. This means that epileptic EEG is more predictable than the Healthy EEG. In this study, feed-forward back-propagation neural network has been used for classification and the training algorithm was Levenberg Marquardt. The results indicate an accuracy which ranges from 60 to 100% depending on the subject. The EEG data used in this study was provided by the university of Bonn and contains recordings from five healthy subjects and five non healthy subjects, the data was recorded at a sampling rate of 173Hz.

Mporas et al. (2014) investigated the performance of a seizure detection technique using as input data streams from electroencephalographic and electrocardiographic (ECG) recordings. Data was collected from the Department of Clinical Neurophysiology from 3 patients with diagnosed idiopathic generalized epilepsy. The EEG (21 electrodes) and ECG data were recorded with a sampling frequency equal to 500 Hz and the recordings were manually annotated by neurological experts. EEG and ECG recordings were frame blocked into 1 second segments.

Table 1 presents the summary of the mentioned related work.

3. Conformal Prediction

This section gives a brief description of the main principles of CP. For more details see Vovk et al. (2005).

Let $A = \{(x_i, y_i)|i = 1, \ldots, N\}$ denote our training set, where x_i is an object given in the form of an input vector or matrix, $R = \{t_1, \ldots, t_c\}$ is the set of possible labels and $y_i \in R$ is the label of the corresponding input vector or matrix. Let $B = \{X_k|k = 1, \ldots, M\}$ denote our test set, where X_k is a test instance (vector or matrix). We define as $C_{k,l} = A \cup \{(X_k, t_l)\}$, where $t_l \in R$, the training set extended with the test example X_k together with candidate label t_l. These sets will lead us to assessing predictions with confidence measures and finding which candidate labels are possible for the test instance X_k given a desired confidence level.

A *nonconformity* score (NCS) is a numerical value assigned to each instance that indicates how unusual or strange a pair (x_s, y_s) is, based on the underlying algorithm, where $s \in \{1, \ldots, N, new\}$ is the index of the sth element in $C_{k,l}$. In particular, the underlying algorithm is trained on the instances belonging to $C_{k,l}$, for each $l \in \{1, \ldots, c\}$ and $k \in \{1, \ldots, M\}$, and the NCM uses the resulting model to assign a NCS $\alpha_s^{k,l}$ to each example in $C_{k,l}$.

For every test example k we have c sequences of NCS denoted as $H_{k,l}$. Every sequence is used to find the p-value of a test example k with a candidate label t_l. Given a sequence $H_{k,l}$ of NCS $\alpha_s^{k,l}$ we can calculate how likely a test instance (X_k, t_l) is with the function:

$$p_k(t_l) = \frac{|\{\alpha_s^{k,l} \in H_{k,l}|\alpha_s^{k,l} > \alpha_{new}^{k,l}\}| + \tau|\{\alpha_s^{k,l} \in H_{k,l}|\alpha_s^{k,l} == \alpha_{new}^{k,l}\}|}{N+1}, \quad (1)$$

where $\alpha_{new}^{k,l}$ is the NCS of the k^{th} example in the test set with candidate label t_l and τ follows uniform distribution (0,1).

Given a pair (X_k, t_l) with a p-value of δ this means that this example will be generated with at most δ frequency, under the assumption that the examples are exchangeable, proven in Vovk et al. (2005).

After all p-values have been calculated they can be used for producing prediction sets that satisfy a preset confidence level $1 - \delta$ (δ is called the significance level). Given the significance level δ, a CP will output the prediction set:

$$\{t_l|p_k(t_l) > \delta\}.$$

We would like prediction sets to be as small as possible. The size of prediction sets depends on the quality of the p-values and consequently on the NCM used.

If we want only a single prediction, or *forced prediction*, the CP outputs the label t_r with

$$r = \arg\max_{l=1,\ldots,c} \; p_k(t_l),$$

in other words the t_l with the highest p-value. This prediction is complemented with measures of *confidence* and *credibility*. Confidence is defined as one minus the second largest p-value. Confidence is a measure that indicates the likelihood of a predicted classification compared to all the other possible classifications. Credibility is defined as the largest p-value. Low credibility means that either the data violate the exchangeability assumption or the particular test example is very different from the training set examples.

4. Preprocessing

From each EEG clip we first extract the several brain wave bands using wavelet analysis. After we extract the brain wave bands we then apply Fast Fourier Transformation. The result is used to extract 7 features.

4.1. Wavelet analysis

Wavelet analysis is a signal processing technique, which is applied in non-periodic and noisy signals. Its based on calculating a correlation among a signal and a basis function, known as the wavelet function. This is calculated for different time steps and frequencies. A wavelet function $\phi(.)$ is a small wave, which means that it oscillates in short periods of time under the following conditions:

- Its energy is finite, that is:
$$\int_{-\infty}^{\infty} |\phi(t)|^2 dt < \infty. \tag{2}$$
- Its admissibility constant C is finite, that is:
$$\int_{0}^{\infty} \frac{|\Phi(u)|^2 du}{u} < \infty, \tag{3}$$
where
$$\Phi(u) = \phi(t) e^{-2\pi i t}. \tag{4}$$

The Discrete wavelet transformation (DWT) calculation is based on the fact than any function can be expressed as a linear decomposition:

$$f(t) = \sum_{l} a_l \phi_l(t), \tag{5}$$

where a_l are real coefficients and $\phi_l(t)$ are real functions.

Thus we derive the DWT:

$$f(t) = \sum_{k} \sum_{j} a_{j,k} \phi_{j,k}(t), \tag{6}$$

where the $a_{j,k}$ are known as the Discrete Wavelet Transform (DWT) of f(t). In figure 1 we show a wavelet analysis of an EEG with sampling frequency=4096 Hz using DWT and inverse wavelet transformation (iDWT) (Gómez-Gil et al., 2014).

In order to have a robust classification model the appropriate filter should be chosen. We have chosen a Chebyshev II filter and a DWT-Haar wavelet based on the findings of Gómez-Gil et al. (2014).

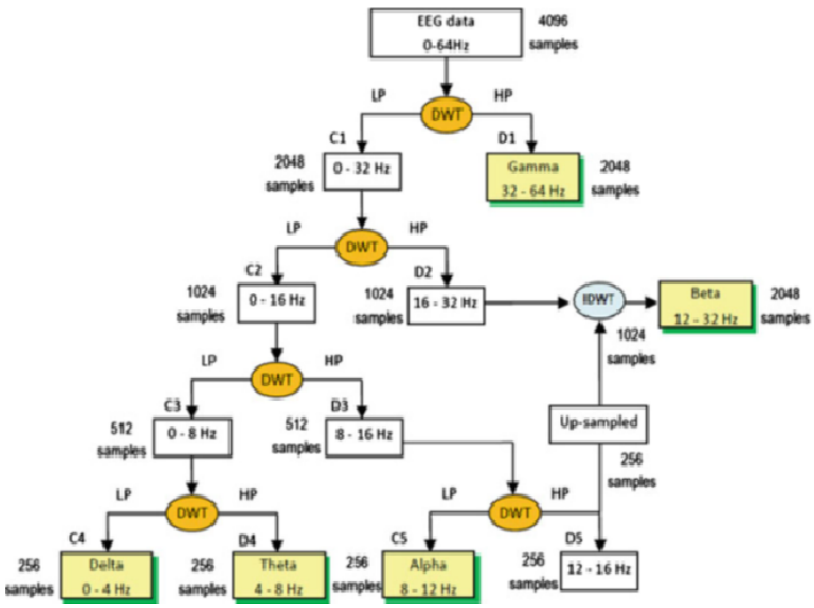


Figure 1: Wavelet analysis of an EEG using a DWT. The yellow blocks correspond to identiers of sub-bands found in a EEG (see Table 1). LP Low-pass filter, HP High-pass filter, C# Approximation coefficient number #, D# Detailed coefficient number# (Gómez-Gil et al., 2014).

4.2. Fast Fourier Transformation

The Discrete Fourier Transformation (DFT) is an algorithm which decomposes a sequence of values into components of different frequencies. It converts a finite list of equally spaced samples of a function into the list of coefficients of a finite combination of complex sinusoids, ordered by their frequencies, that have those same sample values:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}, k = 0, ..., N-1, \tag{7}$$

where X_k is a complex number that encodes both amplitude and phase of a sinusoidal component of function x_n. The sinusoid's frequency is k/N cycles per sample. In our case x_n is a vector component of a vector produced from DWT. Its amplitude and phase are:

$$|X_k| = (\mathrm{Re}(X_k)^2 + \mathrm{Im}(X_k)^2)^{0.5}/N, \tag{8}$$

$$\mathrm{Arg}(X_k) = -i \cdot \ln(X_k/(|X_k|)). \tag{9}$$

4.3. Features

From the amplitude sequence for every sub-band we calculate the mean, std, skewness, kurtosis, max, min and the median.

5. Nonconformity measures for Tree bagging

This section describes the Tree Bagging Conformal Predictor (TB-CP), which uses tree bagging Breiman (1996) as underlying technique combined with the CP framework discussed in Section (3).

5.1. Tree bagging

Tree bagging is an ensemble based method used for classification or regression. It consists of a predefined number of tree classifiers. Each tree is trained using a portion of the training instances with a subset of the available variables. In this study we used the classification tree bagger thus each instance is classified as the label with the higher percentage of voting between the decision trees. All of our tree baggers have been trained in MATLAB (2010) using the default settings of the TreeBagger method and with the number of trees set to 100.

5.2. Nonconformity measures

In this section we provide a description of the NCM we have used in this study. The NCM is based on the Tree bagging classifier described in Section 5.1. Recall from Section 3 that $C_{k,l} = A \cup \{(X_k, t_l)\}$, where $\{t_1, \ldots, t_c\}$ are the possible labels. For each test example X_k TCP generates $C_{k,1}, \ldots, C_{k,c}$ and assigns a NCS to each example in each of the c sets. We denote as $z_s^{k,l}$ the sth element of $C_{k,l}$ and as $\alpha_s^{k,l}$ its NCS, with $s = 1, \ldots, N, new$.

The nonconformity measure we used is:

$$\alpha_s^{k,l} = -D_j^s, \tag{10}$$

where D_j^s is the percentage of trees with prediction t_j for instance s. It should be noted that when $s = new$ we use $y_s = t_l$ (the candidate class as defined 3). After calculating the NCS we calculate p-values and make predictions following the process described in Section 3.

6. Experiments and Results

6.1. Dataset

The dataset we used, available online (Andrzejak et al., 2001), contains EEG recordings for both healthy and epileptic subjects. Specifically it contains five sets (denoted A-E) each one consisting of 100 EEG segments of 23.6-sec in duration. Segments of set A and B were recorded from surface EEG and belong to five healthy volunteers. Set A was recorded with the volunteers having their eyes open while set B was recorded with their eyes closed. Segments from C, D and E were taken from 5 patients. Segments in set D were recorded within the epileptic zone while those in set C from the hippocampal formation of the opposite hemisphere of the brain. Note that sets C and D contain EEG's during seizure free intervals. Segments in set E contain EEG's during seizure activity. For more information refer to Andrzejak et al. (2001) . In our analysis we have used four different classification problems following Tzallas et al. (2007).

- The first consists of three classes, the first class refers to the segments belonging to sets A and B (healthy subjects, 200 segments), the second class contains the segments belonging to sets C and D (seizure free subjects, 200 segments) and the third class contains the segments belonging to set E (seizure subjects, 100 segments).
- The second deals with classifying segments as non-seizures or seizures. The non-seizures class consists of the sets A, B, C and D (400 segments), while the seizure class consists of set E (100 segments).
- The third is similar to the first but the first class consists of set A only (100 segments) and the second class of set C (100 segments) only.
- The forth is similar to the second but the first class consists of set A only and second class of set E only.

6.2. Experimental Setting and Performance Measures

In this section we detail the experiments and results of the proposed approach on the four classification problems discussed in 6.1.

Before feature extraction a preprocessing takes place. In order to have a robust classification model the appropriate filter should be chosen. We have chosen a Chebyshev II filter and a DWT-Haar wavelet, based on the findings of Gómez-Gil et al. (2014). As mentioned by Gómez-Gil et al. (2014) in the filtering process we first remove frequencies greater than 64 Hz. We have applied discrete wavelet analysis to each 23.6 segment and we have extracted the frequencies Delta(0-4hz), Theta(4-8hz), Alpha(8-12hz), Beta(12-30hz), Gamma(30-64hz) as shown in Figure 1. Then in each sub band we applied a fast Fourier transformation (FFT) and extracted the following features: average, standard deviation, skewness, kurtosis, max, min, median.

After extracting the features for each sub-band we have to choose which sub-bands should be included. First we split our dataset into training set and test set each containing 80% and 20% of the total instances respectively. Then we apply a ten fold cross validation process on the training set to choose the appropriate sub-bands in a way that the accuracy is maximized. Because we make an exhaustive search on the sub-band selection the ten fold cross validation was performed $2^5 - 1$ times. This process is repeated 100 times and each time we keep the selected sub bands and the training and test sets fixed. The features of the selected sub-bands are then used for training and testing the Tree Bagging Conformal Predictor.

Due to the fact that the accuracy itself is not a good indication for measuring the performance of a CP we used four probabilistic criteria for evaluating p-values proposed by Vovk et al. (2016). These criteria are divided into two main categories called *Basic Criteria*, which do not take into account the true label, and *Observed Criteria*, which take into account the true label. The two Basic Criteria we used are:

The S ("sum") criterion

$$\frac{1}{M}\sum_{l=1}^{c}\sum_{k=1}^{M} p_k(t_l), \tag{11}$$

where $p_k(t_l)$ is the p-value of the test example X_k with candidate label t_l as in equation (1) In effect the S-criterion is the average sum of all p-values.

The N ("number") criterion

$$\frac{1}{M}\sum_{k=1}^{M}|\{t_l|p_k(t_l) > \delta\}|, \tag{12}$$

which is the average size of the prediction sets with respect to a confidence level $1-\delta$.

The two Observed Criteria we used are:

The OF ("observed fuzziness") criterion

$$\frac{1}{M}\sum_{k=1}^{M}\sum_{l,t_l\neq t_k} p_k(t_l), \tag{13}$$

which uses the average sum of the p-values of the false labels.

The OE ("observed excess") criterion

$$\frac{1}{M}\sum_{k=1}^{M}|\{t_l|p_k(t_l) > \delta, t_l \neq t_k\}|, \tag{14}$$

which represents the average number of false labels included in the prediction sets, with respect to a confidence level $1-\delta$.

For all criteria smaller values indicate more informative p-values.

6.3. EEG Classification results

6.3.1. Accuracy

Table 2 presents the accuracy of the conventional Tree Bagger technique with the selected features on the four classification problems described in Subsection 6.1, while Table 3 reports the accuracy of the corresponding Conformal Prediction technique along with the average confidence and credibility measures it produced. The accuracies reported in these tables are quite high for all problems and comparable to that of the related work presented in Section 2. In comparing the accuracy of the Conformal Predictor technique with that of its conventional counterpart, we observe that the first leads to a very small decrease, which is probably due to some nonconformity scores being equal. However this decrease is not very significant especially since our main interest is on the prediction sets provided by CP.

Table 2: Accuracy of the Underlying model

Classification problem	Accuracy(%)
1	97.27
2	99.53
3	96.17
4	99.55

Table 3: Average accuracy, credibility and confidence of the proposed CP

	Classification Problem	
Accuracy	1	96.92
	2	99.43
	3	95.77
	4	99.43
Average confidence	1	99.33
	2	99.78
	3	99.01
	4	99.47
Average credibility	1	50.82
	2	50.22
	3	50.29
	4	50.13

Table 4: Unobserved criteria

Classification Problem	S criterion	N criterion (per significance level) 0.005	0.01	0.05	0.10
1	0.52	1.51	1.16	0.97	0.91
2	0.50	1.00	0.99	0.95	0.90
3	0.52	1.98	1.40	0.98	0.92
4	0.51	1.47	1.14	0.95	0.90

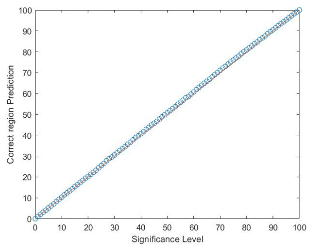
6.3.2. Empirical Validity

Figure 2, presents the percentage of correct region predictions(the percentage of the number of the sets that contain the correct class) as a function of the confidence level for the four classification problems. In all cases the accuracy is equal to the required confidence level, as guaranteed by Conformal Prediction. This shows that we can effectively control the number of errors made and therefore the upper bound on the probability of discarding a seizure segment.

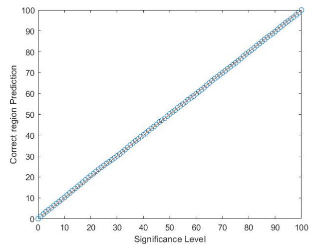
6.3.3. Informational Efficiency

In this study the main objective is to complement the single predictions with probabilistic measures of confidences and provide prediction sets with respect to a confidence level. Here we investigate how informative our p-values are and the practical usefulness of our prediction sets. This is done following the informational efficiency criteria described in Subsection 6.2 and proposed by Vovk et al. (2016).

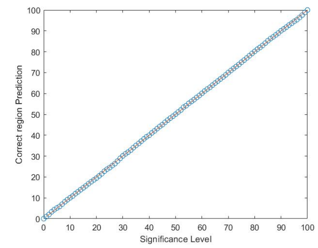
Table 4 presents the values of the two unobserved criteria for the four classification problems. Specifically the second column of the table contains the values of the S cri-



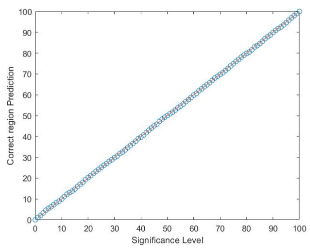
(a) Class. Problem 1



(b) Class. Problem 2



(c) Class. Problem 3



(d) Class. Problem 4

Figure 2: Percentage of correct region predictions of the four problems

Table 5: Observed criteria

Classification Problem	OF criterion	OE criterion (per significance level) 0.005	0.01	0.05	0.10
1	0.011	0.52	0.16	0.02	0.01
2	0.003	0.01	0.00	0.00	0.00
3	0.017	0.99	0.42	0.03	0.01
4	0.008	0.48	0.15	0.00	0.00

terion, while the rest of the columns present the N criterion for the significance levels $0.005, 0.01, 0.05, 0.10$. In the same manner Table 5 presents the values of the two observed criteria. The second column contains the values of the OF criterion, while the rest of the columns give the values of the OE criterion for the significance levels $0.005, 0.01, 0.05, 0.10$. The values of all criteria show the quality of the resulting pvalues, furthermore the values of the N and OE criteria also demonstrate the practical usefulness of the produced prediction sets.

At 99.5% confidence for the two class problems the prediction sets contain on average 1 and 1.5 labels, while the number of false labels is 0 and 0.5 respectively. By lowering the significance level to 99% confidence the prediction sets contain on average approximately one label, while the number of false labels is 0 and 0.15.

For the case of the three labels problem at 99.5% confidence the prediction sets contain on average 1.5 and 2 labels respectively, while the number of false labels is 0.5 and 1 respectively. By lowering the significance level to 99% confidence the prediction sets contain on average 1.16 and 1.40 labels, while the number of false labels equals to 0.16 and 0.42 respectively.

Setting the significance level to 95% for all classification problems the average number of labels in the prediction sets is approximately one while the number of false labels is almost zero.

7. Conclusions

This work examines the application of Conformal Prediction to the problem of detecting seizures in EEG recordings. This is an important task as EEG recordings of seizures are of primary interest in the evaluation of epileptic patients while the analysis of long term EEG recordings by neurophysiologists is a time consuming task. An important advantage of the proposed approach is that it enables control of the percentage of erroneous prediction regions it produces. Therefore by providing the neurophysiologist with all segments of which the prediction regions contain the seizure class we can guarantee an upper bound on the frequency with which a seizure will not be detected.

We examine the performance of the proposed CP approach on four classification problems. Our results show that the accuracy of its forced predictions is comparable to that of other techniques proposed in the literature. We also demonstrate that the prediction regions it produces are well-calibrated, therefore the error rate of these prediction regions can be exactly controlled by setting the required confidence level. Finally our results show that the proposed approach produces practically useful prediction regions for confidence levels as high as 99.5%.

As the datasets used in this study are relatively small, our first future plan is to examine the performance of the proposed approach on larger datasets, such as the ones available at https://www.kaggle.com/c/seizure-detection/data. With the use of larger datasets even higher confidence levels will be possible. Furthermore, we plan on examining and comparing other underlying approaches and nonconformity measures on the particular task.

References

American-EEG-Society. Guideline fifteen: guidelines for polygraphic assessment of sleep-related disorders (polysomnography). american electroencephalographic society. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, 11 1:116–24, 1994.

Ralph Gregor Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E*, 64:061907, Nov 2001. doi: 10.1103/PhysRevE.64.061907. URL https://link.aps.org/doi/10.1103/PhysRevE.64.061907.

Melvin Ayala, Mercedes Cabrerizo, Prasanna Jayakar, and Malek Adjouadi. Subdural eeg classification into seizure and nonseizure files using neural networks in the gamma frequency band. 28:20–9, 02 2011.

Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug 1996. ISSN 1573-0565. doi: 10.1007/BF00058655. URL https://doi.org/10.1007/BF00058655.

Pilar Gómez-Gil, Ever Juárez-Guerra, Vicente Alarcón-Aquino, Manuel Ramírez-Cortés, and José Rangel-Magdaleno. *Identification of Epilepsy Seizures Using Multi-resolution Analysis and Artificial Neural Networks*, pages 337–351. Springer International Publishing, Cham, 2014. ISBN 978-3-319-05170-3. doi: 10.1007/978-3-319-05170-3_23. URL https://doi.org/10.1007/978-3-319-05170-3_23.

Amir Harati, Silvia Lopez, Iyad Obeid, Joseph Picone, M P. Jacobson, and Steven Tobochnik. The tuh eeg corpus: A big data resource for automated eeg interpretation, 12 2014.

Yatindra Kumar, Mohan Lal. Dewal, and R. S. Anand. Epileptic seizures detection in eeg using dwt-based apen and artificial neural network. *Signal, Image and Video Processing*, 8 (7):1323–1334, Oct 2014. ISSN 1863-1711. doi: 10.1007/s11760-012-0362-9. URL https://doi.org/10.1007/s11760-012-0362-9.

Klaus Lehnertz, Florioan Mormann, Thomas Kreuz, Ralph Gregor Andrzejak, Cristoph Rieke, Peter David, and Cristian E. Elger. Seizure prediction by nonlinear eeg analysis. *IEEE Engineering in Medicine and Biology Magazine*, 22(1):57–63, Jan 2003. ISSN 0739-5175. doi: 10.1109/MEMB.2003.1191451.

MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

Florian Mormann, Ralph Gregor Andrzejak, Christian E. Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314–333, 2007. doi: 10.1093/brain/awl241. URL +http://dx.doi.org/10.1093/brain/awl241.

Iosif Mporas, Vasiliki Tsirka, Evangelia Zacharaki, Michalis Koutroumanidis, and Vasileios Megalooikonomou. Online seizure detection from eeg and ecg signals for monitoring of epileptic patients, 05 2014.

Judith Nottage. Uncovering gamma in visual tasks. 23:58–71, 12 2009.

Nitish V. Thakor and David L. Sherman. *EEG Signal Processing: Theory and Applications*, pages 259–303. Springer US, Boston, MA, 2013. ISBN 978-1-4614-5227-0. doi: 10.1007/978-1-4614-5227-0_5. URL https://doi.org/10.1007/978-1-4614-5227-0_5.

Alexandros T Tzallas, Markos Tsipouras, and Dimitrios I Fotiadis. Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Intell. Neuroscience*, 2007:18:1–18:13, August 2007. ISSN 1687-5265. doi: 10.1155/2007/80510. URL http://dx.doi.org/10.1155/2007/80510.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer Science & Business Media, 2005.

Vladimir Vovk, Valentina Fedorova, Ilia Nouretdinov, and Alexander Gammerman. Criteria of efficiency for conformal prediction. pages 23–39, 2016. doi: 10.1007/978-3-319-33395-3_2. URL http://dx.doi.org/10.1007/978-3-319-33395-3_2.

Elizabeth J. Waterhouse. New horizons in ambulatory electroencephalography. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):74–80, May 2003a. ISSN 0739-5175. doi: 10.1109/MEMB.2003.1213629.

Elizabeth J. Waterhouse. New horizons in ambulatory electroencephalography. *IEEE Engineering in Medicine and Biology Magazine*, 22:74–80, 2003b.