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

William Riddell

October 16, 2023

Comments for Kashi will be written in italics. - William

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 ($\geq$ 85%) when being applied to real-time EEG data. It will make use of a Support Vector Machine (SVM) to classify the preictal periods from interictal and ictal periods, making use of the CHB-MIT Dataset, which contains extracranial EEG signals from 22 patients from the Boston Children's Hospital. (Shoeb 2009) (Goldberger, Amaral, Glass, Hausdorff, Ivanov, Mark, Mietus, Moody, Peng & Stanley 2000 (June 13))

2 Background Review

There has been extensive research into applying ML models to EEG data; (Shoeb & Guttag 2010) (Chakraborti, Choudhary, Singh, Kumar & Swetapadma 2018)(Kumar & Kolekar 2014) (Shen, Wen, Song & Li 2022) (Gupta & Pachori 2019) (Samiee, Kiranyaz, Gabbouj & Saramäki 2015) (Zabihi, Kiranyaz, Rad, Katsaggelos, Gabbouj & Ince 2015) (Wang, Wang, Liu, Chang, Kärkkäinen & Cong 2021) (Zarei & Asl 2021) (Li, Zhou, Liu, Zhang, Geng, Liu, Wang & Shang 2021) (Shoeb 2009) (Wong, Simmons, Rivera-Villicana, Barnett, Sivathamboo, Perucca, Ge, Kwan, Kuhlmann, Vasa et al. 2023). Although there have been fewer attempts to apply ML models to identify the preictal state in a real-time setting with high accuracy, although they do exist (Usman, Usman, Fong et al. 2017).

Dataset CHB-MIT (Shoeb 2009) (Goldberger et al. 2000 (June 13)) was taken from Boston Children's Hospital. It is a long-term dataset with recordings of 22 paediatric subjects, 5 male and 17 female, who have intractable seizures. The dataset contains 23 EEG signals positioned on the scalp in the international 10-20 EEG system (Sharbrough 1991). The dataset is labelled with the timestamps of the onset and ending of each seizure. Over the course of several days the onsets and ends of a total of 182 seizures were annotated onto the long-term EEG data (Shoeb 2009) (Goldberger et al. 2000 (June 13)). This dataset is utilized by many papers when working on real-time epileptic seizure detection, there have been many different approaches with detection accuracy reaching 92.23% (Usman et al. 2017)

I will be expanding on this section, adding maybe 2 or 3 more papers looking at predicting seizures (and updating the accuracy figure above, because higher accuracies have been achieved).

Usman et al. was able to achieve a preictal detection rate of 92.23% against the CHB-MIT dataset. (Usman et al. 2017) Usman noted that poor Signal-to-Noise Ratio (SNR) within the dataset lead to a large amount of false positives and posed as a issue during his research. To combat this he combined the many EEG signals into a single surrogate signal and then utilized Empirical Mode Decomposition (EMD) to increase the SNR. Usman et al. then broke the surrogate signal down and "extracted multiple features including entropy, approximate entropy, and Hjorth parameters." In Usman's et al. papers, it was observed that both statistical and spectral features played a part in detecting the differences between interictal and preictal states so he also extracted spectral moments, and statistical moments to aid with his classification. He finally utilized a Support Vector Machine (SVM) taking in a continuous EEG signal broken into non-overlapping one second sections and classifying if it was part of a preictal period (Usman et al. 2017).

When a seizure occurs, or during the preictal state, the spike rate and variation in the EEG signals change (Lange, Lieb, Engel Jr & Crandall 1983) (Truccolo, Donoghue, Hochberg, Eskandar, Madsen, Anderson, Brown, Halgren & Cash 2011) "spike rate is used as the indicator to anticipate seizures in Electroencephalogram (EEG) signal. Spikes detection step is used in EEG signal during interictal, preictal, and ictal periods followed by a mean filter to smooth the spike number. The maximum spike rate in interictal periods is used as an indicator to predict seizures." (Slimen, Boubchir & Seddik 2020). Using this method, Slimen et al. was able to achieve a 92% accuracy. *I need to link this back with the ML part and expand on this paper more.*

3 Methodology

3.1 Approach (Description of the research and development methodology, e.g. Software development model, requirement gathering method, test and evaluation process)

I will be utilizing the open-source CHB-MIT dataset (Shoeb 2009) (Goldberger et al. 2000 (June 13)) as it contains all ictal periods which is ideal for the prediction of epileptic seizures. I will be following the preprocessing stages that Usman et al. used during their research. This involves converting the

EEG signals into a single signal through the use of applying an averaging filter and the Common Spatial Pattern (CSP) algorithm to increase the SNR. Once I have applied these algorithms I will perform EMD which breaks the surrogate signal into its oscillatory functions, which are amplitude, period, and frequency. This is also known as its Intrinsic Mode Functions (IMF).

Usman et al. also state that the noise generally affects the high-frequency components (Usman et al. 2017), due to this I will select the suitable IMF frequencies and extract them for classification. I will finally be using a SVM to classify the data between interictal and preictal states.

I will be programming the preprocessing and the ML model in Python 3 (Van Rossum & Drake 2009), using librarys such as TensorFlow (Abadi, Agarwal, Barham, Brevdo, Chen, Citro, Corrado, Davis, Dean, Devin, Ghemawat, Goodfellow, Harp, Irving, Isard, Jia, Jozefowicz, Kaiser, Kudlur, Levenberg, Mané, Monga, Moore, Murray, Olah, Schuster, Shlens, Steiner, Sutskever, Talwar, Tucker, Vanhoucke, Vasudevan, Viégas, Vinyals, Warden, Wattenberg, Wicke, Yu & Zheng 2015) which I will be using to construct my SVM and Numpy (Harris, Millman, van der Walt, Gommers, Virtanen, Cournapeau, Wieser, Taylor, Berg, Smith, Kern, Picus, Hoyer, van Kerkwijk, Brett, Haldane, del Río, Wiebe, Peterson, Gérard-Marchant, Sheppard, Reddy, Weckesser, Abbasi, Gohlke & Oliphant 2020) to write in my averaging filter and the CSP algorithm. I will also be using the "EMD-signal" library (Laszuk 2017) to split the surrogate signal into its IMF. I will be version managing this project using the software "Git" , and hosting the repository on "Github".

For testing purposes *write in testing methodology, including suitable cross-validation for SVM (required research)*

4 Project management

4.1 Activities: tasks required to complete each objective

asdf

4.2 Schedule i.e. Gantt or other, showing activities, deadlines

asdf

4.3 Data management plan (e.g. Google folder for project logs, reports, literature etc)

asdf

4.4 Deliverables

asdf

Acronyms

AI Artificial Intelligence. 1

CSP Common Spatial Pattern. 4

EEG Electroencephalogram. 1–4

EEGs Electroencephalograms. 1

EMD Empirical Mode Decomposition. 3, 4

IMF Intrinsic Mode Functions. 4

ML Machine Learning. 1, 2

SNR Signal-to-Noise Ratio. 3, 4

SVM Support Vector Machine. 2–4

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. (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*

Chakraborti, S., Choudhary, A., Singh, A., Kumar, R. & Swetapadma, A. (2018), 'A machine learning based method to detect epilepsy', *International Journal of Information Technology* **10**, 257–263.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K. & Stanley, H. E. (2000 (June 13)), 'PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals', *Circulation* **101**(23), e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

Gupta, V. & Pachori, R. B. (2019), 'Epileptic seizure identification using entropy of fbse based eeg rhythms', *Biomedical Signal Processing and Control* **53**, 101569.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A.,

del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. & Oliphant, T. E. (2020), 'Array programming with NumPy', *Nature* **585**(7825), 357–362.
URL: *https://doi.org/10.1038/s41586-020-2649-2*

Kumar, A. & Kolekar, M. H. (2014), Machine learning approach for epileptic seizure detection using wavelet analysis of eeg signals, *in* '2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)', IEEE, pp. 412–416.

Lange, H. H., Lieb, J. P., Engel Jr, J. & Crandall, P. H. (1983), 'Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy', *Electroencephalography and clinical neurophysiology* **56**(6), 543–555.

Laszuk, D. (2017), 'Python implementation of empirical mode decomposition algorithm', https://github.com/laszukdawid/PyEMD.

Li, C., Zhou, W., Liu, G., Zhang, Y., Geng, M., Liu, Z., Wang, S. & Shang, W. (2021), 'Seizure onset detection using empirical mode decomposition and common spatial pattern', *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 458–467.

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*

Samiee, K., Kiranyaz, S., Gabbouj, M. & Saramäki, T. (2015), 'Long-term epileptic eeg classification via 2d mapping and textural features', *Expert Systems with Applications* **42**(20), 7175–7185.

Sharbrough, F. (1991), 'American electroencephalographic society guidelines for standard electrode position nomenclature', *Clin Neurophysiol* **8**, 200–202.

Shen, M., Wen, P., Song, B. & Li, Y. (2022), 'An eeg based real-time epilepsy seizure detection approach using discrete wavelet transform and machine learning methods', *Biomedical Signal Processing and Control* **77**, 103820.

Shoeb, A. H. (2009), Application of machine learning to epileptic seizure onset detection and treatment, PhD thesis, Massachusetts Institute of Technology.

Shoeb, A. H. & Guttag, J. V. (2010), Application of machine learning to epileptic seizure detection, *in* 'Proceedings of the 27th international conference on machine learning (ICML-10)', pp. 975–982.

Slimen, I. B., Boubchir, L. & Seddik, H. (2020), 'Epileptic seizure prediction based on eeg spikes detection of ictal-preictal states', *Journal of biomedical research* **34**(3), 162.

Truccolo, W., Donoghue, J. A., Hochberg, L. R., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Brown, E. N., Halgren, E. & Cash, S. S. (2011), 'Single-neuron dynamics in human focal epilepsy', *Nature neuroscience* **14**(5), 635–641.

Usman, S. M., Usman, M., Fong, S. et al. (2017), 'Epileptic seizures prediction using machine learning methods', *Computational and mathematical methods in medicine* **2017**.

Van Rossum, G. & Drake, F. L. (2009), *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA.

Wang, X., Wang, X., Liu, W., Chang, Z., Kärkkäinen, T. & Cong, F. (2021), 'One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial eeg', *Neurocomputing* **459**, 212–222.

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* .

Zabihi, M., Kiranyaz, S., Rad, A. B., Katsaggelos, A. K., Gabbouj, M. & Ince, T. (2015), 'Analysis of high-dimensional phase space via poincaré section for patient-specific seizure detection', *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24**(3), 386–398.

Zarei, A. & Asl, B. M. (2021), 'Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy based features of eeg signals', *Computers in Biology and Medicine* **131**, 104250.