

Wavelet Time Scattering Based Classification of Interictal and Preictal EEG Signals

Afreda A. Susu^{1*}, H.A. Agboola¹, C. Solebo², F.E.A. Lesi³ and D.S. Aribike¹

¹Department of Chemical and Petroleum Engineering, University of Lagos, Lagos, Nigeria

²Department of Potters Bar Clinic, Elysium Healthcare, Great Ormond Street Institute of Child Health, London, United Kingdom

³Department of Paediatrics, University of Lagos, Lagos, Nigeria

Abstract

If it were possible to reliably identify the preictal brain state from dynamical changes in EEG data of epilepsy patients, then the age long problem of actualizing a fully automated closed-loop seizure – warning or seizure-prevention system that is clinically deployable would have been resolved. Accordingly, through feature engineering, a great deal of effort has been invested over the discovery of EEG features or measures that are always indicative of the preictal brain state. However, this has proven to be difficult, time consuming and apparently unsuccessful. Therefore, lately, attention has shifted to feature learning-methods that automatically learn and extracts useful discriminatory features from raw data. This paper studies the efficacy of wavelet time scattering learned EEG features for interictal and preictal EEG classification. Wavelet time scattering network developed in Matlab and two different EEG datasets: CHB-MIT scalp EEG and AES intracranial EEG datasets were used for the study. The learned interictal and preictal EEG features were used to train and evaluate a simple binary support vector machine classifier. Three different classification accuracy results namely ordinary cross validation, true cross validation and test classification accuracy results were reported for the analysis. Mean classification accuracy values of 93.15%, 97.57% and 91.33% were obtained respectively for the scalp EEG while mean classification accuracy values of 98.33%, 100% and 96.73% were obtained respectively for the intracranial EEG. A general comparison showed that the combination of wavelet time scattering learned EEG features and a simple binary support vector machine classifier performed equally or even better than deep convolutional neural networks in EEG classification tasks. Finally, wavelet time scattering has proven to be a very good EEG feature learner and may greatly improve the sensitivity and specificity of seizure prediction algorithms.

Keywords: Electroencephalogram • Interictal • Preictal EEG • Wavelet time scattering

Introduction

Epilepsy is a chronic brain disorder that constitutes a major public health concern. It affects more than 50 million people worldwide [1]. The prevalence of epilepsy is particularly high in developing countries especially Latin America and several African countries, notably Liberia, Nigeria and the United Republic of Tanzania [2]. The hallmark of epilepsy is recurrent and spontaneous seizures which are caused by parts of the brain eliciting abnormally synchronous electrical activity. These seizures not only disrupt normal living but can also cause mental and physical damage and in extreme cases, even death.

Generally, the causes of epilepsy can be classified into three broad categories: genetic, cryptogenic (unknown or hidden causes) and others (head trauma, brain tumors etc.). Experts believe that genetic predisposition combined with environmental conditions lead to epilepsy in some patients. The affected genes are often those that control the excitability of neurons in the brain [3].

In the majority of epileptic cases, accurate diagnosis of the disease can be made with treatment in the form of regular use of Anti-Epileptic Drugs (AEDs) but there are concerns about the side effects of these drugs. Also quite a number of epileptic patients suffer from drug resistant epilepsy and may require surgical measures which involve excision of relatively large amount of brain tissue. Apart from the fact that surgery raises concern about neurological disability that may result by the removal of either normal or functionally necessary tissue, there have been reported cases of seizures in quite a number of patients who had resection [4]. In summary, AEDs have side effects and refractory epilepsy has defied existing treatment protocols.

Therefore, researchers are currently looking for alternative therapeutic strategy for epilepsy.

One good strategy is seizure prediction. If seizures are predicted well ahead of time, patients with refractory epilepsy will have ample time to prepare and guide against injuries and sudden deaths. In addition, since anti-epileptic drugs would forthwith be administered on-demand (i.e. after the prediction of an impending epileptic seizure), dose – related side effects in patients placed on anti-epileptic drugs will be greatly reduced. Furthermore, other emerging interventional therapeutic approaches such as electrical neurostimulation, optogenetics, drug perfusion and focal cooling require devices whose animation will be triggered by a reliable, accurate and timely seizure prediction algorithm. Such algorithm can be used for the interruption of seizure-generating mechanisms and, also, to avert impending seizures.

Since the pioneering works of Viglione and colleagues aimed at predicting seizure through EEG data analysis, a huge amount of studies have since been carried out but to date this problem has not been satisfactorily solved. This has been a great source of concern to the international research community such that biennially, the International Workshop on Seizure Prediction (IWSP) is held. The IWSPs are a forum that brings together an international interdisciplinary group of epileptologists, engineers, physicists, mathematicians, neurosurgeons and neuroscientists with the goal of developing engineering-based epilepsy treatments [5].

The main steps involved in seizure prediction workflow include EEG data acquisition and preprocessing, Feature extraction and design of preictal state identification method (i.e. classification or thresholding). The hypothesis that there exists a transition state (preictal) between the

*Corresponding Author: A.A. Susu, Department of Chemical & Petroleum Engineering, University of Lagos, Lagos, Nigeria; E-mail: alfredasusu@gmail.com

Copyright: © 2020 Susu AA, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: October 07, 2020; **Accepted:** October 21, 2020 ; **Published:** October 28, 2020

interictal (normal) and the ictal (seizure) brain states is central to the seizure prediction idea and there are a number of clinical evidences that support this hypothesis. These evidence include increases in cerebral blood flow, cerebral oxygenation and many others [6]. If reliable seizure prediction from dynamical changes in the EEG is possible, then, the dream of actualizing a clinically deployable closed-loop seizure – warning/aborting system would be possible. Accordingly, a great deal of effort has been invested over characteristic EEG features that are always indicative of the preictal brain state. In fact, feature extraction step has been described as the most difficult steps in the seizure prediction workflow [7]. Feature engineering which is the process of using domain knowledge to manually extract features has been the traditional approach to EEG feature extraction for seizure prediction. The main disadvantage of feature engineering is that it is error prone because it is tedious, time consuming and dependent on domain knowledge. Feature learning or representation learning is an automated feature extraction scheme that improves upon the standard workflow by automatically extracting meaningful and useful features. In particular, deep learning models such as deep Convolutional Neural Networks (CNNs) nowadays provide state-of-the-art solutions to many problems in computer vision or image classification, speech recognition, natural language processing, etc. These models can learn data representation at different levels of abstraction and extract complex features from input raw data. However, CNNs require large number of labeled learning data and computational needs owing to the large number of hyperparameters to be learned in the network. In addition, features extracted by CNNs can be difficult to interpret [8].

Interestingly, a technique which addresses the challenges of deep CNNs has been proposed. This technique is called group invariant scattering [9]. Group invariant scattering is also referred to as wavelet time scattering or scattering transform. The scheme uses multi-layered network involving fixed wavelet kernel based transform. Physiological signals often portray certain variabilities that are not really important for classification task (i.e. the variabilities are not useful in determining the class of the signals). Example of such variabilities include shifting and stretching in time and transposition in frequency. Since those things that are important for a given signal classification task are usually unknown, scatter transform takes a conservative approach and creates representations of the raw data that are invariant to variabilities that don't affect the class of the signals while still preserving as much information in the data so as to keep other variabilities that are important in determining class of signals. An immediate advantage of scatter transform is that it allows us to construct classification models that don't require as much training data [10].

In this paper, we investigate the capacity of scattering transform to automatically extract EEG features that are efficient in predicting seizure occurrence by identifying a preictal brain state. Since the success or failure of a seizure prediction algorithm is highly dependent on how well and how consistent its classifier can correctly classify interictal and preictal EEG data epochs of epileptic patients, the efficacy of the obtained scattering features for seizure prediction may, in the first instance, be accessed by using the scattering features to build and evaluate a simple binary classifier such as a linear support vector machine. The efficacy of wavelet time scattering has already been demonstrated in unsupervised anomaly sensing based seizure detection and unsupervised anomaly sensing based seizure prediction algorithms [11]. This paper provides an introductory framework for the use of scattering transform in supervised anomaly sensing based seizure prediction algorithms. The remaining part of the paper is organized as follows. A brief description of the workings of wavelet time scattering is given in section II. Section III gives detailed description of the methods used while results and discussion are presented in section IV. We give our conclusion and future directions in section V.

Scattering transform

Scattering transform or wavelet time scattering is a technique used to derive low-variance features from real-valued time series data or signals. Its historic frame of reference starts with the Fourier transform which is often referred to as the canonical signal processing technique. A major setback

of the Fourier transform is its inability to localize frequency information contained in signals. As a result, the technique exhibit very high instability to signal deformations at high frequency. This means that the spectrogram representation of a signal and its slightly flustered version through high frequency deformation will look different even though the two signals still look very similar. This instability property of Fourier transform to signal deformation is attributed to the non-localized support property of sine wave which happens to be the major building block of the Fourier transform [12]. To fix this problem, the wavelet transform concept was developed [13]. In wavelet transform, signals are decomposed using a dictionary of wavelets having localized support property but with variant dilation and thus, the emerging representation exhibits high frequency components localization of signals.

In finding good data representation for pattern classification or recognition problems, there exists another important property which is often desired in the representation. This property is called translation invariance. A representation or transformation exhibits translation invariance if under the transformation a signal and its shifted versions have the same representation in the feature space. Wavelet transform lacks this property and it is said to be translation covariant, that is, when a signal is shifted its wavelet coefficients are also shifted. This makes signal classification difficult as a signal and its time shifted version will be assigned to different classes.

The need for building a signal representation that displays translation invariance and stability under deformation properties led to the development of the scattering transform [14]. The fundamental building block of the wavelet time scattering is the Morlet wavelet which is derived from Gaussian windowed sinusoid. A wavelet time scattering framework processes data in stages. The output of one stage becomes the input for the next stage. Each stage consists of three operations, namely, convolution (using wavelets), nonlinearity (by taking modulus) and averaging (using scaling function). In what follows we provide a general description of the steps involved in using wavelet scattering network for feature extraction. Detailed mathematical description of the wavelet scattering framework can be found elsewhere [15].

The scattering transform generates features in an iterative manner (Figure 1). An input signal, y is first averaged using wavelet low pass filter (i.e. convolve the signal with the scaling function, ψ). The results, $y \cdot \psi$, are referred to as the first order scattering coefficients, $Sc(0)$ or layer zero scattering features. With the averaging operation, high frequency details in the signal is lost. The lost details in the first step are captured in the subsequent layers by performing a continuous wavelet transform of the signal to yield a set of scalogram coefficients. A nonlinear operator (a modulus) is applied on the scalogram coefficients and then the output is filtered with the low pass filter to yield a set of layer 1 scattering coefficients, $Sc(1)$.

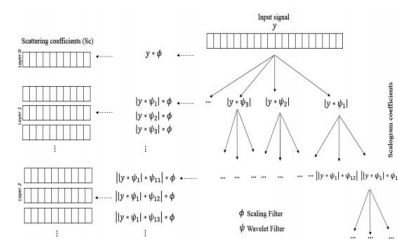


Figure 1. Wavelet Time Scattering Framework. The sequence of edges from the root to a node is referred to as a path. The tree nodes are the scalogram coefficients. The scattering coefficients are the scalogram coefficients convolved with the scaling function. The set of scattering coefficients are the low-variance features derived from the data.

The same process is repeated to obtain the layer 2 scattering coefficients, $Sc(2)$. The output of the scalogram coefficients in the previous layer becomes the input to the operation in the next layer. Then we apply

the same modulus operator and filter the output with the wavelet low pass function to yield the layer 2 scattering coefficients. We can have more than three layers in the scattering network but in practice the energy dissipates with every iteration so three layers are appropriate for most applications [16]. The coefficients are critically down sampled to reduce computational complexity of the network. These coefficients which can be visualized and interpreted are collectively referred to as the scattering features. A wavelet scattering network may be referred to as a deep network because it performs the three major tasks that make a deep network: convolution, nonlinearity and pooling. Convolution is performed by wavelets, the modulus operator serves as the nonlinearity and filtering with wavelet low pass filter is analogous to pooling.

Methodology

EEG datasets

Two separate EEG datasets: CHB-MIT scalp EEG dataset from PhysioNet and American Epilepsy Society (AES) intracranial EEG dataset from Kaggle were obtained and used for the study. We provide a brief description of each dataset and the extracted EEG data for this study [17].

CHB-MIT dataset

The CHB-MIT database consists of 24 scalp EEG (sEEG) recordings from 23 patients (i.e. one patient has two different recordings) suffering from intractable epileptic seizures. In total the recordings span approximately 982 hours and contains 198 seizures. All signals were sampled at 256 samples per second with 16-bit resolution over 23 electrodes. The International 10-20 system of EEG electrode positions and nomenclature was used for the recordings. The data is grouped into cases with each case containing between 9 and 42 continuous .edf (European data format) files from a single patient. Information about the elapsed time in seconds from the beginning of each .edf file to the beginning and end of each seizure contained in it is also made available in the dataset. The EEG data can be accessed through the PhysioNet website: <http://physionet.org/physiobank/database/chbmit/>.

The EEG data files which were recorded in the European data format (Edf) were converted into Matlab files through BIOSIGtoolbox for EEGLAB in the Matlab environment. 25 minutes' postictal duration and 60 minutes' preictal duration were assumed. In order to allow for therapeutic intervention, 5 minutes of data immediately preceding seizure was not included in the preictal data [18]. The remaining data were then taken as the interictal. Furthermore, to avoid excessive mixing of pre-seizure, seizure and post seizure EEG data, only patients with at least 2 seizure segments separated by at least a 2-hr period had their data extracted and used for the study [19]. This led to the exclusion of 7 EEG recordings of 6 patients from the study. The number of 60 minutes' preictal EEG data blocks in a patient's data was dictated by number of leading seizure events in the whole EEG recordings of the patient. Equal number of 60 minutes' interictal EEG data blocks were extracted for the patient. The interictal EEG data blocks considered are those recorded at least four hours away from any seizure event. Figure 2 is a Matlab stacked plot showing few samples of interictal and preictal scalp EEG data blocks of one of the patients[20].

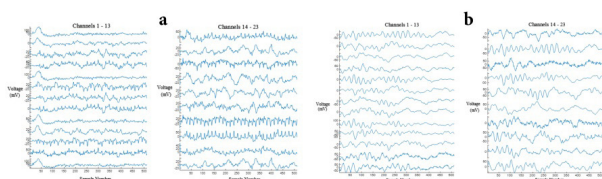


Figure 2. (a) Interictal and (b) Preictal sEEG data blocks of one patient in the CHB – MIT dataset.

Patients' datasets

In preparation for feature extraction, an interictal EEG data block and a preictal EEG data block for a patient were randomly selected and row

concatenated to obtain a data matrix. The number of data matrices for a patient make up the number of datasets for the patient.

Class labels

After obtaining the patients' datasets, a matrix named EEGclassLabels was created. EEGclassLabels is a 2k-by-1 (k is the number of EEG channels) cell array of class labels, one for each row data in the patients' datasets. The two class labels are 'INT' (interictal) and 'PRE' (preictal).

Training and test patients' datasets

Each patient's datasets were randomly split into two sets – the training and test sets. To achieve this, a Matlab function was created. This function takes as inputs a dataset and EEGclassLabels and outputs two datasets (TrainingDataset and TestDataset) along with a set of labels (TrainingLabels and TestLabels) for each. Each element of TrainingLabels and TestLabels contains the class label for the corresponding row of the patients' dataset matrices. 70% of the data in each class was assigned to TrainingDataset while the remaining 30% was held out for testing and was assigned to Test Dataset [21].

AES dataset

National (American) Institutes of Health, the Epilepsy Foundation, and the American Epilepsy Society organized an international competition tagged "American Epilepsy Society Seizure Prediction Challenge" [22]. Its goal was to identify the best model for discriminating between preictal and interictal iEEG clips. iEEG data of five canine subjects with naturally occurring epilepsy and two human subjects with refractory epilepsy were provided for the competition. However, we only accessed data of the two human subjects. For human subject 1, 50 interictal and 18 preictal data training data clips were made available while human subject 2 had 42 interictal and 18 preictal training data clips. Data was sampled at 5000 Hz and each iEEG data clip is ten minutes long. 10 interictal and 10 preictal ten minutes' training data clips were randomly selected for each patient and used for the study. To reduce computational complexity, the iEEG data clips were down sampled to 1000 Hz. Random row concatenation of interictal and preictal EEG data clips were then performed to create ten datasets for each of the subjects. Lastly, creation of class labels and training and test datasets were carried out in the same manner described for the CHB-MIT EEG data. Figure 3 shows Matlab plots of few samples of randomly selected interictal and preictal iEEG signals for the two human subjects.

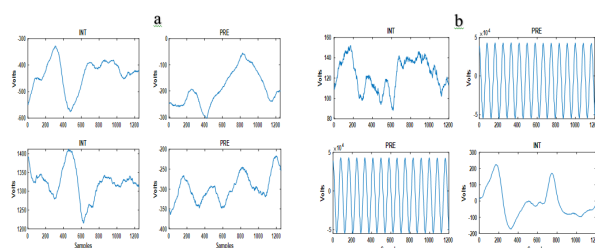


Figure 3. Random interictal (INT) and Preictal (PRE) iEEG data samples of (a) human subject 1 and (b) human subject 2.

Feature extraction

The wavelet time scattering network used for feature extraction was designed using the wavelet toolbox in Matlab. The toolbox uses the Gabor or analytic Morlet wavelet function for signal decomposition. The key parameters to specify are the scale of the time invariant, the number of wavelet transforms or number of wavelet filter banks, and the number of wavelets per octave in each of the wavelet filter banks. In what follows, we give a brief description of each parameter and their typical values.

Invariance scale (t)

Scattering framework is invariant to translations up to the invariance scale which is a duration. The invariance is provided in the framework by application of the scaling filter therefore the time support of the scaling

function does not exceed the size of the invariant. The time support of the wavelet cannot also exceed the invariance scale therefore, the invariance scale also affects the spacing of the center frequencies of the wavelets in the filter banks. These considerations suggest that the choice of the invariance scale is key to obtaining a good representation and therefore, must be carefully chosen. Choosing a suitable invariant scale would require a good understanding of the dynamical changes in the signal which is lacking in the case of EEG or iEEG. Since our task is to classify as interictal or preictal, 1-hour sEEG or 10-minute iEEG segments which represents signal from one sEEG/iEEG channel in the patients' datasets we chose an invariant scale of 1 hour for the sEEG data and 10-minute for the iEEG data.

Number of layers or wavelet filter banks (n)

The choice of number of filter banks in the network is usually dictated by the fact that energy in the current layer is substantial enough for another useful successive layer. 2-layer networks have been shown to be sufficient for many applications particularly audio signal classification and intracranial EEG signal processing thus two wavelet banks (Figure 4) were used in the network [23].

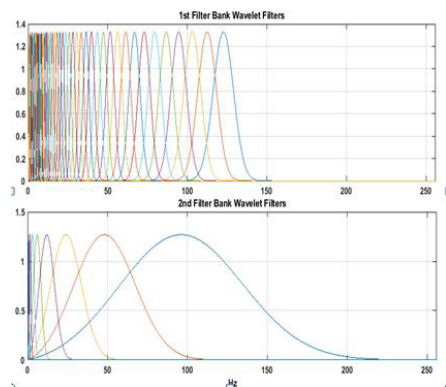


Figure 4. Wavelet filter banks used in the scattering transform network for sEEG.

Quality factors (q)

An advantage of continuous wavelet transform over discrete wavelet transform is the added flexibility of analyzing signals at intermediary scales within each octave. This often allows for a fine scale analysis. The number of wavelet per octave in each of the wavelet filter banks is referred to as quality factor. The wavelet transform discretizes the scales using the specified number of wavelet filters. prescribed 8 and 1 wavelets per octave in the first and second layers of the scattering network respectively for audio and speech signal processing. These choices of quality factors were adopted in the present work.

The scattering transform network $S(t,n,q)$ was applied separately to each training/test set in the patients' datasets. The network treats each row (i.e. EEG data from a channel) in the training/test set as a single signal. The number of scattering paths and time windows in the representation for each signal depend on the choice of t,n and q . With the values $t=1$ hr, $n=2$, $q=[8,1]$ for sEEG and $t=10$ min, $n=2$, $q=[8,1]$ for iEEG the outputs of the transform are tensors 1034-by-4-by- k and 951-by-5-by- k respectively which are indexed by scattering path, time window and k where k is the number of signals (i.e. rows or EEG channels) in each patient's training/test sets. Each page of a tensor corresponds to the scatter transform of a signal therefore, an sEEG signal results in a feature matrix of dimension 1034-by-4 while an iEEG signal results in a feature matrix of dimension 921-by-5. In order to obtain a matrix compatible with the classification algorithm, each multi-signal scattering transform (i.e. tensor) was reshaped to a matrix where each column corresponds to a scattering path and each row is a scattering time window. This way, feature matrices of dimensions 4 k -by-1034 and 5 k -by-921 were obtained respectively for sEEG and iEEG training/test sets. Note that each signal in sEEG and iEEG training/test sets has 4 and 5 different scattering time window representations, respectively. Therefore, the class label entries in TrainingLabels and TestLabels were modified to

match the number of scattering windows. In Figures 5a and 5b we have shown sample scattering features derived from one minute interictal and preictal EEG signals.

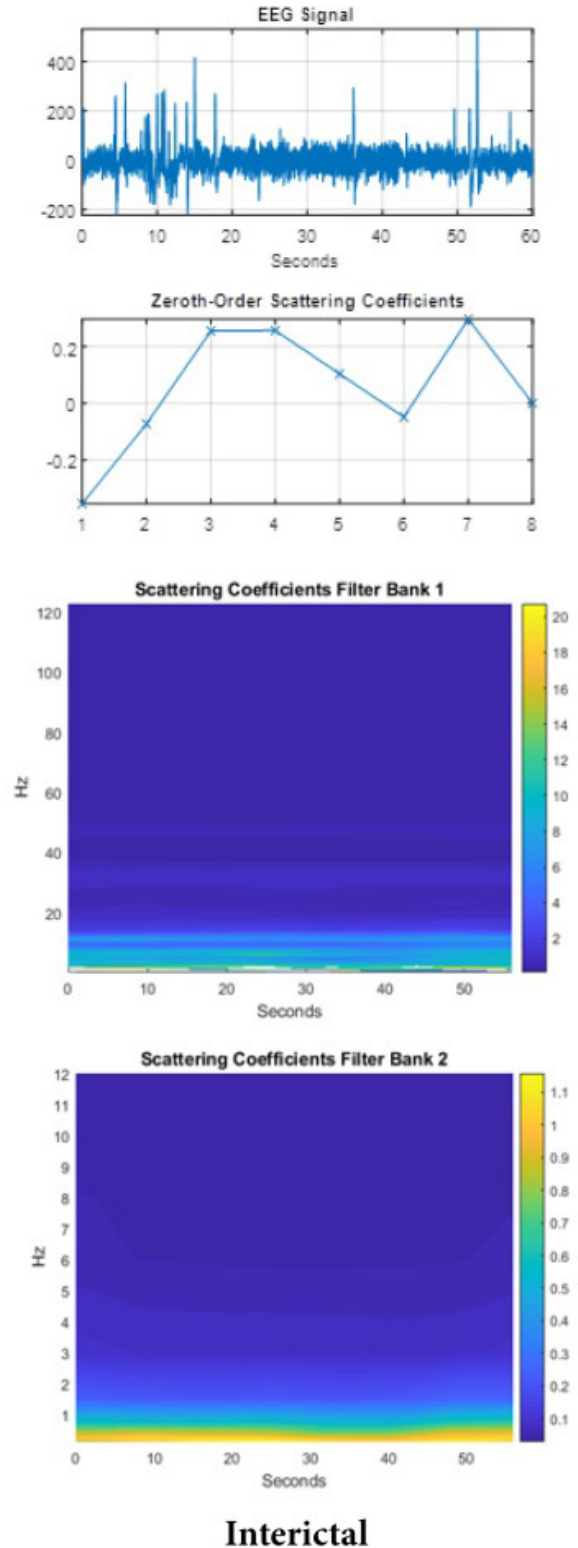
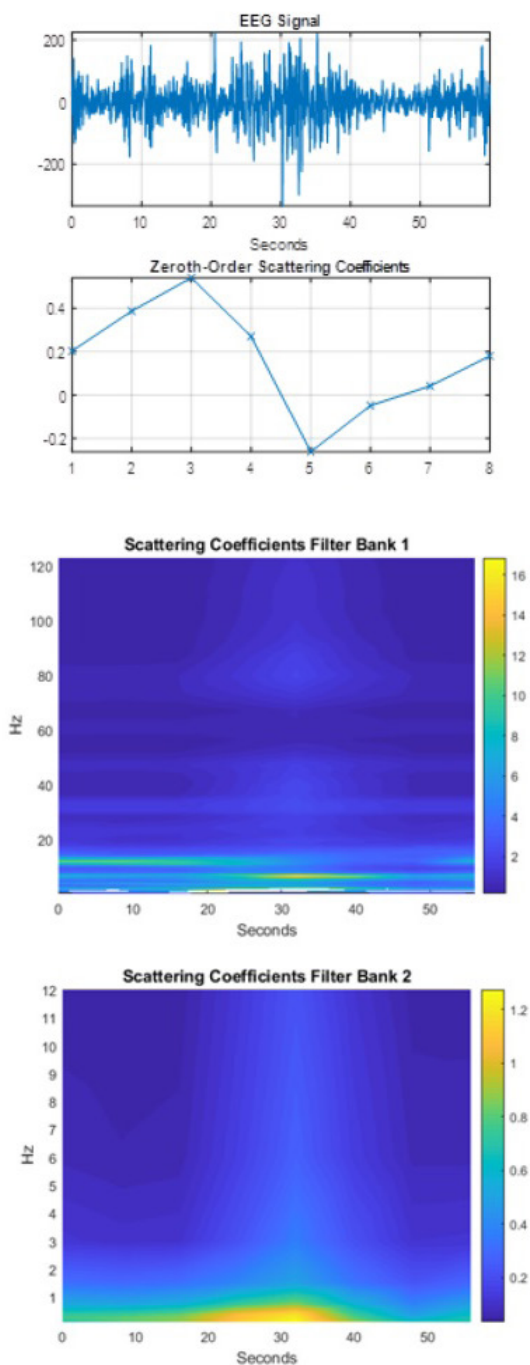


Figure 5(a). Scattering features derived from 1 min interictal sEEG signal of one patient.



Preictal

Figure 5(b). Scattering features derived from 1 min preictal sEEG signal of one patient.

Classifier design

If it is assumed that each signal, $y(t)$ in the training/test sets lies in the Hilbert space H_p where p is the number of samples in the signal, then wavelet time scattering can be viewed as a map, M that transforms $y(t)$ from H_p to $F=[f_i \in R^{1034}, i=1,2,3,4]$ in the case of sEEG and $G=[g_j \in R^{921}, j=1,2,3,4,5]$ for iEEG. The f_s of i and g_s of j are the different scattering time window representations of each signal. The main task here

is to design a method for the classification of interictal and preictal signals, $y(t)$ using their new representations in the F and G domains. Since the dimensionalities of the domains of F and G are quite high, we employed a linear binary support vector machine classification model in order to reduce computational cost. Furthermore, deploying a simple classifier will further show the efficacy or otherwise of wavelet time scattering as a good feature extractor for interictal and preictal EEG classification. Moreover, it has been observed that simple unsupervised feature extraction algorithms, when properly tuned, can generate representations of the data that allow even basic classifiers, such as a linear support vector machine, to achieve state-of-the-art performances [24].

For the classification task, two analyses were performed, the cross validation and held out classification analyses. The latter utilizes both the training and test sets while the former utilizes only the training sets.

Cross validation analysis

The cross validation analysis fits and evaluates a linear binary support vector machine using all the scattering data from each training sets (i.e. 70% of each dataset). There are 4/5 (sEEG/iEEG) scattering sequences for each signal in the entire scattering data. The classification accuracy is estimated in two different ways using 5-fold cross validation. The first is referred to as Ordinary Cross Validation (OCV) analysis. It classifies each scattering window from a signal separately as preictal or interictal EEG. The second approach which is called True Cross Validation (TCV) analysis uses majority vote scheme on the individual scattering windows to make a single classification on all the scattering window representations of a signal as preictal or interictal EEG [25].

Held out analysis (Test analysis)

This fits a linear binary support vector machine only to scattering data obtained from the training sets and then uses that model to make predictions on the scattering data obtained from 30% held out test set. In the same vein, the majority vote scheme was used on the individual scattering windows to make a single classification on all the scattering window representations of a signal as preictal or interictal EEG.

The performance metric used to evaluate the classifier is classification accuracy. It is defined as follows:

$$\text{Accuracy (\%)} = \frac{\text{True Positives} + \text{True negatives}}{\text{True Positives} + \text{False Negatives} + \text{True Negatives} + \text{False Positives}} \times 100$$

where true positives and false positives are the number of samples correctly and incorrectly classified as preictal samples respectively while true negatives and false negatives are the number of samples correctly and incorrectly classified as interictal samples respectively. The cross validation and held out analyses gave rise to three different sets of classification accuracy values namely, Ordinary Cross Validation (OCV), True Cross Validation (TCV) and test (Test) classification accuracy values [26].

Results

Results of the classification experiments detailed in the last section are presented in Table 1. The upper section gives results for patients in the CHB-MIT scalp EEG dataset while the lower section gives results for the two human subjects in the AES intracranial EEG dataset. The reported classification accuracy value for each patient is the average of the classification accuracy values obtained from each dataset created for the patient. The number of datasets for a patient in the CHB-MIT dataset corresponds to the number leading seizure segments contained in the patient's EEG recording. However, each of the two patients in the AES dataset has ten datasets. Classification accuracy values obtained in the three different classification schemes are presented for each patient. The last row of the upper and lower sections of Table 1 gives the averages of the accuracy values for each classification schemes.

Data source	Patient id case		No of patients dataset	Average classification accuracy (%)		
				Ordinary CV	True CV	Test
CHB – MIT sEEG	1	1	2	89.56	95.86	83.12
	2	2	2	91.22	98.45	90.56
	3	3	2	94.16	96.43	95.15
	4	4	3	88.72	92.34	89.44
	5	5	3	93.54	99.01	88.54
	6	6	4	95.23	100	94.34
	7	7	1	92.34	98.88	90.05
	8	9	2	96.23	99.45	91.56
	9	10	3	90.98	98.76	91.22
	10	12	4	94.12	96.66	89.46
	11	13	2	89.45	92.57	88.34
	12	14	2	97.23	100	95.03
	13	15	4	89.35	96.23	87.32
	14	18	2	96.12	100	94.43
	15	20	2	97.43	100	97.19
	16	22	2	90.32	94	91.23
	17	24	3	97.56	100	95.66
	Avg			93.15	97.57	91.33
AES iEEG	1	1	10	96.65	100	95.46
	2	2	10	100	100	98
	Avg			98.33	100.00	96.73

Table 1. Classification Accuracy results.

OCV, TCV and test accuracy results

CHB – MIT dataset: The mean interictal and preictal EEG classification accuracy for the ordinary cross validation experiment is 93.15%. The best OCV accuracy values are obtained for patients 6, 9, 14, 18, 20 and 24 with OCV accuracy values >95% while patients 1, 4, 13 and 15 recorded the least OCV accuracy values with 89.56%, 88.72%, 89.45% and 89.35% respectively. The OCV accuracy values is >90% in nearly all the patients.

The true cross validation classification experiment gave interictal and preictal EEG classification accuracy mean of 97.57%. This is remarkably very high. All the interictal and preictal EEG signals in the datasets of some patients specifically, patients 6, 14, 18, 20 and 24 are correctly classified. Furthermore, the TCV classification accuracy value is >92% for all the patients.

A mean test classification accuracy value of 91.33% was obtained from the test classification experiment. The highest test classification accuracy result of 97.19 was realized in patient 20 while the lowest test accuracy value of 83.12% was obtained in patient 1. Test accuracy values >95% was recorded in approximately 24% of the patients.

AES dataset: The results of the OCV, TCV and Test classification experiment ran on the AES dataset showed that patient 1 recorded OCV, TCV and Test classification accuracy values of 96.65%, 100% and 95.46% respectively while for patient 2 OCV, TCV and Test classification accuracy values of 100%, 100% and 98% respectively were obtained. For the two patients the mean OCV, TCV and Test classification accuracy values are 98.33%, 100% and 96.73%.

Comparison of OCV, TCV and test classification results

Figure 6a is a plot which compares OCV, TCV and Test classification accuracy results obtained for each of the patients in the CHB-MIT sEEG database. It clearly shows that for all patients, results from the TCV classification experiment are consistently higher than the corresponding values obtained from OCV and Test classification experiments. Although the OCV accuracy values are higher than the Test accuracy values in most of the patients, there are occasional overlaps specifically the Test OCV accuracy values are higher in patients 3, 4, 9 and 16. OCV, TCV and Test accuracy values are very close in seven patients namely patients 3, 4, 11,

12, 15, 16 and 17. A box and whiskers plot comparing the spread of the OCV, TCV and Test classification results across all the 17 patients is shown in Figure 6b. The maximum, median and the minimum accuracy values clearly shows that the TCV accuracy results exhibit very low variability. On the other hand, OCV accuracy values show moderate variability while Test accuracy results gave the most variability.

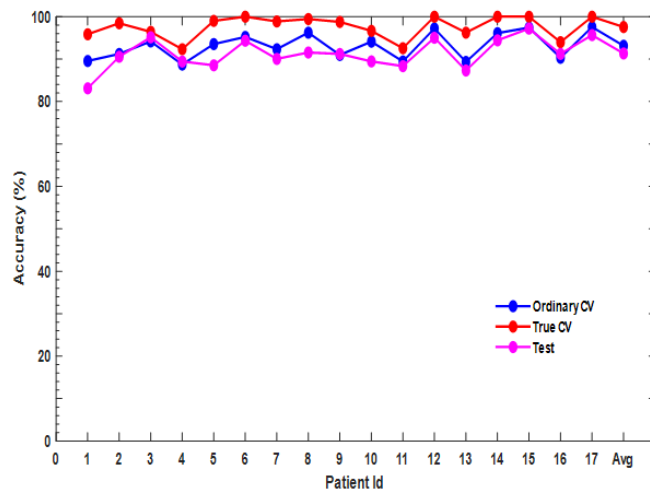


Figure 6(a). Comparison between OCV, TCV and Test classification accuracy values.

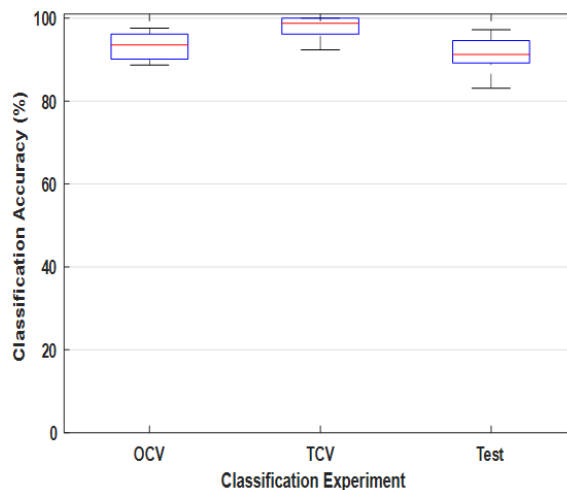


Figure 6(b). Comparison between the spread of OCV, TCV and Test classification accuracy values.

Discussion

Deep CNNs versus scattering transform plus simple classifier for classification task

EEG analysis has been an important tool in neuroscience and neural engineering and many of the analytical tools used in EEG studies have used machine learning to uncover relevant information from neural activities. Specifically, deep convolutional neural networks have recently been employed in several EEG classification task. Deep CNNs can automatically uncover useful discriminatory features for any classification task from raw input data but they have lots of tunable hyperparameters thus requiring huge training dataset and computational resources. On the other hand, group invariant scattering or wavelet time scattering can learn useful data representation from raw input data with few example data from each data class. In this study, sEEG and iEEG features learned through wavelet time scattering were combined with a linear support vector machine in order to classify interictal and preictal EEG signals. In Table 2 we compare the results obtained in this work with performance outcomes of deep CNNs on different EEG classification tasks reported in recent studies. The

classification tasks covered include emotion recognition, motor imagery, mental workload, seizure detection, event related potential detection, and sleep stage scoring. The CNNs' architecture, activation function and data input formulation used in these tasks vary widely.

Reference	Classification task	No. of subjects	Length of EEG data	Performance outcome (accuracy), %
Antoniades et al.	Seizure detection	25	0.68 hr	87.5
Ullah et al.	Seizure detection	5	4097 Samples	99
Antoniades et al.	Seizure detection	18	6 hr	89
Acharya et al.	Seizure detection	10	3.3 hr	88.7
Vilamala et al.	Sleep scoring	19	304 hr	86
Wei et al.	Seizure detection	13	336 hr	90
Abbas & Khan	Motor imagery	9	1.44 hr	61
Tabar & Halici	Motor imagery	9	0.77 hr	75.1
Pereira et al.	Event related Potential	66	25.7 hr	81
Liu et al.	Event related Potential	-	27.7 hr	
Moon et al.	Emotion recognition	32	21.3 hr	
Ang & Guan	Mental workload	120	20 hr	
Jiao et al.	Mental workload	15	1 hr	90
Zouth et al. (2018)	Seizure detection (iEEG)	21	-	96.7
	Seizure detection (sEEG)	23	-	95.6
	Seizure detection (sEEG)	17	43hr	97.6
This work	Seizure detection (iEEG)	2	40hr	100

Table 2. General Comparison between the performance of deep CNNs and scattering features + SVM classifier on EEG classification tasks.

The number of subjects studied and length of EEG data analyzed also vary. Furthermore, the performance metric reported is the highest accuracy achieved in the classification tasks. The TCV classification accuracy values which were the ones reported for the current study are obviously higher than the accuracies obtained from other studies. This observation is quite remarkable considering the fact that a very simple classification algorithm was deployed for the classification task. Moreover, in the dataset preparation stage, data from all EEG channels were taken as interictal/preictal signals (i.e. no channel selection). This might not be true for patients whose seizures are categorized as focal seizures. Focal seizures are usually confined to a particular region of the brain and therefore, only recording electrode attached to the specific region will pick up preictal signals of upcoming seizure events. On the other hand, generalized seizures affect the whole brain region and as such all recording electrodes can pick up preictal signals of upcoming generalized seizures. Since no information concerning the type of seizures were made available in the MIT-CHB and AES datasets we assumed generalized seizures for all patients. Given this assumption, one would expect that our method will exhibit a relatively low performance but the reverse is the case. This shows that wavelet time scattering is not only good at learning useful features but also very robust to little errors.

OCV, TCV and test accuracy values

In order to access the utility of wavelet time scattering as a good feature extractor for interictal and preictal EEG brain state identification, we designed a simple classification algorithm using EEG features derived from wavelet time scattering. Three different classification schemes were used to access the performance of the classifier and hence suitability of wavelet time scattering for the classification task. The TCV classification scheme gave the best accuracy values across all patients and it is closely followed by OCV classification scheme. Although the Test classification scheme gave the least classification accuracy values, these values cannot be described as too low. This observation (i.e. discrepancy in classification accuracies) may be explained in the light of the amount of training data the classifier was exposed to in the classification schemes. In the TCV and OCV classification schemes the classifier trained on every interictal and preictal signal. Since the classifier had a feel of every single training data, it tends to create a decision boundary/surface that supports good generalization. However, the reverse is the case for Test classification scheme. Only a fraction of the interictal and preictal signals were presented for training the classifier resulting in low generalizing ability. For instance, Figures 7a and 7b are scatter plots of the training data (showing the support vectors) in an already trained SVM classifier for a patient in the CHB-MIT dataset. In order to obtain a 2D display, the classifier was trained using scattering coefficient sequences along only two scattering paths. Scattering coefficients in corresponding positions along the paths were plotted against each other for each data point. In Figure 7a the SVM classifier was trained with scattering coefficients from five interictal and five preictal signals while the SVM classifier in Figure 7b was trained with twice as much signals. The support vectors are placed in the black circles. It is observed that the classifier in Figure 7b having learned from more data examples has identified more support vectors. Support vectors are data points that lie on or cross the boundary between the two data classes (i.e. interictal and preictal) therefore, they dictate the shape of the decision boundary learned by the classifier. They are also in some manner instrumental to how well the classifier generalizes to data points outside those used to train the classifier. This observation also shows that, although wavelet time scattering can learn useful data representation for classification task from few examples from each data class, having a fairly large example data from the data classes may improve the quality of features learned using wavelet time scattering.

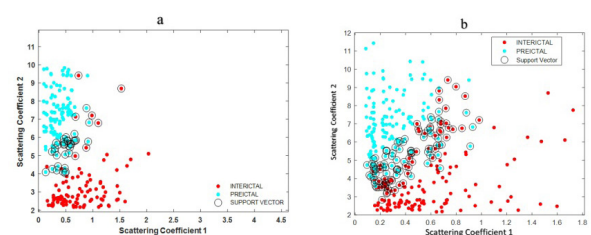


Figure 7. Scatter plots of the training data (showing the support vectors) of an SVM classifier trained with (a) five (b) ten interictal and preictal EEG signals.

sEEG vs. iEEG

Two different forms of EEG data, scalp EEG and intracranial EEG data, were used in this study. Although data from only two patients were available for analysis in the AES iEEG dataset, the average classification accuracy results obtained are higher for iEEG data than the sEEG data. Two factors may be responsible for this observation. Firstly, it may be explained in terms of the advantages of iEEG recordings over sEEG recordings. iEEG recordings has high signal to noise ratio. Furthermore, it is a localized recording of the brain activity which minimizes unwanted interferences from other brain sites on the signals recorded from the region of interest. However, it has been argued that although scalp EEG recording

cannot provide localized neuronal potential activities, it can present a more generalized spatiotemporal view of brain's dynamical system. The classification accuracies for the iEEG data may also be higher if the seizure events in the two patients are generalized. This means that preictal changes in the recordings are present in all the recording channels making it easy for the wavelet time scattering network to learn more discriminatory features for the classification tasks.

Implication of classification results for seizure prediction algorithms

Algorithms that aim at the identification of preictal brain state sufficiently long before electrographic seizure onset are referred to as seizure prediction algorithms. The classifier is the component of a seizure prediction algorithm that decides which EEG signal epochs are preictal and then triggers other post processing operations. This makes interictal and preictal EEG classification system the backbone of any seizure prediction algorithm. Efforts at accurately classifying interictal and preictal EEG signals had seen many researchers using domain knowledge to extract EEG features in the hope that these hand-crafted EEG features would serve as consistent preictal brain state markers. However, when it appeared that no single feature or group of features could serve as a consistent preictal brain state marker, the idea of building classification systems that can automatically learn useful discriminatory features from raw input data became popular. Various deep learning architectures such as Convolutional Neural Networks, Deep Belief Networks and Recurrent Neural Networks that are capable of feature learning have been used in many seizure prediction algorithms. Although these seizure prediction algorithms displayed very promising prediction performances, they are however highly computationally expensive and depend upon very large training datasets thereby requiring high power consumption for their operations. But, if seizure prediction algorithms eventually proved to be successful, they would most likely be implemented in a portable and implantable device which must operate on low power in order to drastically reduce routine maintenance and remain convenient for the patients and care givers.

Interestingly, wavelet time scattering is an automatic feature learner with relatively simple configuration and architecture which is capable of revolutionizing the manner in which the herculean task of actualizing a robust and consistent seizure prediction algorithm is being pursued. We have demonstrated in this study that the combination of wavelet time scattering based EEG features and very simple machine learning algorithms can produce excellent interictal and preictal EEG classification accuracy results for epileptic patients. Hopefully, leveraging the highlighted advantages of wavelet time scattering could pave the way for very powerful and consistent seizure prediction algorithms.

Conclusion

If implemented within a closed-loop intervention system equipped with efficient seizure-aborting strategies, seizure prediction algorithms may prove useful as an alternative therapeutic strategy for epilepsy. On the other hand, an interictal and preictal EEG classification system that is highly sensitive and specific is essential for a successful seizure prediction algorithm. The efficacy of a relatively new and simple feature learner – wavelet time scattering for interictal and preictal EEG classification was studied. Features learned through wavelet time scattering from raw sEEG and iEEG data were used to train and evaluate a linear support vector machine classifier. The three different classification experimental schemes carried out resulted in very high classification accuracy values despite using a very crude and simple linear classifier. Therefore, research efforts that may produce even better results include extensive parameter sensitivity analysis of the wavelet time scattering network and investigation of other classification algorithms which may be optimized for each individual patient.

Finally, the results obtained here when properly harnessed may have significantly positive impact on the realization of the long awaited clinically deployable seizure prediction algorithm. Therefore, an important

future direction of this study is the development of a seizure prediction algorithm which leverages our interictal and preictal EEG classification efficiency.

Acknowledgements

We acknowledge with gratitude the efforts of individuals at the Massachusetts Institute of Technology (MIT) and Children Hospital Boston (CHB) who made the sEEG data used in this work available. A team of investigators from the Children Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT) created and contributed the database to PhysioNet (The PhysioNet web site is a public service of the PhysioNet Research Resource for Complex Physiologic Signals). The clinical investigators from CHB include Jack Connolly, REEGT; Herman Edwards, REEGT; Blaise Bourgeois, MD. The investigators from MIT include Ali Shoeb, PhD and Professor John Guttag. We equally acknowledge with thanks the efforts of every individual at the American Epilepsy Society (AES) and Kaggle who made the iEEG data used in this work accessible.

References

1. PD, Adelson, Nemoto E, Scheuer M and Painter M et al. "Noninvasive Continuous Monitoring of Cerebral Oxygenation Preictally Using Near-infrared Spectroscopy: A Preliminary Report." *Epilepsia* 40 (1999): 1484–1489.
2. Zubair Ahmad, Muhammad, Awais Mehmood Kamboh, Sajid Saleem and Amir Ali Khan. "Mallat's Scattering Transform Based Anomaly Sensing for Detection of Seizures in Scalp EEG." *IEEE Access* 5(2017): 16919–16929.
3. Alotaiby, Turki N, Saleh A Alshebeili, Faisal M Alotaibi and Saud R Alrshoud. "Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals." *Comput Intell Neurosci* (2017): 231–240.
4. Andén, Joakim and Stéphane Mallat. "Deep Scattering Spectrum" *IEEE Trans Signal Process* 62(2014): 4114–4128.
5. Balestrieri, Randall and Behnaam Aazhang. "Robust Unsupervised Transient Detection with Invariant Representation Based on the Scattering Network." *ArXiv Preprint* (2016).
6. Bandarabadi, Mojtaba, César A Teixeira, Jalil Rasekhi, and António Dourado. "Epileptic Seizure Prediction Using Relative Spectral Power Features." *Neurophysiol Clin* 126(2015): 237–48.
7. CW, Baumgartner and Leutmezer F "Preictal SPECT in Temporal Lobe Epilepsy: Regional Cerebral Blood Flow is Increased Prior to Electroencephalography-Seizure Onset". *J Nuclear Medicine* 39 (1998): 978–982.
8. Bruna, Joan and Stéphane Mallat. "Invariant Scattering Convolutional Networks" *IEEE Trans Pattern Anal Mach Intell* 35(2013): 1872–1886.
9. Charles K. Chui. "An introduction to wavelets" *Academic Press* (1992).
10. Coates, Adam, Honglak Lee and Andrew Ng. "An Analysis of Single-layer Networks in Unsupervised Feature Learning." *AISTATS* 14(2011).
11. L. Cohen. "Time – Frequency Analysis, Prentice-Hall New Jersey 1995." *Control Engineering Practice* 5(1995): 292–294.
12. Delorme, Arnaud and Scott Makeig. "EEGLAB: An Open Source Toolbox for Analysis of Single-trial EEG dynamics" *J Neuroscience Methods* 134(2014): 9–21.
13. Kiral-Kornek, Isabell, Subhrajit Roy, Ewan Nurse and Benjamin Mashford et al. "Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System." *EBioMedicine* 27(2018): 103–111.
14. Kaggle. "American Epilepsy Society Seizure Prediction Challenge." (2014).
15. Levin, Kuhlmann, Grayden David B, Cook Mark J and Burkitt Anthony N et al. "Proceedings of the 7th International Workshop on Seizure Prediction." (2015): 1.
16. Mallat, Stéphane. "Group Invariant Scattering." *Communications in Pure and Applied Mathematics* 65(2012): 1331–1398.
17. Mallat, Stéphane. "Understanding Deep Convolutional Networks." *Philos Trans A Math Phys Eng Sci* 374(2016): 1–16.

18. Myers, Mark H, Akshay Padmanabha, Gahangir Hossain and Amy L de Jongh Curry et al. "Seizure Prediction and Detection Via Phase and Amplitude Lock Values." *Front Hum Neurosci* (2016).
19. Wang, Ning, and Michael R Lyu. "Extracting and Selecting Distinctive EEG Features For Efficient Epileptic Seizure Prediction." *IEEE J Biomed Health Inform* 19(2014): 123–128.
20. N, Senanayake and Roman GC. "Epidemiology of Epilepsy in Developing Countries." *Bull World Health Organ* 71(1993): 247-258.
21. Keller, Simon S, G Russell Glenn, Bernd Weber and Barbara AK Kreilkamp et al. "Preoperative Automated Fibre Quantification Predicts Postoperative Seizure Outcome in Temporal Lobe Epilepsy." *Brain* 28(2016): 203 – 210.
22. Talmon, Ronen, Stéphane Mallat, Hitten Zaveri and Ronald R Coifman "Manifold Learning For Latent Variable Inference in Dynamical Systems." *IEEE Trans Signal Process* 63(2015): 3843-3856.
23. "Causes of Epilepsy." *The University of Chicago Medical Centre* (2018).
24. S S, Viglione and G O Walsh "Epileptic Seizure Prediction." *Electroencephalogr Clin Neurophysiol* 39(1975): 435-436.
25. S, Viglione, Ordon V and Risch F. "A Methodology for Detecting Ongoing Changes in the EEG Prior to Clinical Seizures." (1970).
26. "Epilepsy Fact Sheets." *World Health Organization* (2019).

How to cite this article: Afreda Susu A, Agboola HA, Solebo C and Lessi F EA et al. "Wavelet Time Scattering Based Classification of Interictal and Preictal EEG Signals." *J Brain Res* 3 (2020): 115.