The Application of Machine Learning Algorithms for Epileptic Seizure Prediction and Detection using EEG Data

Neha Hulkund, Tesla STEM High School, Redmond, WA: s-nhulkund@lwsd.org

ABSTRACT
Epilepsy is a chronic brain disorder affecting 1% of the population worldwide. Its primary symptoms, seizures, occur without warning and can often be dangerous. Epilepsy is diagnosed with the use of the electroencephalogram (EEG), measuring the bursts of electrical activity associated with seizures. The objective of this study is to apply machine learning algorithms to predict seizures before they occur and diagnose epilepsy using EEG data. Preprocessing was done using bandpass filters and discrete wavelet transform for feature extraction of energy and entropy of the data on selective electrodes. The k-Nearest Neighbors classification algorithm was utilized to differentiate between preictal, ictal, and interictal segments of the data. It was able to detect seizures with a 99% accuracy, 93% sensitivity, and 95% specificity and predict seizures 3 minutes before they occured with a 96% accuracy, 90% sensitivity, and 91% specificity. With the application of the algorithm in medical wearable devices, seizure onset can be predicted, improving quality of life for epileptic patients.

1 INTRODUCTION

Epilepsy is a chronic brain disorder affecting the nervous system, prevalant all over the world. In 2009, the World Health Organization reported that 50 million people worldwide had the disease, affecting nearly 1% of the population [7]. In the United States, epilepsy is the third most common neurological disease, affecting over 3 million people [8]. Over 80% of this population live in third world countries, with a misdiagnosis rate of 23% [9]. Epilepsy may develop from traumatic brain injury, abnormal brain development, imbalances in neurotransmitters, and as a result of strokes, among other causes [11]. Its primary symptom is the occurence of epileptic seizures. Epilepsy is defined by two or more unprovoked seizures [7]. Seizures are classified by two types: focal and generalized. Focal seizures only occur in certain parts of the brain, either one hemisphere or part of a lobe, whereas generalized seizures affect the whole brain [8]. The primary form of treatment for epilepsy includes anticonvulsant drugs as medication.

However, in as many as 40% of patients who receive treatment, one or multiple drugs prove ineffective [9]. These untreatable seizures are known as medically intractable seizures and have significant negative impacts on the patients they affect [11]. Epilepsy is most commonly diagnosed with an electroencephalograph (EEG) test, an electrophysical method to monitor brain activity that detects changes in neural oscillations, indicating seizures and spams [5]. Generally, the diagnosis of epilepsy requires a clinician to evaluate the EEG seizure data after extensive days of monitoring, which can take up to 5 weeks to evaulate [4]. The machine learning methods detailed in the paper allow for the diagnosis of epilepsy to be made without a trained clinician, allowing for real time analysis of EEG signals predicting when a seizure is going to occur so patients or caregivers can be warned ahead of time, allowing for them to act before an attack.

2 MATERIALS

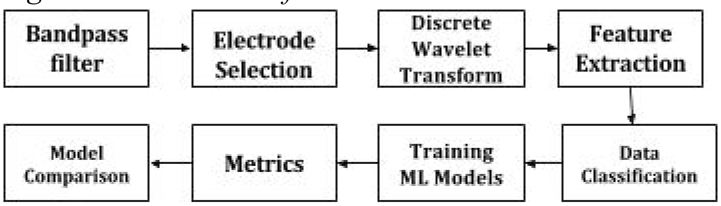
The dataset of epilepsy data was obtained from the open source CHB-MIT Scalp Database, recorded at the Children's Hospital Boston from pediatric patients [10]. EEG data was taken from 22 patients, containing between 9 and 24 files of hour long EEG data, some containing seizure data. There are a total of 173 events classified as seizures within the dataset. Signals were recorded at a sampling frequency of 256 Hz with a 16 bit resolution with 23 electrode channels per patient.

3 METHODS

It is helpful to split the research conducted into two portions: seizure detection and seizure prediction. Seizure detection is the detection of epileptic seizures while they are occurring within the dataset for the diagnosis of epilepsy and its severity based on the number of seizures per time interval [2]. Seizure prediction is the ability to predict beforehand when seizures are going to occur [3][6]. For example, seizure prediction should identify a patient that the seizure is about to occur 3 minutes before it does. Both types of research were conducted but with different methodologies, as further explained in this section.

3.1 Preprocessing

Figure 1: Flowchart of methods



Significant preprocessing is required on the EEG data to eliminate artifacts from the data before applying the machine learning algorithms. A bandpass filter of 0.5-40 Hz, the frequencies for seizure activity, was applied to reduce noise. Five electrodes were selected per patient, used across all of the patient's data to account for the localization of seizures. Discrete Wavelet Transform (DWT) [3], a modern signal analysis technique, was applied to each of the 5 electrode signals at decomposition level 5 to decompose the signal into different subbands by passing the signal through a series of high pass and low pass filters, generating approximation and detailed coefficients used later for feature extraction.

3.2 Feature Extraction

Using the coefficients from DWT, relative (Figure 2) and total energy (Figure 3) were calculated .

(1) The equation below represents the wavelet energy at each decomposition level i=1, 2, ...L, where L is the maximum level of decomposition. With this dataset, L=5, with 5 separate subbands of frequency. E_A is used to calculate the energy using the approximation coefficients and E_D is used to calculate the energy using detailed coefficients.

$$E_{D_i} = \sum_{j=1}^{N} |D_{ij}|^2, \quad i = 1, 2, 3, \ldots, L$$

$$E_{A_i} = \sum_{j=1}^{N} |A_{ij}|^2, \quad i = L$$

(2) *Figure 2: Total energy*

$$E_{Total} = \left(\sum_{i=1}^{L} E_{D_i} + E_{A_L} \right)$$

(3) *Figure 3: Relative Energy*

$$E_r = \frac{E_j}{E_{Total}}$$

$$\text{where } E_j = E_{D_{i=1 \ldots L}} \text{ or } E_{A_{i=L}}$$

3.6 Data Classification

For seizure prediction, data was classified into 3 classes: interictal (no seizure), preictal (before seizure), and ictal (seizure), labeled 0, 1, 2, respectively [6]. The preictal period was labeled in different intervals of 30 seconds, 1 minute, 2 minutes, and 3 minutes.

For seizure detection, data was classified into 2 classes: interictal and ictal, labeled 0 and 1, respectively.

3.7 KNN Algorithm

The k-Nearest Neighbors algorithm is a supervised classifier, used with k=1 on a split of 20% training and 80% test data. A 5 fold cross-validation was used to protect against the overfitting. A binary kNN classifier was used for seizure detection and a multi class kNN classifier used for seizure prediction.

4 RESULTS

The performance of the seizure prediction algorithms was measured by three metrics: sensitivity, specificity, and total classification accuracy.

- Sensitivity is calculated by the total number of correctly classified positive patterns.
- Specificity is measured by the number of correctly classified negative patterns. [1]
- Total classification accuracy is calculated by the total number of correctly identified patterns.

4.1 Seizure Detection

Using the kNN algorithm, the following metrics were calculated for the binary classification of EEG seizures. This shows seizure detection, or epilepsy diagnosis.

Patient #	Accuracy	Specificity	Sensitivity
1	98.5%	93.5%	92.3%
2	99.1%	96.7%	91.2%
3	97.5%	95.4%	90.5%
4	95.2%	92.9%	92.3%
5	99.0%	96.3%	94.3%
6	99.1%	97.3%	94.5%
7	98.2%	92.1%	92.3%
8	98.7%	95.2%	93.7%
9	100.0%	97.8%	94.5%
10	99.1%	96.7%	93.6%
11	100.0%	95.7%	93.5%
12	98.5%	93.4%	93.5%
13	99.0%	95.6%	93.5%
14	98.0%	95.7%	93.4%
15	99.0%	96.5%	94.6%
Average	**98.6%**	**95.4%**	**93.2%**

4.2 Seizure Prediction

A multi class kNN was used for seizure prediction before it occurred, using different intervals of the preictal period.

Metric	30 sec	1 min	2 min	3 min
Accuracy	99.2%	98.5%	98%	**96.4%**
Specificity	95.6%	93.9%	93.7%	**93.2%**
Sensitivity	92.1%	91.2%	90.1%	**90.4%**

5 DISCUSSION

For seizure detection, the k-Nearest Neighbors classification algorithm was utilized to differentiate between ictal and interictal segments of the data. It was able to detect seizures with a 99% accuracy, 95% specificity, and 93% sensitivity. This means that it will be able to detect 9 out of 10 seizures, with a false positive rate of 10%. This is a significant increase to the accuracy of doctor's diagnosis at 77%. This allows for the diagnosis of epilepsy to be made computationally in a doctor's office without a trained clinician available. This will decrease the rates of misdiagnosis if these algorithms are implemented upon a clinical EEG. However, these metrics can be further improved with further preprocessing of the data through Independent Component Analysis (ICA) or further feature extraction of entropy of the signal.

For seizure detection,the algorithm is able to predict seizures 3 minutes before they occurred with a 96% accuracy, 93% specificity, and 90% sensitivity before a significant decline in quality of prediction. This is a great increase in prediction in comparison to previous latencies of 15 to 30 seconds. The increased warning time will allow for patients or caregivers to call for help or respond appropriately to a warning in a dangerous situation. With the application of the algorithm in medical wearable devices, seizure onset can be predicted, improving quality of life for epileptic patients.

6 CONCLUSIONS

Because of the high accuracy of prediction, the utilization of machine learning algorithms for epilepsy diagnosis and prediction is able to significantly change the way healthcare is administered for the disease. If implemented at a large scale, the cost of the diagnosis could be significantly reduced from the current $3000 required for extended monitoring. Due to the the high accuracy, specificity, and sensitivity, the algorithms implemented in a device can be used for computational diagnosis of epilepsy without a trained clinician, having the potential to decrease the rates of undiagnosis of the disease.

The seizure detection algorithm can also be used to detect the number of seizures occurring per hour for accurate prescription of anti-seizure drugs in order to avoid underdiagnosis or overdiagnosis. Furthermore, the device can predict when seizures are about to occur with a prediction time of 3 minutes, allowing for patients or caregivers to react appropriately. This has the potential to reduce the impact of epileptic attacks by proving warning time to avoid fatal accidents or receive the appropriate treatment, overall improving the patient's' quality of life.

REFERENCES

1. Cunha, J. P. S., Choupina, H. M. P., Rocha, A. P., Fernandes, J. M., Achilles, F., Loesch, A. M., … Noachtar, S. (2016). NeuroKinect: A novel low-cost 3Dvideo-EEG system for epileptic seizure motion quantification. *PLOS ONE*, *11*(1), e0145669. doi:10.1371/journal.pone.0145669
2. Debener, S., Minow, F., Emkes, R., Gandras, K., & Vos, de (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology.*, *49*(11), 1617–21.
3. Müller, K., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2007). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of neuroscience methods.*, *167*(1), 82–90.
4. Noachtar, S., & Rémi, J. (2009). The role of EEG in epilepsy: A critical review. *Epilepsy & behavior : E&B.*, *15*(1), 22–33.
5. Schiff, S., Casa, M., Caro, D., Aprile, D., Spinelli, G., Rui, D., … Montagnese, S. (2016). A low-cost, user-friendly electroencephalographic recording system for the assessment of hepatic encephalopathy. *Hepatology (Baltimore, Md.).*, *63*(5), 1651–9.
6. Gajic, D., Djurovic, Z., Di Gennaro, S., & Gustafsson, F. (2014). Classification of EEG Signals for Detection of Eplieptic Seizures Based on Wavelets and Statistical Pattern Recognition. *Biomedical Engineering: Applications, Basis and Communications*, *26*(02), 1450021. doi:10.4015/s1016237214500215
7. Epilepsy. (n.d.). Retrieved July 15, 2017, from http://www.who.int/mediacentre/factsheets/fs999/en/
8. Seizure types. Retrieved February 27, 2017, from Epilepsy Society.
9. Ferrie, C. D. (2006). Preventing misdiagnosis of epilepsy. *Archives of Disease in Childhood*, 91(3), 206–209.
10. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; 2000 (June 13).