Electroencephalogram analysis for Automatic Epileptic Seizure detection method using PyEEG

Thara D K\textsuperscript{a}, Dr B G Premasudha\textsuperscript{b}

\textsuperscript{a}Department of ISE, CIT, Gubbi, Karnataka, India;  
\textsuperscript{b}Department of CSE, SIT, Tumkur, Karnataka, India

Abstract: Epilepsy is a precarious nervous system disorder caused by unprovoked recurrent seizures. Most commonly used tool for diagnosing epilepsy is EEG (Electroencephalogram). Manual diagnosis of epilepsy is a laborious task for the neurologists. The diagnosis of epilepsy can be improved by developing an EEG based automatic seizure detection system which acts as a supporting system for neurologists to make more accurate decisions during diagnosis. Automatic detection system is of great help in clinical practice. This paper presents classification of epileptic and non-epileptic seizure events using Deep learning technique. In this study we used PyEEG, open source software for feature extraction. We used five different features which include fractal dimension features like, Petrosian Fractal Dimension (PFD) and Higuchi Fractal Dimension (HFD). Statistical features like Detrended Fluctuation Analysis (DFA), Fisher Information (FI) and entropy feature Singular Value Decomposition Entropy (SVDEn). These features are then fed as input to support vector machine classification model. And in this study we confirm that our model shows good accuracy of 95%.

Keywords: EEG, Seizure, Epilepsy, Feature extraction, Support vector machine, PyEEG.

1. INTRODUCTION

Electroencephalogram [3] is the most commonly used brain-imaging technique to read the electrical signals exchanged between neurons inside the brain. Electrical signals are recorded by placing electrodes on the scalp of the patient. For people who experience epilepsy, if EEG is performed within 24-48 hours of the occurrence of the seizure, the abnormality can be observed in 70% of the individuals. Signals with spikes and sharp wave patterns can be observed in EEG recordings [6]. EEG also has excellent time resolution. Also it is inexpensive, portable and directly records the neural activity inside the brain. Because of these properties of EEG, it is the one, which is very widely used tool by many neurologists to diagnose epilepsy. In traditional approach, physicians use manual method to analyse EEG recordings. Manual method is time consuming and requires trained experts and physicians to understand lengthy EEG recordings. Lots of researches have been carried out in developing a computer aided diagnosis (CAD) system. This CAD system helps physicians in making accurate decisions about the diagnosis of epilepsy. It can, not only help in detecting epilepsy but also can be extended to predict epileptic seizure events before actually it occurs [4]. This CAD system can help reduce failing to take correct decisions while diagnosing epilepsy.

Many algorithms have been used and proposed by so far for feature extraction and classification tasks. Feature extraction is an important step to achieve good results in the classification. EEG data consists of both time domain features and frequency domain features. Fourier transform [5] is the method most commonly used for frequency domain feature extraction. In this work we are using a freely available python module PyEEG to perform feature extraction.

Yannick Roy et al [1], have given an exhaustive review of deep learning based EEG analysis, the works published from 2010 to 2018. They say that more than half of the research works carried out so far for EEG analysis have used publicly available datasets. 40% of the works are done using convolutional neural networks; nearly 14% have used recurrent neural network methods. Before deep learning methods could become so popular, the research works were done using statistical and machine learning algorithms. Helen T.et al [2], have proposed a methodology that can be embedded into hardware for seizure detection. They used...
machine learning algorithm support vector machine for seizure monitoring in epilepsy patients. They used butterworth bandpass filters to preprocess the data. They obtained 95% classification accuracy.

1.1. Contribution and organization of the paper
In this work we perform EEG analysis using deep learning approach. We use PyEEG for feature extraction. It is the only sophisticated feature extraction technique available in python and suitable for machine learning and deep learning approaches. We perform classification task, classifying seizure events from non-seizure events. This is a three class classification problem. We got classification accuracy 95%. The rest of the paper is as follows: Section II presents the methodology of the experiment. Section III describes the details of the dataset and various features used for the experiment. Section IV discusses the results of the experiment followed by conclusion.

1.2. PyEEG
It is a python module used for extracting feature from electroencephalogram. It comprises of two types of functions, preprocessing functions and feature extraction functions (figure 1). Preprocessing functions do not have any value where as feature extraction functions return the feature value of the eeg data. For using PyEEG we need to just download (https://code.google.com/archive/p/pyeeg/downloads) PyEEG and save it as pyeeg.py under current python path. In the program we can directly import PyEEG just like importing other modules. In PyEEG, EEG time series data are represented as a numpy array or list. Because of this reason PyEEG cannot read and load ASCII files and EDF files. If we use EDF file or ASCII file we need to use other open source tools like EEG LAB to convert files before feeding eeg time series data to PyEEG. We can extract the following features from EEG data [10].

<table>
<thead>
<tr>
<th>Features available in PyEEG</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Intensity Ratios</td>
<td>1-D Vector</td>
</tr>
<tr>
<td>Petrosian Fractal Dimension</td>
<td>Scalar</td>
</tr>
<tr>
<td>Higuchi Fractal Dimension</td>
<td>Scalar</td>
</tr>
<tr>
<td>Hjorth mobility &amp; complexity</td>
<td>2-D Vector</td>
</tr>
<tr>
<td>Spectral Entropy</td>
<td>Scalar</td>
</tr>
<tr>
<td>SVD Entropy</td>
<td>Scalar</td>
</tr>
<tr>
<td>Fisher Information</td>
<td>Scalar</td>
</tr>
<tr>
<td>Approximate Entropy</td>
<td>Scalar</td>
</tr>
<tr>
<td>Sample Entropy</td>
<td>Scalar</td>
</tr>
<tr>
<td>Detrended Fluctuation Analysis</td>
<td>Scalar</td>
</tr>
</tbody>
</table>

II. METHODOLOGY:
In this experiment we used Jupyter notebook to run our model. We used libraries like numpy, pandas and Matplotlib. For feature extraction we used PyEEG. We used dataset collected in Bonn University Germany. The details of the dataset are available under section III. As a first step, we loaded files from folders. We used all five sets for the experiment.

We created a separate table for each set. Now each table consists of 100 channels with 4097 data points. For each set, from each channel we extracted 5 features (DFA, HFD, SVD_Entropy, Fisher Information and PFD). This step results in 5 feature matrix. Each matrix consisted of 100 rows and 6 columns including class label. Set A & B has the class label 1 representing healthy data, Set C & D has class label 0 representing transition period and Set E has class label -1 representing unhealthy data. The three feature matrixes are then combined into one big table. This matrix is split into train and test subsets. Train set is given as input to the SVM [9] model for training purpose. After that test set is given for classification. SVM was able to classify the input 96% accuracy. Steps used in the experiment are shown in the figure 2 below.
III. Dataset

The dataset used in the experiment was taken from University Bonn, Germany. The dataset consists of 5 sets (A, B, C, D & E). In each set there are 100 channels with 4097 EEG data points. Data in set A & B is extra-cranial healthy EEG. Set C and D consist of Intra-cranial EEG collected during interictal periods. Set E consists of intracranial EEG collected over ictal periods. Below figure 3 shows a sample of Set E data.

![Figure 3: Sample of Set E segments.](image)

EEG recordings are described in the form of five frequency sub-bands namely Delta, theta, alpha, beta and gamma with 0-4Hz, 4-8 Hz, 8-12Hz, 12-30Hz and >30 Hz respectively [7][8]. Features such as PFD, HFD, DFA, FI and SVD_Entropy are extracted from each of the sub-band. The features extracted are then used for classifying seizure events from non-seizure events. We extracted five features for EEG data analysis. Five features are described below:

**Detrended Fluctuation Analysis (DFA)** [10]: DFA is useful for identifying the association in the time series data. To calculate DFA, the time series data is first integrated. DFA takes three arguments.

dfa(P, Ave=None, R=None)

Where, P is the time series 1-d array, Ave is the average value of P and R is the list of box sizes in integer. The return value of dfa is an integer which indicates the slope of the fitting line denoted by Alpha. Higher the Alpha higher is the signal complexity.

**Higuchi Fractal Dimension (HFD)** [10]: HFD builds k new time series from the actual series. Length of each time series is computed and then the average length is computed. HFD takes two arguments.

hfd(X, Kmax)

Where, X is the time series and Kmax is the number of times average length is computed for each time series.

**Singular Value Decomposition (SVD)** [10]:

svd_entropy (P, Tau, DE, L=None)
P: series data, with lag tau and embedding dimension DE (default)
L: list which includes normalized singular values of a matrix. When L becomes none, it performs as shown below to generate singular spectrum:

First, generate an embedding matrix from P, Tau and DE using pyeeg function embed_seq(): M = embed_seq(P, Tau, DE)

Second, using svd, a scipy.linalg function, decompose the embedding matrix M and get a list of singular values:

\[ L = \text{svd}(M, \text{compute}_uv = 0) \]

\[ L : \text{L} = \text{sum}(L) \]

At last, normalize L: L /= sum(L)

**Fisher Information** [10]: Fisher information is a scalar feature. Equation used to calculate fisher information is,

\[ I = \sum_{i=1}^{M-1} \frac{\sigma_{i+1}^2 - \sigma_i^2}{\sigma_i^2} \]

**Petrosian Fractal Dimension (PFD)** [10]: Equation used to calculate PFD is,

pfdd(P, Q = None)

Compute Petrosian Fractal Dimension of a time series from either two cases below:
1. P : time series, by default it is of type list.
2. Q: first order differential sequence of P
In case 1, D is computed using Numpy's difference function.

IV. Experimental Results

**Dataset features extracted from PyEEG:**

Figure 4, 5 and 6 shows the values of the features extracted from set B, E and C respectively. We extracted 5 PyEEG features from the input EEG data. All 5 features are extracted
for all 100 channels of all the sets. Detrended Fluctuation Analysis values are stored in the variable f1. Higuchi Fractal Dimension values are stored in the variable f2. Svd_entropy values are stored in the variable f3. Fisher Information values are stored in the variable f4 and Petrosian Fractal Dimension values are stored in the variable f5.

Figure 4: Set B features with class label 1.0 (Healthy)

Figure 5: Set E features with class label

Figure 6: Set C features with class label

Support vector machine is used for classifying epileptic EEG from non-epileptic EEG. This is a 3-class classification problem. Class 1 represents healthy EEG with label 1. Class 2 represents EEG in the transition period i.e. Inter-ictal period with label 0 and class 3 represents ictal EEG with label -1. Figure 7 below shows the visualization of the dataset plotted using Matplotlib. We can observe from the figure that signals (green) with voltage level around 500mv to 1000mv are the evidence of the occurrence of seizure. Blue lines represent healthy signals and orange lines represent signals in inter-ictal periods.

Figure 7: Visualization of the data set showing healthy, inter-ictal and seizure activity.

The features are extracted from all the sets (DFA, HFD, SVD_Entrophy, Fisher Information and PFD) to create feature class matrix. Each of the feature class matrices is then concatenated. The concatenated matrix now consists of 300 rows, out of which 33% of the rows are used as test set and rest of the rows are used as train set. Test set consisting of 99 rows is then provided as input to SVM for classification. Our model was able to classify the test set with an accuracy of 96%. Accuracy is calculated using confusion matrix,

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP</td>
<td>TN</td>
<td></td>
</tr>
</tbody>
</table>

Formula used for accuracy is given below:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

Where, TP is True Positive
TN is True negative
FP is false positive
FN is False negative

Results obtained from the experiment are given below:

\[
\begin{bmatrix}
90 & 2 \\
0 & 2 \\
5 & 0
\end{bmatrix}
\]

As per the results, classification accuracy of the experiment obtained is 96%.

Conclusion

In this experiment, we tried to develop a method for epilepsy diagnosis using machine-learning algorithm, support vector machine. It is a three-class classification problem, classifying ictal, inter-ictal and non-ictal EEG. For the experiment, we used most commonly used dataset from Bonn University. We used PyEEG open source module for feature extraction. We extracted 5 features from each channel. When features are fed to the classifier, it classified the given test input with an accuracy of 96%. In future, we plan to use deep learning algorithms for classifying EEG data as they do not require feature-extracting techniques.
REFERENCES


[6] Andriana Theodorakopoulou, Machine learning data preparation for epileptic seizures Prediction Technological Educational Institute of Crete, Department of Informatics Engineering, June 2017


