Abstract

Epilepsy is one of the most common and yet diverse set of chronic neurological disorders. This excessive or synchronous neuronal activity is termed “seizure”. Electroencephalogram signal processing plays a significant role in detection and prediction of epileptic seizures. In this paper we introduce an approach that relies upon the properties of wavelets for seizure detection. We utilise the Maximum Overlap Discrete Wavelet Transform which enables us to reduce signal noise. Then from the variance exhibited in wavelet coefficients we develop connectivity and communication efficiency between the electrodes as these properties differ significantly during a seizure period in comparison to a non-seizure period. We use basic statistical parameters derived from the reconstructed noise reduced signal, electrode connectivity and the efficiency of information transfer to build the attribute space. We have utilised data that are publicly available to test our method that is found to be significantly better than some existing approaches.

Keywords: EEG, Graph Theory, MODWT, Wavelet variance, Decision trees

1. Introduction

Epilepsy is described by recurring seizures caused by abnormal activity in the brain. This maybe considered a chronic disorder of the central nervous system that exposes individuals to experiencing recurrent seizures. A seizure is a transient irregularity in the brain’s electrical processes that may also produce disruptive physical symptoms (Zandi et al., 2010). Electroencephalogram (EEG) is a signal which represents the electrical activity of neurons within the brain. The signal is acquired from the surface of the scalp and these signals are non-stationary, that is its statistical properties vary over time. The important frequencies from the physiological viewpoint lie in the range of 0.1 to 30 Hz. (Shrestha et al., 2019). A common classification problem is seizure detection, where non-seizure and seizure EEG records of patients need to be identified.

It is not an onerous task to find a number of papers and differing methods that seek to determine the onset of seizure. The nearest-neighbour classifier was used on EEG features extracted in both time and frequency domains to detect the onset of epileptic seizures (Qu and Gotman, 1997). Also a method using various derived statistical parameters as extracted features has been proposed (Siddiqui et al., 2018). One problem with these methods is that
all the information in the signal is being considered. Some of that information may simply be noise and may also contain specific frequency bands that are not related to the event that is to be classified. In another study the application of a method based on energy extraction at various frequency sub-bands using the wavelet transform was implemented (Elsa Jacob et al., 2018). This avoids processing redundant data by selecting the required sub-bands / wavelet decomposition levels which contain relevant information. Statistical parameters of the energy at each level were derived and used as attributes for a Support Vector Machine (SVM) classifier. Using this method with statistical parameters of energy, at different frequency levels, and in this case with 21 electrodes (arranged in the international “10-20” format, as an example see Figure 1.), the amount of computation required increases quickly if we require additional frequency bands to better segregate the signal or additional statistical parameters to fully describe the energy feature. This results from the need to compute the statistical parameters for the energy within each additional sub-band/wavelet decomposition level required and of course for the signal from each electrode.

Figure 1: 10 - 20 Electrode positioning; developed from Helwig (2018).

The EEG signal contains several spectral components. Wavelet analysis allows frequency separation of such components (Percival and Walden, 2000). Use of Wavelets in EEG analysis is quite often seen and is well described, with the Discrete Wavelet transform (DWT) frequently encountered (Faust et al., 2015). (For brief overview of DWT see Appendix A.3.1). The wavelet transform is effective for representing different aspects of non-stationary signals such as trends, discontinuities, and where signal patterns are recurring, here other signal processing approaches are unlikely to be as effective (Daubechies, 1990). Some limitations of the DWT are the requirement of the signal to be of length \(2^J\) where \(J \in \mathbb{Z}^+\) and is the upper level of wavelet decomposition. Also as \(J\) increases the
number of wavelet coefficients within each additional level systematically reduces, so to derive meaningful statistical parameters at these upper levels of decomposition might not always be possible.

Our proposed method minimises the use of non-required frequency ranges within the signal, reconstructs the signal to its original length and overcomes some of the possible shortfalls of the DWT. It also uses the interaction between subsets of electrodes during an epoch. This interaction or connectivity can be seen to vary significantly between the non-seizure and seizure periods. We utilise statistical parameters from our reconstructed signal and electrode connectivity as extracted features for classification. The organisation of this paper is as follows; Section 2 outlines some related works to our proposed method, Section 3 presents our method, Section 4 outlines experimental results together with compared works. Concluding remarks are presented in Section 5, followed by an Appendix which includes a Glossary.

2. Related Work

Time and frequency domains are commonly used in feature extraction methods for EEG signals. To analyse the frequency domain features, the Discrete Fourier Transform (DFT) has been applied, followed by a form of power spectral density analysis of the EEG signals (Lee et al., 2014). Decomposing a signal in terms of its frequency content, using the DFT which relies on sinusoids, results in fine resolution within the frequency domain. However the Fourier series representation is not very effective at time resolution (Vikash, 2015). Another method of decomposing signals is the DWT, where in a wavelet representation, we represent the signal in terms of functions that are localised both in time and frequency. Here we have an adaptive time frequency window, which provides good time resolution at high frequencies, and good frequency resolution at low frequencies (Polikar, 1994).

Siddiqui et al. (2018) generated nine statistical parameters from EEG data to extract time domain features and reduce dimension. The duration of data (Epoch) was also altered, starting at 10 sec, and reducing in steps to 0.025 secs. The classifiers chosen there were decision forests, an ensemble of decision trees. A concern with this methodology is that one is using all the available information or frequency bandwidth within the signal. Such techniques as DFT or DWT were not used to minimise the inclusion of some sections of the signal’s bandwidth, it is quite possible that additional noise\(^1\) may have been introduced. Hence the likelihood of introducing error into the classification of the EEG signal.

EEG signals have also been decomposed into time–frequency representations using the DWT (Parvez and Paul, 2013; Faust et al., 2015). This enables the EEG signals to be decomposed into several sub-bands, this decomposition is called Multi Resolution Analysis (MRA). By Parseval’s Theorem, the percentage distribution of energy features in the EEG signal may be extracted at each of these different sub-bands (Omerhodzic et al., 2013). The features extracted from the wavelet coefficients at various levels or different frequency bands can be used to determine the characteristics of the signal. Use of the DWT continues to exhibit improving results on seizure detection (Faust et al., 2015). One concern here is that with the DWT, as we increase the levels of decomposition to extract the relevant frequency bandwidths we have fewer wavelet detailed coefficients \(\{d_{i,j}\}\), within these higher

\(^1\) Here we define noise as frequency ranges unrelated to a seizure event.
levels of decomposition\(^2\). That is \(\{d_{1,j}, d_{2,j}, \ldots, d_{k,j}\} \in \mathcal{D}_j\) has fewer elements as \(j\) increases: \(0 \leq j \leq J\). Hence the difficulty in deriving meaningful statistical parameters at the higher decomposition levels.

Chen et al. (2017) uses various different DWTs for seizure detection, also choosing from nine statistical features derived from each of the wavelet generated frequency bands then classifying (via SVM), the multi-channel EEG recordings from Goldberger et al. (2000). The empirical evidence shown there outlines that the choice of mother wavelet does not have a significant effect on seizure detection, however it is very sensitive to decomposition level if the features of non-seizure/seizure EEGs exhibit significant difference in the different frequency bands. In this particular study each Epoch of EEG data used was 20 seconds in duration. Of concern here is that many frequency bands and the resulting DWT coefficient features could be redundant, causing inaccuracy and unnecessary high computational cost. If we decompose to 6 levels: \(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_6 \) and \(\mathcal{S}_6\) and use all the wavelet coefficients within each of these levels to derive statistical parameters, the result set of attributes to be used for classification becomes quite large for multivariate EEG data.

While use of the DWT within EEG analysis is often encountered, however use of the variance within the wavelet coefficients resulting from the transform, mainly consider this wavelet variance as a selected feature only. This is usually the standard deviation of the detail wavelet coefficients at each decomposition scale or a linear combination of the variance across the different wavelet decomposition scales (Janjarasjitt, 2010; Ashok and Mahalakshmi, 2017). However little if any consideration is given to the covariance between different signals/nodes, which may be calculated similarly as the wavelet variance, this is called wavelet covariance, see definition 5 Appendix A.2, as well as Equations 9, 10 and 11. The covariance (and wavelet correlation, see Equation 12) can be used to derive electrode/node connectivity during an epoch to provide additional features for classification, especially if the connectivity between electrodes/nodes alters significantly between the non-seizure and seizure states. As an example this is shown in Figure 2, highlighting an epoch for both non-seizure and seizure states, displaying an electrode’s number of connections to other electrodes. It can be seen that during the non-seizure state, there exists numerous connections between the electrodes, i.e. 6 electrodes have connections ranging from 5 to 10 other electrodes and another 6 electrodes have connections ranging from 15 to 20 other electrodes. Where as in comparison to non-seizure here no electrode has more than eight connections to other electrodes and only one electrode has a maximum of eight connections at the chosen wavelet correlation threshold.

The discrimination of EEG events, using statistical parameters derived from multivariate data as attributes are likely to include considerable information unrelated to the event under study, noticed as noise and may induce error. However using the DWT in an attempt to reduce noise by selecting the required frequency bands (wavelet decomposition levels), problems may still arise. One obvious issue is length of signal, if not a multiple of \(2^J\) where \(J \in \mathbb{Z}^+\), then additional components would need to be added in as extra coefficients or they may be simply omitted. Either option may introduce bias, hence possibly confuse classification.

\(^2\) Here \(d_{i,j}\) represents the \(i^{th}\) detail wavelet coefficient at the \(j^{th}\) scale or level and \(\mathcal{D}_j\) represents all the detail coefficients at level \(j\). Similarly for the smooth wavelet coefficients, \(s_{i,j}\) and \(\mathcal{S}_j\).
3. Our Proposed Technique

As we are interested in the classification of neurological events, labelled as either non-seizure or seizure states, where the data is presented in a multivariate time series format. Here we present a novel approach for feature extraction and classification using the inherent capabilities of the Maximum Overlap discrete Wavelet Transform (MODWT, Appendix A.3.2) to:

♦ decompose and then reconstruct the EEG signal minimising the noise components
♦ construct a record of inter-electrode connectivity and develop a metric for communication efficiency of electrodes \{ from wavelet covariance. \}

From these wavelet derived methods, we build our attribute space and apply an ensemble classifier to the transformed data.

The scenario we are considering is to have a database consisting of labelled EEG signals from a number of previous or existing patients. We subset the signal into smaller periods of time (Epochs) and label each epoch (non-seizure or seizure), determined by where the epoch occurs in time, within the signal. We construct a model from this training data, and apply the model on incoming unlabelled EEG data, an epoch at a time, from a patient to classify/detect non-seizure and seizure events.

When using raw data and deriving statistical parameters as attributes, then our method if using same number of derived statistical parameters, we notice slightly more attributes per record. However, we reduce the number of attributes when compared to standard DWT decomposition methodology.
3.1 Main Steps

We outline the main steps of our method in this subsection, then explain each of these steps in greater detail in the following subsections. With the raw data chosen then after subsetting the data for each patient into the required time duration or epoch, we begin a three step process to attain the required attributes, then combine all into a data set followed by the 4th step, Classification.

Step 1. Preparation of Attributes from statistical parameters of reconstructed signal.

Step 2. Preparation of Attributes from Connectivity of Electrodes

Step 3. Preparation of Attributes from Global Efficiency of Electrodes

Step 4. Assigning Class labels and Training Classifiers

3.2 Step 1: Preparation of Attributes from statistical parameters of reconstructed signal.

With data, originating from an EEG skull cap sampling method, see Figure 1 for an example of the standard EEG “10-20” skull cap electrode configuration. The data appears as a time series or signal from each electrode. The non-seizure and seizure periods are previously identified. We segment the signal into epochs of suitable duration and attach class labels. Any epoch with part of its duration within the identified seizure period is defined as seizure, elsewhere non-seizure. As an example see Figure 3. Each segment along the X-Axis within a left and right arrow represents a epoch. The segment on the left hand end, entirely falls during a non seizure event and hence is labelled as non-seizure. The 2nd and 3rd epochs from the left, either partially or fully occur during a seizure event and hence are labelled as seizure. We apply to each epoch the wavelet transform.

The decomposition function of the wavelet transform may be represented as a tree of low and high pass filters (LPF & HPF), with each step further decomposing the low pass filter, Figure 4 displays such decomposition to 3 levels. In this example the resultant transform would consist of the levels \{H1, H2, H3 & L3\} and the usual nomenclature of the MODWT for these levels is \{\widetilde{D}_1, \widetilde{D}_2, \widetilde{D}_3 & \widetilde{S}_3\} respectively. Each of these MODWT levels consists of wavelet coefficients, with \{d_{i,j}\} ∈ \widetilde{D}_j as detail coefficients and \{s_{i,J}\} ∈ \widetilde{S}_J for smooth coefficients. Figure 5 shows the wavelet coefficients resulting from a 3 level MODWT, at each of the detail and smooth levels, across time for an epoch.

EEG waves are conventionally grouped into frequency ranges and named: delta, theta, alpha, beta and gamma waves, where each set of these waves occupies a different non overlapping frequency bandwidth. Then dependent upon the sampling frequency of the initial data, we select our level of wavelet decomposition to align as close as possible with the frequency bandwidths required, to analyse the EEG signal (Elsa Jacob et al., 2018). However, some frequency bands are unlikely to be related to the EEG event we seek to classify, as it has been noted that EEG signals do not have any useful frequency components above 30 Hz (Subasi, 2007). Therefore after wavelet decomposition, we omit or smooth the

3. Similar to Section 2 for the DWT here \widetilde{D}_j and \widetilde{S}_J represent the MODWT levels j or J : 0 ≤ j ≤ J
non-related frequency bands and reconstruct our signal using the inverse wavelet transform\(^4\). Figure 6 shows an original signal compared to the reconstructed signal where MODWT level \(D_1\), has been omitted prior to applying the inverse wavelet transform.

For each electrode (say, 1 to \(n\)), after decomposition and reduction of unwanted noise, the inverse transform produces a reconstructed signal of same length as the original data. We then take statistical parameters being: minimum, maximum, mean, standard deviation and normalised energy\(^5\) from each electrode’s reconstructed signal and form the attributes. We have for each epoch; say epoch \(i\) \(min_{i1}, max_{i1}, mean_{i1}, std_{i1}, normE_{i1}, min_{i2}, max_{i2}, mean_{i2}, \ldots, mean_{in}, std_{in}, NormE_{in}\). We use the statistical parameters to form a list of attributes representing a single epoch. Table 1 represents the data set produced in this step. The following steps add additional attributes to the data set as explained later. A data set is generated from the EEG signal of each patient in order to create patient specific training data set. One may also create a training data set combining all patients. Algorithm 1 in Section 4 outlines this Step 1 in a structured procedure format.

![Figure 3: Data from a single electrode, with seizure duration defined and Epochs labelled.](image)

### 3.3 Step 2: Preparation of Attributes from Connectivity of Electrodes

Using the MODWT to decompose the time series from each electrode per epoch, we derive wavelet variance, and similarly wavelet covariance and hence wavelet correlation between each of the transformed signals from each electrode. Wavelet variance is calculated at each level of wavelet decomposition and is equal to the variance of the wavelet coefficients at that level.

---

4. The inverse wavelet transform is fully documented (Percival and Walden 2000 Chapter 5).
5. Normalised energy \((NormE)\) here is the sum of the squares of the time series values divided by the number of elements in the time series, i.e. \(\text{Norm.Energy} = \frac{\sum_i X_i^2}{\text{number of elements}}\).
level. The sum of the wavelet variances from all decomposition levels equals the sample variance of the time series, similarly for wavelet covariance between each of the transformed signals. Covariance provides a measure of joint variation for two random sequences. The covariance of a random sequence with itself would be the variance. see Appendix A.3.2, A.3.3 and (Percival and Walden, 2000, Chapter 8).
Wavelet correlation which is derived from wavelet covariance see Equation 12, provides us with a metric allowing us to determine association (joint variability) between the transformed signals from each electrode to another, during an epoch. By setting a user defined lower bound on the correlation value between two electrodes, we may define if a connection (or association) exists between these electrodes. If the correlation value between two transformed signals is greater than or equal to the user defined lower bound then we conclude that a connection exists between the two during the epoch.

We then represent those connections in an adjacency matrix where a connection is represented by the value 1. See Table 2 as an example of an adjacency matrix. Here, Electrode 1 (Elec-1) and Electrode 2 (Elec-2) are connected since the correlation between the signals, as calculated by Equation 12 is higher than the user defined lower bound. From the derived adjacency matrix we take the sum of the rows. In Table 2, Electrode 1 is directly connected to two other electrodes i.e Elec-2 & Elec-3, therefore sum of the row is 2. That is the degree of connection is equal to 2. Electrode 3 is directly connected to five other electrodes, therefore the sum of row is 5.

6. The value of this lower bound may be derived by the number of connections available determined by that specific value, given that we need to derive statistical parameters from these connection values.
Hence the degree of connection is equal to 5.

For each electrode in our network we now have a degree of connection or number of direct connections during the epoch. Figure 2 is a representative example of this. From these row sums (or degree of connection) we derive the statistical parameters: mean, maximum, 1st Quartile, 3rd Quartile and skewness. We represent these as: meanC, maxC, ..., skewC. This provides us with additional attributes for the epoch in question. These additional attributes for each epoch are combined with the data set outlined in Step 1 (as shown in Table 1) increasing its dimension. The updated data set is shown in Table 3. A procedural format of this Step 2 is shown in Algorithm 1.

Table 2: Example: Adjacency matrix

|      | Elec-1 | Elec-2 | Elec-3 | Elec-4 | Elec-5 | Elec-6 | Elec-7 | ... | Elec-n | Sum |
|------|--------|--------|--------|--------|--------|--------|--------|-----|--------|-----|
| Elec-1 | 0      | 1      | 1      | 0      | 0      | 0      | 0      | ... | 0      | 2   |
| Elec-2 | 1      | 0      | 1      | 0      | 0      | 1      | 1      | 0   | 0      | 4   |
| Elec-3 | 1      | 1      | 0      | 1      | 1      | 1      | 0      | 0   | 0      | 5   |
| Elec-4 | 0      | 0      | 1      | 0      | 0      | 0      | 0      | 0   | 0      | 1   |
| Elec-5 | 0      | 0      | 1      | 0      | 0      | 0      | 0      | ... | 0      | 1   |
| Elec-6 | 0      | 1      | 1      | 0      | 0      | 0      | 1      | 0   | 0      | 3   |
| Elec-7 | 0      | 1      | 0      | 0      | 0      | 1      | 0      | 0   | 0      | 2   |
| ...   | ...    | ...    | ...    | ...    | ...    | ...    | ...    |     | ...    |     |
| Elec-n | 0      | 0      | 0      | 0      | 0      | 0      | 0      | ... | 0      | 0   |

Table 3: Example: Data set resulting after Step 2

|     | Elec-1 |     |     |     |     |     | ... |     |     |     |     |     |     |     |     |
|-----|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     | min_{11}, max_{11}, ..., normE_{11} | ... |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Epoch 1 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Epoch 2 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| ... |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Epoch n |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

3.4 Step 3. Preparation of Attributes from Global Efficiency of Electrodes.

Section 3.3 shows that from the wavelet coefficients we have a methodology for the construction of an adjacency matrix derived from the correlation between the electrodes. From that adjacency matrix we are able to construct a visual representation of the network or Graph. Figure 7 shows a sub-graph derived from only the first seven electrodes of Table 2. In Figure 7 an electrode is represented as a numbered node and a pair of electrodes/nodes having a correlation above the threshold are connected through an edge between them. Figure 7 has seven nodes/vertices (i.e. \(N = 7\)) and nine edges. Each single edge has an assumed unit length of 1. From this constructed network diagram (and hence the adjacency...
matrix) we derive how “efficiently”\textsuperscript{7} the electrodes are communicating across their network of connections. We calculate the global efficiency ($E_{glob}$) for each node.

For every electrode $i$ we compute the reciprocal of shortest path length ($R_{i,j}$) to each other electrode $j$. For example the shortest path length from electrode 5 to Electrode 7 is three, hence $R_{5,7}$ is $\frac{1}{3}$. This is shown in the corresponding cell (encircled) in Table 4. For each electrode $i$ we then then compute $\sum_{j=1}^{N} \frac{R_{i,j}}{N-1}$ where $N$ is the number of nodes and $R_{i,i} = 0$. The last row of Table 4 shows these values. We call these values global efficiency ($E_{glob}$) values of an electrode $i$ for an epoch, see definition 7.

![Figure 7: A Visual representation of Table 2](image)

Table 4: Example: **Global Efficiency**

|        | Elec1 | Elec2 | Elec3 | Elec4 | Elec5 | Elec6 | Elec7 |
|--------|-------|-------|-------|-------|-------|-------|-------|
| Elec1  | 0     | 1     | 1     | 1/2   | 1/2   | 1/2   | 1/2   |
| Elec2  | 1     | 0     | 1     | 1/2   | 1/2   | 1     | 1     |
| Elec3  | 1     | 1     | 0     | 1     | 1     | 1     | 1/2   |
| Elec4  | 1/2   | 1/2   | 1     | 0     | 1/2   | 1/2   | 1/3   |
| Elec5  | 1/2   | 1/2   | 1     | 1/2   | 0     | 1/2   | 1/3   |
| Elec6  | 1/2   | 1     | 1     | 1/2   | 1/2   | 0     | 1     |
| Elec7  | 1/2   | 1     | 1/2   | 1/3   | 1/3   | 1     | 0     |
| \textit{Sum}/(N-1) | 2/3 | 5/6 | 11/12 | 5/9 | 5/9 | 3/4 | 11/18 |

\textsuperscript{7} Efficiency of node $i$ is related to the sum of the inverses of the minimum distances between node $i$ and all other nodes in the network.
From these global efficiency values for an epoch as shown in the last row of Table 4 we derive statistical parameters: mean, maximum, 1st Quartile, 3rd Quartile & skewness. We represent them as Mean\(_G\), Max\(_G\), ...Skew\(_G\) for the \(i\)th epoch. These attributes are again added to the data set as shown in Table 5. Again this step in procedural form is seen in Algorithm 1.

### Table 5: Example: Data set resulting after Step 3.

| Individual | Step 1 | Step 2 | Step 3 | Global Efficiency |
|------------|--------|--------|--------|-------------------|
|             | Elec-1 to \(n\) | Connectivity |          |                   |
| Epoch 1     | \(min_{11}, max_{11}, \ldots, normE_{1n}\) | \(meanC_1, maxC_1, \ldots, skewC_1\) | \(MeanG_1, MaxG_1, \ldots, SkewG_1\) |
| Epoch 2     | \(min_{21}, max_{21}, \ldots, normE_{2n}\) | \(meanC_2, maxC_2, \ldots, skewC_2\) | \(MeanG_2, MaxG_2, \ldots, SkewG_2\) |
| ...        |        |        |        |                   |
| Epoch \(n\) | \(min_{m1}, max_{m1}, \ldots, normE_{mn}\) | \(meanC_m, maxC_m, \ldots, skewC_m\) | \(MeanG_m, MaxG_m, \ldots, SkewG_m\) |

### Table 6: Example: Data set resulting from the 3 Steps and attaching label

| Individual | Step 1 | Step 2 | Step 3 | Label          |
|------------|--------|--------|--------|----------------|
|             |        | Connectiv |       |                 |
|             |        |        |        |                 |
| Epoch 1     | \(min_{11}, max_{11}, \ldots, normE_{1n}\) | \(meanC_1, maxC_1, \ldots, skewC_1\) | \(MeanG_1, MaxG_1, \ldots, SkewG_1\) | non-seizure     |
| Epoch 2     | \(min_{21}, max_{21}, \ldots, normE_{2n}\) | \(meanC_2, maxC_2, \ldots, skewC_2\) | \(MeanG_2, MaxG_2, \ldots, SkewG_2\) | seizure         |
| ...        |        |        |        |                 |
| Epoch \(n\) | \(min_{m1}, max_{m1}, \ldots, normE_{mn}\) | \(meanC_m, maxC_m, \ldots, skewC_m\) | \(MeanG_m, MaxG_m, \ldots, SkewG_m\) | non-seizure     |

### 3.5 Step 4. Assigning Class labels and Training Classifiers

In this step we first assign the class labels to the epochs. That is for every row/epoch in Table 5 we assign an additional column with a class label, see Table 6. If the epoch entirely falls within a non-seizure period then the epoch is labelled as non-seizure (as also explained in Section 3.2 and Figure 3) otherwise the epoch is labelled as seizure. As the consequence of misclassification of a seizure record as a non-seizure record can be much higher than a non-seizure record with a seizure prediction, we use a cost sensitive classifier. Thus we aim to reduce the misclassification of seizure records with non-seizure records (i.e. False Negative predictions). Once we train our cost sensitive classifier on the training data set, being Table 6, we are ready to predict/detect seizure in the new unlabelled signal. When we receive new unlabelled signal we first convert into a record as shown in Table 5. The trained classifier is then used to label the record.

### 4. Experimental Results

#### 4.1 Data

We used data from Goldberger et al. (2000), which was collected at the Children’s Hospital Boston, Massachusetts (MIT). This consists of EEG recordings from paediatric subjects with intractable seizures. There are recordings from 24 different investigations (cases) from 23 subjects (5 males, 17 females and 1 with gender not specified). The International 10-20 system of EEG electrode positions (for 23 electrodes) and nomenclature was used for...
Algorithm 1: EEG Analysis via MODWT decomposition and Wavelet Correlation

**input**: A Signal input $DB_I$, of individuals’ EEG recordings

**output**: $A_{i,j}$ : Output for individual per epoch from the Cost sensitive Classifier.

Initialization: Build transformed database

$I_j \leftarrow DB_I$ //Download initial data file per individual

for All individuals $I_j$ do

$I_{i,j} \leftarrow I_j$ //Segment individual’s data into 1 second Epochs

for Each Individual by Epoch $I_{i,j}$ do

**Step 1.**

for Each Electrode $I_{i,j,e}$ do

$\tilde{W}_i \leftarrow I_{i,j,e}$ // Apply MODWT, decompose to required levels

$I_{i,j,e,R} \leftarrow \tilde{W}_i$ // Omit noise levels, use Inverse MODWT to reconstruct reduced signal

$Sp_i \leftarrow I_{i,j,e,R}$, // Derive statistical parameters from new signal

end

$RW_{i,j,e} \leftarrow Sp_1, Sp_2, \ldots, Sp_n$ //Form attributes from statistical parameters

end

for Each Individual by epoch $I_{i,j}$ do

**Step 2.**

for Each Electrode $I_{i,j,e}$ do

$\tilde{W}_i \leftarrow I_{i,j,e}$ // Apply MODWT, decompose to first wavelet detail level

end

for Each electrode pair $\{I_{i,j,\tau}, I_{i,j,\zeta}\}$ do

$\text{cov}\{\tilde{W}_{i,\tau}, \tilde{W}_{i,\zeta}\} \leftarrow \tilde{W}_{i,\tau}, \tilde{W}_{i,\zeta}$ // derive wavelet covariance between electrodes

$\rho_{\tau\zeta} \leftarrow \text{cov}\{\tilde{W}_{i,\tau}, \tilde{W}_{i,\zeta}\}$ // derive correlation from standardised covariance

end

$Thresh \leftarrow 0 < t_h \leq 1$ // determine correlation threshold level $t_h$

$Adj_{ee} \leftarrow 1$ where $\rho \geq Thresh$ // construct adjacency matrix

$Sp_i \leftarrow \sum r_i$ // derive statistical parameters from row sums of adj matrix

$C_{i,j} \leftarrow Sp_1, Sp_2, \ldots, Sp_k$ // Use statistical parameters to form attributes

$RW_{i,j,e}, C_{i,j} \leftarrow C_{i,j}$ // combine with attributes from Step 1.

**Step 3.**

$Eglob_i \leftarrow Adj_{ee}$ // From Adjacency matrix derive Global efficiency

$Sp_i \leftarrow Eglob_i$ // derive statistical parameters

$G_{i,j} \leftarrow Sp_1, Sp_2, \ldots, Sp_k$ // Use statistical parameters to form attributes

$RW_{i,j,e}, C_{i,j}, G_{i,j} \leftarrow G_{i,j}$ // combine with attributes from Step 2

end

**Step4.** $DB_T \leftarrow RW_{i,j,e}, C_{i,j}, G_{i,j}, label$ // add label and build database (for training classifier upon)

end

$A_{i,j} \leftarrow DB_T$ //Apply Cost sensitive Classifier
Grant and Islam

these recordings with the data sampled at 256Hz. More details about the data set may be found from https://archive.physionet.org/pn6/chbmit/. For this study we used a single file from each of 12 individuals, these individuals were identified in the MIT data by the case numbers; 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 23 and 24. The cases were selected on the basis of

- the data across cases was reported with same identical contiguous electrode sequence, i.e. \(\{FP1, F7, T7, \ldots, FT10, P8\}\).
- each case was a different individual
- limited selection to files up to two hours in duration
- no seizure event within the first 10 minutes.

Most of the selected files were approx one hour in duration. As the files have 256 data points per second for each electrode signal, therefore one hour equates to \(3600 \times 256 = 921600\) data points, per electrode.

A few files were two hours in length, for these files we selected either the first or second hour depending upon which hour had the first seizure event. This raw data is presented with indicated periods of seizure, we subset the data into one second intervals (i.e. an epoch length of one second) and attached labels indicating the class of that epoch, i.e. non-seizure or seizure. We have approx 1 hour of recording from each patient. For two of the patients there were no seizures recorded within the hour sample, we still included these two into our training set. Of these two, one patient simply had no seizures recorded in files of up to two hours duration, the other patient’s files had the sequence in which the electrodes were reported, altered before first file with a seizure occurred.

### 4.2 Synopsis of results and comparisons

In this study True Positive (TP) refers to events where epochs labelled seizure were classified as seizure. False Negative (FN) refers to events where seizure labelled epochs were classed as non-seizure. True negative (TN) refers to epochs labelled non-seizure and classified non-seizure and False Positive (FP) refers to a epoch labelled non-seizure and classified seizure. We use the following metrics or indicators: Accuracy, Recall, Precision and f-score. They are defined as,

\[
Accuracy\% = \frac{TP + TN}{TP + FN + TN + FP} \times 100
\]

\[
Recall = \frac{TP}{TP + FN}
\]

\[
Precision = \frac{TP}{TP + FP}
\]

\[
f-score = \frac{2}{(Recall)^{-1} + (Precision)^{-1}} = 2 \times \frac{Precision \times Recall}{Precision + Recall}
\]

We use these indicators to compare our methodology to the other methods. As we are interested in correctly classifying seizure we use a cost sensitive/class imbalanced algorithm.
Table 7: Penalty Cost Structure for CSForest.

| Predicted Value | Actual Value |   |   |
|-----------------|--------------|---|---|
|                 | Positive     | TP = 1 | FP = 1 |
|                 | Negative     | FN = 15 | TN = 0 |

Here we choose CSForest (Siers and Islam, 2015) and set the cost penalty structure as shown in Table 7.

We arrived at this cost structure by training on all individuals in the transformed data set, except individual 1. From this initial reduced data set, we implemented a 66% training 34% test split on this data set for validation of the CSForest model (i.e. we repeated this training using 66% of data for training and testing on 34%, altering the penalty cost, with the objective of increasing the f-score while maintaining a high Recall value). As this penalty cost structure shown in Table 7 provided us with a close to optimal result, in regards to f-score\(^8\), we left the CSForest cost structure unchanged throughout the rest of experiment across all individuals in our data set.

We also compared our results to the two other methods outlined in Section 4.3 and used leave-one-out cross validation for all methods. By using the macro average method, see Definition 6 Appendix A.2, for determining an average f-score across individuals per method, our method returned a higher average f-score than the other two comparative methods, as shown in Table 8. For the f-score results at an individual level see Table 9. Our technique (labelled Method No. 3 in Tables 8 & 9) usually returns better results in respect to this balanced f-score across these individuals. We choose the f-score as it weights both Precision and Recall equally.

Table 8: Average f-score results via Method.

| Method No. | Method origin              | Average f-score |
|------------|----------------------------|-----------------|
| 1          | Siddiqui et al. (2018)     | 0.37            |
| 2          | Chen et al. (2017)         | 0.42            |
| 3          | Our Proposed Technique     | **0.58**        |

Remark 1. There were two other individuals included in our data set and neither of these had a seizure state in the file selected. Since there is no seizure record, all three classification methods all provided near perfect results in regards to accuracy and f-score for these two individuals, (not shown).

\(^8\) The f-score is the harmonic mean of precision and recall, sometimes called a balanced f-score.
Table 9: f-scores via Individual across Methods

| Method Number | f-scores | Method Number | f-scores |
|---------------|----------|---------------|----------|
| Ind. No.      | 1        | 2             | 3        |
| 1             | 0.136    | 0.178         | 0.646    |
| 2             | 0.081    | 0.062         | 0.326    |
| 3             | 0.213    | 0.042         | 0.385    |
| 4             | 0.039    | 0.0           | 0.185    |
| 23            | 0.018    | 0.0           | 0.395    |

4.2.1 Metrics chosen

In our situation we wish maximise True Positives, i.e. those with seizure classified as seizure as well as minimise False Negatives, i.e. a seizure classified as a non-seizure. The metric Recall is important here. Recall provides a measure of the True Positives, divided by the sum of True Positives and False Negatives. Alternatively, Precision which is equal to True positives divided by the sum of True Positives and False Positives. Here False Positives are those in a non-seizure state but classified as seizure. Therefore while we aim to detect all seizures as seizures, classifying non-seizures as seizures is at worst, a waste of resources. However classifying seizures as non-seizures may lead to catastrophic outcomes.

Hence we consider Recall slightly more relevant than Precision for this exercise as Precision does not take into account False Negatives. Precision does however consider False Positives. Table 10 which highlights all the chosen indicators at an individual level, demonstrates that our proposed method (Method Number 3), returns higher values of Recall across nearly all individuals. In some instances Precision values from the comparative methods are higher than from our method, i.e. Individual No 1. This could simply be a situation of a small proportion of the overall number of seizures classified as seizures but with no non-seizures classified as seizures. There is no indication in this Precision metric of how many False Negatives were generated in the classification process. This is why we have chosen f-score as our main indicator of overall performance, to balance out Recall and Precision. Accuracy, shown in Tables 10 & 12 is a misleading metric for such imbalanced data. In our final data set of all individuals we have approx 42816 records of which 822 are labelled seizure. i.e. 1.92%. Hence if we simply classify all records as non-seizure then our Accuracy will be 98.08%.

4.2.2 Alternative Cost Structure within Classifier

If a major concern was False Positives, we could alter our Recall and Precision values by changing the penalty cost structure within the CSForest classifier. In comparison to those values shown in Table 7, by reducing the penalty cost on False Negatives we may reduce the Recall value, conversely while increasing the False Positive cost we may increase the Precision value. As an example of such penalty costs see Table 11.
Table 10: Statistical results via Individual across Methods

| Ind. No. | Metric   | 1     | 2     | 3     | Ind. No. | Metric   | 1     | 2     | 3     |
|----------|----------|-------|-------|-------|----------|----------|-------|-------|-------|
| 1        | Accuracy%| 98.94 | 98.87 | 99.03 | 5        | Accuracy%| 97.47 | 98.00 | 98.56 |
|          | Recall   | 0.073 | 0.098 | 0.780 |          | Recall   | 0.776 | 0.379 | 0.621 |
|          | Precision| 1     | 1     | 0.552 |          | Precision| 0.581 | 1     | 0.900 |
|          | f-score  | 0.136 | 0.178 | 0.646 |          | f-score  | 0.664 | 0.550 | 0.735 |
| 2        | Accuracy%| 74.77 | 97.46 | 94.35 | 7        | Accuracy%| 96.48 | 98.87 | 99.28 |
|          | Recall   | 0.488 | 0.037 | 0.598 |          | Recall   | 0.330 | 0.567 | 0.732 |
|          | Precision| 0.044 | 0.200 | 0.224 |          | Precision| 0.327 | 1     | 0.986 |
|          | f-score  | 0.081 | 0.062 | 0.326 |          | f-score  | 0.328 | 0.724 | 0.840 |
| 3        | Accuracy%| 97.94 | 98.72 | 96.89 | 8        | Accuracy%| 95.39 | 95.61 | 92.08 |
|          | Recall   | 0.208 | 0.021 | 0.729 |          | Recall   | 0.241 | 0.018 | 0.735 |
|          | Precision| 0.217 | 0.1   | 0.261 |          | Precision| 0.476 | 1     | 0.33  |
|          | f-score  | 0.213 | 0.042 | 0.385 |          | f-score  | 0.32  | 0.037 | 0.455 |
| 4        | Accuracy%| 98.67 | 98.36 | 94.83 | 10       | Accuracy%| 97.77 | 99.61 | 99.61 |
|          | Recall   | 0.02  | 0.0   | 0.429 |          | Recall   | 0.82  | 0.77  | 0.813 |
|          | Precision| 0.50  | 0.0   | 0.118 |          | Precision| 0.420 | 1     | 0.963 |
|          | f-score  | 0.039 | 0.0   | 0.185 |          | f-score  | 0.556 | 0.87  | 0.881 |
| 23       | Accuracy%| 96.89 | 96.86 | 96.25 | 24       | Accuracy%| 98.91 | 98.97 | 98.81 |
|          | Recall   | 0.009 | 0.0   | 0.389 |          | Recall   | 0.50  | 0.520 | 0.580 |
|          | Precision| 1     | 0     | 0.40  |          | Precision| 0.641 | 0.667 | 0.569 |
|          | f-score  | 0.018 | 0.0   | 0.395 |          | f-score  | 0.562 | 0.584 | 0.574 |

Table 11: Alternative Penalty Cost Structure for CSForest.

| Predicted Value | Positive | Negative |
|-----------------|----------|----------|
| Positive        | $TP = 1$ | $FP = 2$ |
| Negative        | $FN = 2$ | $TN = 0$ |

Applying the CSForest classifier to the transformed data for the individuals No. 1 and No. 8, but using the cost structure as shown in Table 11, we note from the results in Table 12 that the Recall, Precision and f-score have altered when compared to Table 10.

4.3 Comparative methods

We compared our classification results to two other published methods, using the exact same raw data we had used, i.e one second epochs from the MIT data set, with results mentioned in Section 4.2, Tables 8, 9 & 10. The two works we compared to are as follows:

Method 1. Siddiqui et al. (2018) used nine statistical parameters, i.e. (Min, Max, . . . , Skewness) derived from each electrode signal to build the attribute space, resulting in $9 \times 23 = 207$
Table 12: Statistical results via Individual with Alternative Cost Structure

| Ind. No. | Metric    | Result | Ind. No. | Metric    | Result |
|----------|-----------|--------|----------|-----------|--------|
| 1        | Accuracy% | 98.89  | 8        | Accuracy% | 96.36  |
|          | Recall    | 0.024  |          | Recall    | 0.191  |
|          | Precision | 1.0    |          | Precision | 1.0    |
|          | f-score   | **0.048** |          | f-score   | **0.321** |

attributes plus label per record, then applied an ensemble classifier to data, in this case being SysFor (Islam and Giggins, 2012).

**Method 2.** Chen et al. (2017), utilised the DWT to decompose the electrode signal into different frequency bands, selected a subset of these frequency bands and derived statistical parameters from the wavelet coefficients in the chosen frequency bands (wavelet decomposition scales) then applied the classifier SVM, (via a leave one out methodology), to their transformed data. This was repeated for different mother wavelets, (i.e. Daubechies, Haar, Symlets, etc ...). For comparison we used this publication’s methodology for the Haar wavelet, being six frequency bands together with three statistical parameters: Minimum, Standard deviation and Skewness. This resulted in $6 \times 3 \times 23 = 414$ attributes plus label per record.

**4.4 Application of our proposed Method in the The Experiments**

For this experiment, we used open source software, being R (R Core Team, 2019) for all Graph Theory methods, Wavelet and Statistical analysis. WEKA (Hall et al., 2009) was utilised for the Classification step.

**4.4.1 Step 1: Preparation of Attributes from statistical parameters of reconstructed signal.**

From the initial data that had been segmented into one second epochs, we initially applied the wavelet transform$^9$. As the original data had been sampled at 256Hz, using the MODWT decomposition we could decompose the signal into six frequency bands to match to the advised bandwidths, our decomposition/frequency bandwidth range shown in Table 13.

We observed that after such wavelet decomposition, only minor signal energy was apparent in the first two wavelet decomposition bands$^{10} \tilde{D}_1 \& \tilde{D}_2$, which represent 32Hz to 128Hz. This was noticed across many electrodes, both in non-seizure and seizure states, as an example from one individual for one electrode see Figure 8.

Signal energy at each sub-band (Crystal), e.g. $\tilde{D}_4 \sim$ Alpha, is defined similarly as in total energy in signal (time series), see Definition 2, Appendix A.2

---

9. Using wavelet filter “s8” which had been previously used successfully in Achard (2012) & Chen et al. (2017).
10. These sub-bands are also called Crystals
Table 13: Frequency Band decomposition

| Level | Frequency Band  |
|-------|----------------|
| \( \tilde{D}_1 \) | 64-128Hz High Gamma |
| \( \tilde{D}_2 \) | 32-64Hz Gamma |
| \( \tilde{D}_3 \) | 16-32Hz Beta |
| \( \tilde{D}_4 \) | 8-16Hz Alpha |
| \( \tilde{D}_5 \) | 4-8Hz Theta |
| \( S_5 \) | 0-4Hz Delta |

As an example: Energy of Alpha sub-band:

\[
\text{Energy}\sqrt{\tilde{D}_4} = \sum_{i=1}^{n} |d_{1,4}|^2 \text{, where } d_{1,4}, d_{2,4}, \ldots, d_{n,4} \in \tilde{D}_4
\]  

Figure 8: MODWT Decomposition of 1 second Epochs

After eliminating the first two wavelet levels (Crystals), we then applied the inverse wavelet transform to reconstruct the signal, where the non-relevant frequency ranges had been omitted.

*feature extraction.* After application of the inverse wavelet transform on the reduced set of wavelet coefficients, for the epoch, this resulted in a time series of length 256 points (1 second epoch). For each of these reconstructed signals, per epoch we derived the statistical
parameters and followed the procedure as outlined in Section 3.2. This provides in total for the 23 electrodes, 115 attributes for each epoch. Table 1 is an example of the output derived here.

4.4.2 Step 2: Preparation of Attributes from Connectivity of Electrodes

By applying the MODWT to our data, using same wavelet filter as in Section 4.4.1, for the time series from each electrode, per epoch we derived wavelet covariance and hence wavelet correlation between each time series. We then constructed an adjacency matrix highlighting the direct connections between the electrodes. This was performed using only the wavelet coefficients from the first level of MODWT decomposition, i.e. \( d_{i,1} \in \tilde{D}_1 : i \leq n \) and setting a lower bound on the correlation value of 0.125. Using the lower bound value we determined if a connection exists between the electrodes if the correlation value is greater than or equal to the lower bound\(^{11}\). We represented those connections in an adjacency matrix, as described in Section 3.3.

**feature extraction** From the derived adjacency matrix we calculate the sum of the rows. For each electrode in our network we now have a degree of connections or number of direct connections during the epoch. From the row sums (**or degree of connection**) we derive the statistical parameters and follow the procedure as described in Section 3.3. For each epoch this provides us with another five attributes for each record. Table 3 is an example of this.

The data initially chosen consisted of results from 23 electrodes or channels. As one channel was the reverse of the other, i.e., channel 3: T7-P7, channel 19: P7-T7, we dropped channel 19 when calculating the wavelet covariances and hence from the adjacency matrix construction, we had 22 row sums to derive the statistical parameters from, per epoch.

4.4.3 Step 3. Preparation of Attributes from Global Efficiency of Electrodes

From Section 3.3 where wavelet covariance was shown to provide a methodology for the construction of an adjacency matrix derived from the wavelet correlation between the electrodes, we derive how “efficiently” the electrodes are at communicating across their network. This provides us with the global efficiency per node, as described in Section 3.4.

We performed this calculation using the adjacency matrix derived in Step 2 of our experiment, (**Section 4.4.2**), for each one second epoch and 22 electrodes (**nodes**).

**feature extraction** For each electrode in our network we now have an efficiency value, \( E_{glob_j} : 0 \leq E_{glob_j} \leq 1 \) or how well that electrode communicates or transfers information within the network during the epoch. As we have these values, 22 in total, for each epoch, we derive from them the statistical parameters as described in Section 3.4. This provides us with an additional five parameters. We then place these attributes directly following the attributes derived in Step 1 and Step 2 of our method, see Table 5 as an example.

These three steps combined together provides a total 115 + 5 + 5 = 125 attributes.

---

11. The value 0.125 was chosen as with wavelet decomposition level 1, this provided over 150 electrode connections during the non-seizure state and over 30 connections during the seizure state, per epoch.
4.4.4 Step 4. Assigning Class Labels and Training Classifiers

From the initial three steps in our method, we have constructed a new data set from the original data, where we have transformed our multivariate EEG data into a univariate list of attributes per Epoch and attached labels as described in Section 3.5, where Table 6 is an example of this. Our constructed database of 12 patients which equates to approx 42,816 records, each record of 125 attributes plus label. We applied the CSForest classifier to our data, with the cost structure as outlined in Section 4.2, testing each individual against the training data set constructed from the other 11 individuals in this study, i.e. leave one out cross validation method. This classification was undertaken via WEKA. The results of the three different methods per individual are shown in Table 10.

5. Conclusion

In this study, we investigated the use of wavelets and their inherent ability to provide analysis into the time frequency space. We utilised a wavelet transform, that is defined for all lengths of sample, able to provide selection of relative frequencies as well as deriving correlation between signals to construct different indicators. Together with some basic methodology from graph theory, we applied these to an open source data set and compared our results with two other published methods which we applied to exactly the same data. Our method compared well, providing overall better results, in regards to f-score and Recall.

Our methodology of utilising three steps (consisting of Wavelet applications and Graph theory) to construct the attributes to which the classifier will be applied, does require some additional computation, but returns a similar number of attributes when compared to usual methods that construct statistical parameters for attributes derived from raw data. However this method returns considerably less attributes when compared to methods deriving statistical parameters from wavelet coefficients across different decomposition levels.

During classification we kept the initial penalty cost structure within the cost sensitive classifier unchanged. This could be altered by determining the optimal cost structure each time the training data set is altered or after inclusion of additional data. The chosen data had 22 electrode positions that could be used to derive node correlation and connectivity. This limited the number of connections available at higher levels of wavelet decomposition without considerably lowering the correlation threshold. These higher levels of wavelet decomposition may have been better suited for seizure discrimination however we did not use owing to the limit on derived connections. For electrode connectivity calculation we used the first Crystal $D_1$ as this decomposition level provided a larger number of connections across the electrodes, especially during seizure, in comparison to the other wavelet levels.

Also we did not weight the distance between the nodes, simply gave a derived connection a distance value or path length of one. Further investigation using data sets with many more nodes together with spatial coordinates which would enable weighting for distance apart, may prove better at determining electrode connectivity especially in the surrounding locality of seizure, to provide enhanced discrimination between the different EEG states.

Conditional upon acceptance, the associated R code will be made available via Github.
Appendix A.

A.1 Glossary

\( A_{i,j} \) Output from classifier, per epoch

\( Adj_{ee} \) Adjacency Matrix

\( C_{i,j} \) List of attributes, Step 2

\( DB_T \) Database of transformed data

\( d_{i,j} \) \( i^{th} \) detail wavelet coefficient at \( j^{th} \) level

\( D_j \) \( j^{th} \) level of detail coefficients, DWT

\( \tilde{D}_j \) \( j^{th} \) level of detail coefficients, MODWT

\( DWT \) Discrete Wavelet transform

\( EEG \) Electroencephalogram

\( Eglob_i \) Global Efficiency value, electrode \( i \)

\( FN \) False Negative

\( FP \) False Positive

\( G_{i,j} \) List of attributes, Step 3

\( HPF \) High Pass filter

\( Hz \) Hertz, One cycle per second

\( I_j \) Individual’s data

\( I_{i,j} \) Individual’s data by epoch

\( I_{i,j,e} \) Individual’s data by epoch by electrode

\( I_{i,j,e,R} \) Individual’s reconstructed data

\( LPF \) Low Pass filter

\( MODWT \) Maximum Overlap Discrete Wavelet Transform

\( MRA \) Multi Resolution Analysis

\( \rho_{\tau\zeta} \) Correlation between the two electrodes \( (\tau, \zeta) \) per epoch

Quartiles values that divide a list of numbers into quarters

\( r_i \) a row in Adjacency matrix

\( R_{i,j} \) Reciprocal of shortest path between two electrodes

\( RW_{i,j,e} \) List of attributes, Step 1

\( S_{i,J} \) \( i^{th} \) smooth wavelet coefficient at \( J^{th} \) level

\( S_J \) \( J^{th} \) level of smooth coefficients, DWT

\( \tilde{S}_J \) \( J^{th} \) level of smooth coefficients, MODWT

\( Sp_i \) a set of statistical parameters

Skewness Measure of asymmetry of a probability distribution.

\( SVM \) Support Vector Machine, a classifier

\( t_h \) derived threshold level

\( TN \) True Negative

\( TP \) True Positive

\( \muV \) MicroVolt, one millionth of a volt

\( W \) A vector of DWT wavelet coefficients

\( W_j \) vector of detail DWT coefficients, \( j^{th} \) level

\( V_J \) vector of smooth DWT coefficients, \( J^{th} \) level
A.2 Definitions

Definition 1 Let a sequence $X_1, X_2, \ldots, X_n$ represent a time series of $n$ elements, denoted as $\{X_t : t = 1, \ldots, n\}$, $X_t \in \mathbb{R}$.

Definition 2 The energy within a time series $X_t$ is the sum of the squared norm of $X_t$,

$$\| X_t \|^2 = \langle X, X \rangle = \sum_{t=1}^{n} X_t^2$$

Definition 3 let $\overline{X}$ denote the sample mean of $X_t$

$$\overline{X} \equiv \frac{1}{n} \sum_{t=1}^{n} X_t$$

Definition 4 let $\hat{\sigma}_X^2$ denote the sample variance of $X_t$

$$\hat{\sigma}_X^2 \equiv \frac{1}{n} \sum_{t=1}^{n} (X_t - \overline{X})^2$$

Definition 5 let $\text{cov}\{X_t, Y_t\}$ denote the covariance of two random variate sequences $X_t$ and $Y_t$

$$\text{cov}\{X_t, Y_t\} = \sum_{t=1}^{n} \frac{(X_t - \overline{X})(Y_t - \overline{Y})}{n}$$

Definition 6

macro average Recall (MAR) = $\frac{R_1 + R_2 + \ldots + R_N}{N}$

macro average Precision (MAP) = $\frac{P_1 + P_2 + \ldots + P_N}{N}$

macro average f-score = $2 \times \frac{\frac{1}{\text{MAR}} + \frac{1}{\text{MAP}}}{2} = 2 \times \frac{\text{MAR} \times \text{MAP}}{\text{MAR} + \text{MAP}}$

see Torgo (2014)

Definition 7 In a Graph $G$ (or network) the global efficiency measures how the information is propagating in the entire network. Efficiency = the inverse of the harmonic mean of the minimum path length $L_{jk}$ between a node $j$ and all the other nodes $k$ in the graph.

$$E_{\text{glob}} = \frac{1}{N-1} \sum_{k \in G} \frac{1}{L_{jk}}$$

as shown in Achard (2012)
A.3 Wavelets

A.3.1 DWT

The discrete wavelet transform (DWT) returns a data vector of the same length as that of the input. Usually, in this new vector many data points are almost zero. This transform provides an additive decomposition of $J+1$ levels.

$$X_t = \sum_{j=1}^{J} D_j + S_J$$  \hspace{1cm} (2)

where

- $D_j$ $j^{th}$ level of wavelet detail coefficients, related to changes at scale $j$
- $S_J$ $J^{th}$ level of wavelet smooth coefficients, related to averages at scale $J$
- $\{X_t : t = 1, \ldots, n\}$ where $n = 2^J : J \in \mathbb{Z}^+$, $X_t \in \mathbb{R}$,

The DWT is a linear transform which decomposes $X_t$, into $J$ levels giving $n$ DWT coefficients; the wavelet coefficients are obtained by premultiplying $X$ by $W$

$$W = WX$$  \hspace{1cm} (3)

- $W$ is a vector of DWT coefficients ($j$th component is $W_j$)
- $W$ is $n \times n$ orthonormal transform matrix; i.e., $W^T W = I_n$, where $I_n$ is $n \times n$ identity matrix
- inverse of $W$ is its transpose, $\implies W W^T = I_n$
- $\therefore W^T W = W^T WX = X$

$W$ is partitioned into $J+1$ subvectors

- $W_j$ has $n/2^j$ elements \(^{12}\)
- $V_J$ has one element \(^{13}\)

the synthesis equation for the DWT is:

$$X = W^T W = [W_1^T, W_2^T, \ldots, W_J^T, V_J^T] \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_J \\ V_J \end{bmatrix}$$  \hspace{1cm} (4)

Equation 4 leads to additive decomposition which expresses $X$ as the sum of $J+1$ vectors, each of which is associated with a particular scale $T_j$

$$= \sum_{j=1}^{J} W_j^T W_j + V_J^T V_J \equiv \sum_{j=1}^{J} D_j + S_J$$  \hspace{1cm} (5)

\(^{12}\) note: $\sum_{j=1}^{J} \frac{n}{2^j} = \frac{n}{2} + \frac{n}{2^2} + \cdots + 2 + 1 = 2^j - 1 = n - 1$

\(^{13}\) If we decompose $X_t$ to level $I \leq J$ then $V_I$ has $n/2^I$ elements.
• \( \mathcal{D}_j \equiv W_j^T W_j \) is portion of synthesis due to scale \( \mathcal{T}_j \), called the \( j \)th 'detail'

• \( \mathcal{S}_j \equiv V_j^T V_j \) is a vector called the 'smooth' of the \( J \)th order

A.3.2 MODWT

However another variant of the wavelet transform exists called the Maximum Overlap Discrete Wavelet Transform (MODWT), where the number of coefficients returned by the transform at each level of decomposition is the same as that of the input.

• The MODWT is not orthonormal and is highly redundant.

• The MODWT is defined for \( X_t \) on all sample sizes, \( n \) does not need to be a power of two

• The MODWT to level \( J \) has \((J + 1) \times n \) coefficients, the DWT has a total of \( n \) coefficients for any chosen \( J \).

• Similarly to the DWT, the MODWT permits an MRA

• Similarly to the DWT, the MODWT permits an analysis of variance to be based on MODWT coefficients.

As with the DWT, from the MODWT we may also obtain a scale based additive decomposition. Where the \( \tilde{\cdot} \) represents the coefficients are derived from the MODWT.

\[
X_t = \tilde{W}_j^T \tilde{W}_j + \tilde{V}_J^T \tilde{V}_J \equiv \sum_{j=1}^{J} \tilde{D}_j + \tilde{S}_J
\]  

(6)

Also a scale based energy decomposition, (\textit{can be seen as a basis for ANOVA})

\[
\| X_t \|^2 = \sum_{j=1}^{J} \| \tilde{W}_j \|^2 + \| \tilde{V}_J \|^2
\]  

(7)

We may obtain a decomposition of the sample variance, from Definition 4, it follows:

\[
\hat{\sigma}_X^2 = \frac{1}{n} \sum_{j=1}^{J} \| \tilde{W}_j \|^2 + \frac{1}{n} \| \tilde{V}_J \|^2 - \bar{X})^2 \equiv \sum_{j=1}^{J} \nu_X^2(\mathcal{T}_j)
\]  

(8)

We may consider a wavelet variance \( \nu_X^2(\mathcal{T}_j) \) as a portion of \( \hat{\sigma}_X^2 \) related to changes in averages over scale \( \mathcal{T}_j \) i.e., scale by scale analysis of variance. Or the wavelet variance for scale \( \mathcal{T}_j \) may be defined as the variance of \( \tilde{W}_j \)

\[
\nu_X^2(\mathcal{T}_j) \equiv \text{var}\{\tilde{W}_j\}
\]  

(9)
A.3.3 Wavelet Correlation

If we have two sequences (or signals) $X_t \& Y_t$ as per Definition 1, then with wavelet coefficients $\{\tilde{W}_{X,j}\}$ and $\{\tilde{W}_{Y,j}\}$, we have: wavelet covariance

$$\nu_{XY}(T_j) \equiv \text{cov}\{\tilde{W}_{X,j}, \tilde{W}_{Y,j}\}$$

(10)

which similar to wavelet variance, provides a decomposition at scale $T_j$, of the covariance between $\{X_t\}$ and $\{Y_t\}$:

$$\sum_{j=1}^{\infty} \nu_{XY}(T_j) = \text{cov}\{X_t, Y_t\}$$

(11)

Wavelet covariance may be standardised. which yields, wavelet correlation:

$$\rho_{XY} \equiv \frac{\text{cov}\{\tilde{W}_{X,j}, \tilde{W}_{Y,j}\}}{\left(\text{var}\{W_{X,j}\}, \text{var}\{W_{Y,j}\}\right)^{1/2}} = \frac{\nu_{XY}(T_j)}{\nu_X(T_j)\nu_Y(T_j)}$$

(12)

where $\nu_X(T_j)$ and $\nu_Y(T_j)$ are the wavelet variances for $\{X_t\}$ and $\{Y_t\}$, (see Percival and Walden. 2000, Chap 8).

References

S. Achard. brainwaver: Basic wavelet analysis of multivariate time series with a visualisation and parametrisation using graph theory, 2012. URL https://CRAN.R-project.org/package=brainwaver. R package version 1.6.

S. Ashok and P. Mahalakshmi. Wavelet-based feature extraction for classification of epileptic seizure EEG signal. Journal of Medical Engineering and Technology, 41:1–11, 11 2017. doi: 10.1080/03091902.2017.1394388.

D. Chen, S. Wan, J. Xiang, and F. Bao. A high-performance seizure detection algorithm based on discrete wavelet transform (DWT) and EEG. PLOS ONE, 12:e0173138, 03 2017. doi: 10.1371/journal.pone.0173138.

I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf. Theory, 36:961–1005, 1990.

J. Elsa Jacob, G. Nair, T. Iype, and A. Cherian. Diagnosis of encephalopathy based on energies of EEG subbands using discrete wavelet transform and support vector machine. Neurology Research International, 2018:1–9, 07 2018. doi: 10.1155/2018/1613456.

O. Faust, U. R. Acharya, H. Adeli, and A. Adeli. Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure, 26, 01 2015. doi: 10.1016/j.seizure.2015.01.012.

A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. 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, 101(23):e215–e220, 2000. Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.
P. Grant and M. Z. Islam. Clustering noisy temporal data. In *15th International Conference on Advanced Data Mining and Applications (ADMA 2019)*, Lecture Notes in Computer Science, pages 185–194. Springer, November 2019. ISBN 9783030352301. doi: https://doi.org/10.1007/978-3-030-35231-8_13.

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18, 2009.

N. E. Helwig. *eegkit: Toolkit for Electroencephalography Data*, 2018. URL https://CRAN.R-project.org/package=eegkit. R package version 1.0-4.

M. Z. Islam and H. Giggins. Knowledge discovery through SysFor: A systematically developed forest of multiple decision trees. *Conferences in Research and Practice in Information Technology Series*, 121, 01 2012.

S. Janjarasjitt. Classification of the epileptic EEGs using the wavelet-based scale variance feature. *Int J Appl Biomed Eng*, 3:19–25, 01 2010.

S Lee, J Lim, J Kim, J Yang, and Y Lee. Classification of normal and epileptic seizure eeg signals using wavelet transform, phase-space reconstruction, and euclidean distance. *Comput Methods Programs Biomed*, 116(1):10–25, 08 2014. doi: 10.1016/j.cmpb.2014.04.012.

I. Omerhodzic, S. Avdakovic, A. Nuhanovic, and K. Dizdarevic. Energy distribution of EEG signals: EEG signal wavelet-neural network classifier. *Int. J. Biol. Life Sci.*, 6, 07 2013.

Md Z. Parvez and M. Paul. EEG signal classification using frequency band analysis towards epileptic seizure prediction. *16th Int’l Conf. Computer and Information Technology, IC-CIT 2013*, 12 2013. doi: 10.1109/ICCITechn.2014.6997315.

D. Percival and A. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000. doi: 10.1017/CBO9780511841040.

R. Polikar. The Wavelet Tutorial, The engineer’s ultimate guide to wavelet analysis, 1994. URL users.rowan.edu/~polikar/WTtutorial.html.

H. Qu and J. A. Gotman. A patient-specific algorithm for the detection of seizure onset in long-term eeg monitoring: Possible use as a warning devices. *IEEE Trans Biomed Eng*, IT-44(2):115, 1997.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

S. Shrestha, R. Shrestha, and B. Thapa. Implementing neural network and multi resolution analysis in EEG signal for early detection of epilepsy. *SCITECH Nepal*, 14:8–16, 09 2019. doi: 10.3126/scitech.v14i1.25528.
Md K. Siddiqui, M. Z. Islam, and A. Kabir. A novel quick seizure detection and localization through brain data mining on ECoG dataset. *Neural Computing and Applications*, pages 1–14, 03 2018. doi: 10.1007/s00521-018-3381-9.

M. Siers and M. Z. Islam. Software defect prediction using a cost sensitive decision forest and voting, and a potential solution to the class imbalance problem. *Information Systems*, 51:62–71, 07 2015. doi: 10.1016/j.is.2015.02.006.

A. Subasi. EEG signal classification using wavelet feature extraction and a mixture of expert model. expert systems with applications 32, 1084-1093. *Expert Systems with Applications*, 32:1084–1093, 05 2007. doi: 10.1016/j.eswa.2006.02.005.

L. Torgo. An infra-structure for performance estimation and experimental comparison of predictive models in R. *CoRR*, abs/1412.0436, 2014. URL http://arxiv.org/abs/1412.0436.

R. Vikash. Comparison of DFT and wavelet based image modification techniques. *International Journal of Computer Science and Mobile Computing*, 4:61–65, 8 2015. URL https://www.ijcsmc.com/docs/papers/August2015/V4I8201524.pdf.

A. Zandi, M. Javidan, G. Dumont, and R. Tafrershi. Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. *IEEE transactions on bio-medical engineering*, 57:1639–51, 07 2010. doi: 10.1109/TBME.2010.2046417.