Automatic Epilepsy Detection from EEG signals

Vridhi Rohira*
Indian Institute of Technology Gandhinagar
Gandhinagar, Gujarat, India
vridhirohira@iitgn.ac.in

Shivam Chaudhary
Indian Institute of Technology Gandhinagar
Gandhinagar, Gujarat, India
shivamchaudhary@iitgn.ac.in

Sudip Das
Indian Institute of Technology Gandhinagar
Gandhinagar, Gujarat, India

Krishna Prasad Miyapuram
Indian Institute of Technology Gandhinagar
Gandhinagar, Gujarat, India

ABSTRACT

Epilepsy is a neurological condition characterized by recurrent seizures and affects millions of people all over the world. The abnormal brain electrical activity during an epileptic seizure can be seen with an EEG, which is then read by a trained medical professional to diagnose epilepsy. However, this is often time-consuming, expensive, inaccessible, and inaccurate, thus highlighting the need for automated epilepsy prediction. Previous algorithms for this problem only made use of small data sets which lacked variable, clinical grade data. We used the TUEP dataset to extract features through power spectral density and power spectral connectivity. These features were then classified into epileptic vs non-epileptic using a random forest classifier. Our feature extraction methods using power spectral density and spectral connectivity showed accuracies of over 90% in detecting epilepsy.

KEYWORDS

Epilepsy, EEG, Random forest classifier

ACM Reference Format:
Vridhi Rohira, Shivam Chaudhary, Sudip Das, and Krishna Prasad Miyapuram. 2023. Automatic Epilepsy Detection from EEG signals. In *6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD) (CODS-COMAD 2023), January 4–7, 2023, Mumbai, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3570991.3570995

1 INTRODUCTION

Epilepsy is a neurological condition affecting over 50 million people worldwide and is characterized by recurrent seizures. Seizures involve brief episodes of involuntary movement and likely result from the uncontrolled and excessive electrical discharge of neurons. This abnormal cortical excitability underlying epilepsy can be measured using electroencephalography (EEG). EEG is a device that provides a method of studying brain waves employing an electroencephalograph that amplifies and records the brain's electrical activity using electrodes on the scalp.

EEG is often used to diagnose epilepsy because it is portable, relatively inexpensive, non-invasive, and has a high temporal resolution [6]. However, reading these EEG signals by experts is time-consuming, and the number of false negatives tends to be high [8]. Thus, there is a need for automatic epilepsy detection algorithms that will save time and have high accuracy and sensitivity.

2 RELATED WORKS

Several studies have looked at automated epilepsy or seizure detection using EEG signals. One of the studies made uses of a CNN model and was reported by Acharya et al. [1], wherein they used the Bonn University data set and trained an 11-layered CNN using back-propagation methods, yielding an accuracy of 88.67 percent.

Recently, random forest classifiers have been used and have been found to give higher accuracy. In one study, researchers found the accuracy to be as high as 100 percent [5]. Random forest classifiers have also been used for the early detection of seizures using time and frequency features extracted from intracranial EEG (iEEG) of 10 patients from the European Epilepsy Database. The classifier's sensitivity was 93.84 percent [4].

Some researchers have used feature engineering methods such as the power spectral density of EEG frequency bands as input features. These features are then fed to deep learning classifiers to give higher generalisability on smaller data sets to classify cognitive states using EEG signals [3].

However, the data sets used in all the studies mentioned above were relatively small. The Bonn University data set only consists of recordings from 10 patients with epilepsy and five without. Furthermore, they removed artifacts such as eye movements, which reduces the variability and the generalisability of the data. Thus, larger data sets with more variability are ideal, such as the Temple University Hospital (TUH) data set. A study to distinguish seizure and non-seizure events using a hierarchy graph convolution network (HGCN)on the TUH corpus found an accuracy of 95.72 percent Zeng et al. [9].

Further, one must also acknowledge the challenges associated with applying classifiers for automated epilepsy detection. The primary issues concern selecting suitable statistical features from EEG recordings and the time cost associated with computation. Keeping all of this in mind the current research aimed to use a random forest

*Vridhi Rohira is a postgraduate student at the Centre for Cognitive and Brain Sciences, Indian Institute of Technology Gandhinagar.

CODS-COMAD 2023, January 4–7, 2023, Mumbai, India

ACM ISBN 978-1-4503-9797-1/23/01.
https://doi.org/10.1145/3570991.3570995

classifier in a much larger data set with more significant variability. A random forest classifier can be computed quickly which is essential given the high volume of the chosen data set.

3 METHODOLOGY

3.1 Data Set

The current research used a subset of the Temple University Hospital (TUH) EEG Corpus. It includes 14 years of clinical EEG data collected at the Temple University Hospital in Philadelphia and over 30,000 EEG signals collected in clinical settings. Such clinical-grade data is desirable because it is variable and heterogeneous regarding the location of EEG electrodes, clinical environment, equipment, and noise. This provides an advantage since models trained with such data are more likely to perform robustly.

Each EEG signal had either 19 to 31 data channels and was sampled variable at 250 Hz, 256 Hz, 400 Hz, and 512 Hz. Each data was paired with a clinician report, including history and medication.

The TUH EEG Epilepsy Corpus is one subset of TUH EEG. It contains EEG recordings from 100 healthy individuals and 100 patients with epilepsy, resulting in 1799 recordings collected over 570 sessions [7]. The subjects are between 17 to 88 years of age, and the patients were primarily recorded during the interictal period.

3.2 Data Pre-processing

The recordings were first re-sampled to 250 Hz. Furthermore, we applied a high pass filter at 1 Hz, and we used a notch filter at the power line frequency of 60 Hz. Channels that matched the 10-20 montage system were selected, resulting in 8 selected channels that corresponded to 4 pairs of electrodes, one from each hemisphere. Additionally, the recordings were divided into contiguous 10-second windows.

3.3 Feature Extraction

Features were extracted from the frequency domain using power spectral density (PSD) and spectral connectivity. PSD represents the power distribution of EEG series in the frequency domain and was calculated using the Welch method. In the current processing of data, the frequency information of EEG signals from PSD was divided into the following six brain wave bands: (1) delta (1-4Hz), (2) theta (4- 7.5Hz), (3) alpha (7.5-13Hz), (4) lower beta (13- 16Hz), (5) higher beta (16-30Hz), and (6) gamma (30-40Hz). The total power of each of the bands was found for each channel, resulting in an 8X6 matrix. Additionally, spectral connectivity analysis was also computed by calculating the coherence coefficient.

3.4 Random Forest Classifier

The pre-processed data were first shuffled, after which 67 percent of the data was used for training and 33 percent for testing in a random forest classifier algorithm. We used 100 trees and set the maximum depth of trees to 1000. We ran two random forest classifiers, one using the features from power spectral density and the other using the data obtained from spectral connectivity analysis.

4 RESULTS

We found the random forest classifier with coherence coefficients to have an accuracy of 90.87 percent, while the classifier using power spectral densities had an accuracy of 95.73 percent.

5 DISCUSSION AND CONCLUSION

Random forest classifiers provide high accuracy in detecting epilepsy using EEG signals when extensive data with heterogeneity is used for training. The accuracy is better when power spectral density is used than when spectral connectivity is used. However, the accuracy was less than that found by Mursalin et al. [5]. This is likely because of the lack of variability in the Bonn University data set. Additionally, it may be because they extracted features from time and frequency domains using entropy features and then extracted the most important subset of those features using improved correlation-based feature selection (ICSF).

Future research can focus on highlighting other performance indicators, such as AUC and F1-score instead of just accuracy. Moreover, the current research uses the random forest classifier on only one data set. Future research should also run the classifier on a different data set and compare the effectiveness of different classification methods.

Automated epilepsy detection using EEG signals has significant implications. It helps accurately identify the frequency of seizures, which patients often underestimate. The frequency of seizures is one of the most critical factors in deciding treatment methods. Additionally, if an epileptic seizure is detected right at the onset, then medication can be immediately administered and help control the seizure. It can also help prevent sudden unexpected death in epilepsy (SUDEP). Using wearable sensors, machine learning algorithms can help by alerting when a seizure occurs, especially during sleep, and preventing such occurrences [2].

REFERENCES

[1] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli. 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine* 100 (2018), 270–278.

[2] Christoph Baumgartner, Johannes P Koren, and Michaela Rothmayer. 2018. Automatic computer-based detection of epileptic seizures. *Frontiers in neurology* 9 (2018), 639.

[3] Shivam Chaudhary, Pankaj Pandey, Krishna Prasad Miyapuram, and Derek Lomas. 2022. Classifying EEG Signals of Mind-Wandering Across Different Styles of Meditation. In *International Conference on Brain Informatics*. Springer, 152–163.

[4] Cristian Donos, Matthias Dümpelmann, and Andreas Schulze-Bonhage. 2015. Early seizure detection algorithm based on intracranial EEG and random forest classification. *International journal of neural systems* 25, 05 (2015), 1550023.

[5] Md Mursalin, Yuan Zhang, Yuehui Chen, and Nitesh V Chawla. 2017. Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. *Neurocomputing* 241 (2017), 204–214.

[6] Shelagh JM Smith. 2005. EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry* 76, suppl 2 (2005), ii2–ii7.

[7] L Veloso, J McHugh, E von Weltin, S Lopez, I Obeid, and J Picone. 2017. Big data resources for EEGs: Enabling deep learning research. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 1–3.

[8] Neeraj Wagh and Yogatheesan Varatharajah. 2020. Eeg-gcnn: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network. In *Machine Learning for Health*. PMLR, 367–378.

[9] Difei Zeng, Kejie Huang, Cenglin Xu, Haibin Shen, and Zhong Chen. 2020. Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals. *IEEE transactions on cognitive and developmental systems* 13, 4 (2020), 955–968.