Project Proposal
Real-time preictal detection through the application of machine learning to Electroencephalogram signals.

William Riddell

October 18, 2023

Contents

1 Introduction

Over the last 20 years, Artificial Intelligence (AI) has seen a large evolution through the use of Machine Learning (ML); the statistical analysis of data which leads to the unveiling of characteristics and connections. (Awad & Khanna 2015). There has been a large uptake of applying ML techniques to biomedical data, increasing the speed and accuracy of prediction, detection, diagnosis, and prognosis.

Electroencephalograms (EEGs) measure the electrical signals in the brain. EEGs have a great use in giving an insight into the inner workings of the brain, for example allowing us to pick up abnormalities preceding and during their occurrence. "A seizure is a burst of uncontrolled electrical activity between brain cells (also called neurons or nerve cells) that causes temporary abnormalities in muscle tone or movements (stiffness, twitching or limpness), behaviours, sensations or states of awareness." (Medicine n.d.) Due to this, monitoring the brain's electrical activity through the use of an EEG, and applying analysis through an ML model may allow us to detect the preictal period. "An automated accurate prediction of seizures will significantly improve the quality of life of patients and reduce the burden on caregivers" (Acharya, Hagiwara & Adeli 2018)

This project will aim to develop an ML model which triggers an alert if a preictal period is detected. The model will have to achieve a high degree of accuracy ($\geq 90\%$) when being applied to real-time EEG data. Throughout this project I will compare previous attempts using different ML models, and I will evaluate the different datasets available for preictal prediction.

2 Background Review

(Wong, Simmons, Rivera-Villicana, Barnett, Sivathamboo, Perucca, Ge, Kwan, Kuhlmann, Vasa et al. 2023) reviews 10 datasets available to download. It evaluates the way the EEGs were physically setup on the subject, the subjects themselves and the data's properties. Wong et al. also states their opinion on what tasks suit what dataset, with the main two tasks being either detection and prediction.

Dataset
University of Bonn
CHB-MIT Scalp EEG
Melbourne-NeuroVista seizure trial (Neurovista Ictal)
Kaggle UPenn and Mayo Clinic's Seizure Detection Challenge
Neurology and Sleep Centre Hauz Khas
Kaggle American Epilepsy Society Seizure Prediction Challenge
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge
TUH EEG Seizure Corpus (TUSZ)
Siena Scalp EEG
Helsinki University Hospital EEG

Table 1: 10 Datasets within (Wong et al. 2023)

Within these datasets Wong et al. was also able to find the way the EEG nodes were positioned on the subject's cranium, along with whether the EEG nodes were either placed intracranial or extracranial. Wong et al. also the number of channels that are contained in the raw EEG data for each dataset.

Dataset	Number of channels	Placement method	Type of signal
University of Bonn	1	International 10–20 system, Intracranial	Scalp/Intracranial EEG
CHB-MIT Scalp EEG	18	International 10–20 system/Nomenclature	Scalp EEG
Melbourne-NeuroVista seizure trial (NeuroVista Ictal)	16	Intracranial	Intracranial EEG
Kaggle UPenn and Mayo Clinic's Seizure Detection Challenge	16–76	Intracranial	Intracranial EEG
Kaggle American Epilepsy Society Seizure Prediction Challenge	16	Intracranial	Intracranial EEG
Neurology and Sleep Centre Hauz Khas	1	International 10–20 System	Scalp EEG
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	16	Intracranial	Intracranial EEG
TUH EEG Seizure Corpus (TUSZ)	23–31	International 10–20 system / Nomenclature	Scalp EEG
Helsinki University Hospital EEG	19	International 10–20 system	Scalp EEG
Siena Scalp EEG	20/29	International 10–20 system/Nomenclature	Scalp EEG

Table 2: Channel Characteristics (Wong et al. 2023)

Wong et al. also noted along with this data that the "University of Bonn dataset contains a mixture of both scalp and intracranial EEG data where scalp EEG from healthy subjects was taken, while intracranial EEG was taken from subjects with a history of seizures." (Wong et al. 2023). This may present a skew on the ML model during training.

Dataset	Noncontinuous data	Short-term continuous data	Continuous data
University of Bonn	Yes	No	No
CHB-MIT Scalp EEG	No	Yes	Yes
Melbourne-NeuroVista seizure trial (Neurovista Ictal)	N/A	N/A	N/A
Kaggle UPenn and Mayo Clinic's Seizure Detection Challenge	Yes	No	No
Kaggle American Epilepsy Society Seizure Prediction Challenge	Yes	No	No
Neurology and Sleep Centre Hauz Khas	Yes	No	No
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	Yes	No	No
TUH EEG Seizure Corpus (TUSZ)	No	Yes	No
Helsinki University Hospital EEG	No	Yes	No
Siena Scalp EEG	No	Yes	No

Table 3: Temporal properties (Wong et al. 2023)

Wong et al. ordered the datasets into groups, either continuous or non continuous data. For the continuous data they separated out datasets which record for less that 24 hours in a single go, these were classified as "Short-term continuous" data.

Dataset	Number of subjects	Subject type
University of Bonn	10	Human
CHB-MIT Scalp EEG	23	Human
Melbourne-NeuroVista seizure trial (NeuroVista Ictal)	12	Human
Kaggle UPenn and Mayo Clinic's Seizure Detection Challenge	12	Human & Canine
Kaggle American Epilepsy Society Seizure Prediction Challenge	7	Human & Canine
Neurology and Sleep Centre Hauz Khas	10	Human
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	3	Human
TUH EEG Seizure Corpus (TUSZ)	642	Human
Helsinki University Hospital EEG	79	Human
Siena Scalp EEG	14	Human

Table 4: Subject properties (Wong et al. 2023)

Wong et al. also was able to identify the number of subjects within each dataset. Within the two "Kaggle" datasets there are Canine subjects, making them unsuitable for this project.

Within the review, they also produced tables displaying the segment information for each dataset, breaking down the recording length and frequency, along with the number of events and segments. This information should not weight into which dataset suits the idea of preictal prediction so shall be left out in this background review. Wong et al. also discussed the idea of the class imbalance problem, where the number and length of each ictal period is unbalanced. Two datasets, "University of Bonn" and the "Neurology and Sleep Centre Hauz Khas" have addressed this issue and have balanced their data between ictal, preictal, interictal and nonictal periods.

Taking the research into account Wong et al. suggested which dataset suits either prediction or detection. “Since the aim of seizure prediction is to forecast impending seizures, EEG recordings that include preictal and interictal data should be used for the study, while the aim of seizure detection is to detect ongoing seizure events, hence, EEG recordings that contain ictal and interictal data should be used.” (Wong et al. 2023).

Dataset	**Application**
University of Bonn	Seizure detection
CHB-MIT Scalp EEG	Seizure detection/Prediction
Melbourne-NeuroVista seizure trial (NeuroVista Ictal)	Seizure detection/Prediction
Kaggle UPenn and Mayo Clinic’s Seizure Detection Challenge	Seizure detection
Kaggle American Epilepsy Society Seizure Prediction Challenge	Seizure prediction
Neurology and Sleep Centre Hauz Khas	Seizure detection/Prediction
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	Seizure prediction
TUH EEG Seizure Corpus (TUSZ)	Seizure detection/Prediction
Helsinki University Hospital EEG	Seizure detection/Prediction
Siena Scalp EEG	Seizure detection/Predictio

Table 5: Suggested applications (Wong et al. 2023)

There are also various ML approaches which have varying success. (Natu, Bachute, Gite, Kotecha, Vidyarthi et al. 2022) reviews 7 different approaches, along with varying preprocessing steps. Natu et al. begins by reviewing (Parvez & Paul 2015), and comments on their use of a least square based SVM reliably yields a high accuracy. (Parvez & Paul 2015) Also used their SVM against many datasets, showing the robustness of their trained ML model. Parvez et al. also stated the usefulness of EMD within preprocessing. (Lin, Wang, Wu, Jeng & Chen 2009) also used a SVM, although they did not use the same preprocessing steps that Parvez et al. used. Lin et al. was able to achieve a 82.29% accuracy.

Ilyas et al. (Ilyas, Saad, Ahmad & Ghani 2016) was able to achieve a 73.09% and a 68.97% accuracy for their SVM and Logistical Regression (LR) respectively. It is also important to note that Ilyas et al. also utilized many different classifiers in their paper, although their performance was lower than both the SVM and LR approach. The other classifiers used were K Nearest Neighbor (kNN), and an Multi-layer Perceptron Artificial Neural Network (MP-ANN).

(Guo, Rivero, Dorado, Rabunal & Pazos 2010) however, was able to achieve a 97.99% accuracy with their MP-ANN. Guo's et al. developed a three step process, making use of methods such as Discrete Wavelet Transformation (DWT).

3 Methodology

I will be choosing a dataset which is applicable to preictal prediction, it will have to contain all ictal periods. I will then have to research and settle on a preprocessing and then feature extraction method, and implement these in Python3.

After preprocessing and feature extraction has been researched, along side development I can research ML models, comparing their performance for the task. Once a model has been decided on I can likewise implement it in Python3.

I will most likely have to be utilizing library's such as TensorFlow (Abadi, Agarwal & et al. 2015) and Numpy (Harris, Millman & van der Walt et al. 2020). Once the model is trained I need to verify its accuracy and sensitivity.

I will be version managing this project using the software "Git" , and hosting the repository on "Github", and I will be using my own hardware, a Nvidia GeForce GTX 1080, to train the SVM. For testing purposes I will be using the libraray PyTest to unit test my functions, and as mentioned above I will be using LOOCV to test my ML model.

4 Project management

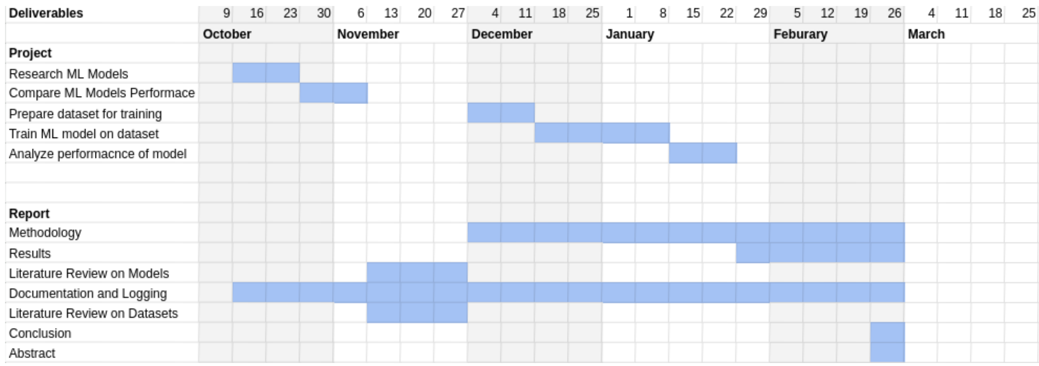
4.1 Activities

1. Data Collection and Preprocessing:
    - Download various applicable datasets.
    - Research available preprocessing methods.

- Create preprocessing pipeline in Python3

2. Feature Extraction:
   - Evaluate other papers to decide on which features to extract.
   - Build my own process in Python3 to extract feature from preprocessed data.
3. Model Development:
   - Research and settle on a final ML model to use.
   - Train and tune the ML model against my selected dataset(s).
4. Testing and Evaluation:
   - Evaluate the ML model using performance metrics.
   - Validate the prediction accuracy and sensitivity with a test dataset.
5. Documentation:
   - Maintain project logs, reports, and literature references within the Git repository.
   - Document the entire process throughout the development.
6. Project Reporting:
   - Create a final project report summarizing the methodology, results, and conclusions.
   - Implement any additional information, such as documentation and sources.

4.2 Schedule



4.3 Deliverables

1. Project

   - Compare the performance of different ML models.
   - Produce an ML model which can accurately predict preictal periods on EEG data.
   - Organized data, including raw data and the preprocessed data, including the relevant literature.

2. Report

   - My methodology, the results, and a conclusion.
   - Logs documenting activities, tasks, and changes throughout the project.
   - A literature review of the subject area, including a review on the different datasets.

Acronyms

AI Artificial Intelligence. 3

DWT Discrete Wavelet Transformation. 9

EEG Electroencephalogram. 3

EEGs Electroencephalograms. 3

EMD Empirical Mode Decomposition. 8

kNN K Nearest Neighbor. 9

LOOCV Leaving One Out Cross-Validation. 9

LR Logistical Regression. 8, 9

ML Machine Learning. 3, 6, 8–11

MP-ANN Multi-layer Perceptron Artificial Neural Network. 9

SVM Support Vector Machine. 8, 9

References

Abadi, M., Agarwal, A. & et al., P. B. (2015), 'TensorFlow: Large-scale machine learning on heterogeneous systems'. Software available from tensorflow.org.
URL: *https://www.tensorflow.org/*

Acharya, U. R., Hagiwara, Y. & Adeli, H. (2018), 'Automated seizure prediction', *Epilepsy & Behavior* **88**, 251–261.

Awad, M. & Khanna, R. (2015), *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Apress.
URL: *https://books.google.co.uk/books?id=qPGQnAEACAAJ*

Guo, L., Rivero, D., Dorado, J., Rabunal, J. R. & Pazos, A. (2010), 'Automatic epileptic seizure detection in eegs based on line length feature and artificial neural networks', *Journal of neuroscience methods* **191**(1), 101–109.

Harris, C. R., Millman, K. J. & van der Walt et al., S. J. (2020), 'Array programming with NumPy', *Nature* **585**(7825), 357–362.
URL: *https://doi.org/10.1038/s41586-020-2649-2*

Ilyas, M., Saad, P., Ahmad, M. & Ghani, A. (2016), Classification of eeg signals for brain-computer interface applications: Performance comparison, *in* '2016 International Conference on Robotics, Automation and Sciences (ICORAS)', IEEE.

Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K. & Chen, J.-H. (2009), Eeg-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine, *in* '2009 IEEE international conference on acoustics, speech and signal processing', IEEE, pp. 489–492.

Medicine, J. H. (n.d.), 'Types of seizures, what is a seizure?'.
URL: *https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/types-of-seizures*

Natu, M., Bachute, M., Gite, S., Kotecha, K., Vidyarthi, A. et al. (2022), 'Review on epileptic seizure prediction: machine learning and deep learning approaches', *Computational and Mathematical Methods in Medicine* **2022**.

Parvez, M. Z. & Paul, M. (2015), 'Epileptic seizure detection by exploiting temporal correlation of electroencephalogram signals', *IET Signal Processing* **9**(6), 467–475.

Wong, S., Simmons, A., Rivera-Villicana, J., Barnett, S., Sivathamboo, S., Perucca, P., Ge, Z., Kwan, P., Kuhlmann, L., Vasa, R. et al. (2023), 'Eeg datasets for seizure detection and prediction—a review', *Epilepsia Open* .