Correlation Between Gamma Event Coupling and Seizures following Traumatic Brain Injury

Sam Vander Dussen

Traumatic Brain Injury (an injury due to trauma to the brain from an outside force), Epilepsy (condition of repeated seizures over time), and Post-Traumatic Epilepsy (condition where repeated seizures occur more than one week post traumatic brain injury) are some of the most outstanding neurological conditions in the United States of America [1]. Every nine seconds, someone in the USA has a traumatic brain injury, over three million people have epilepsy, and one out of every ten people who are hospitalized will have post traumatic epilepsy [1]. The prevalence of these conditions has caused a large amount of research in these areas. One of these research projects is the Epibios study, which is a large collaboration by multiple universities and institutes to look for biomarkers and interventions for each of these conditions.

Dr. Richard Staba and Dr. Neil Harris are a part of and assisting with this research project and allowed me the opportunity to work with them in the study. They each gave me a couple of papers to read through and helped me identify an idea to research. Previous work had been published that showed the connection of gamma event coupling and functional connectivity of the brain [2] and how the brain functional connectivity changes during epileptogenesis (the onset of epilepsy) [3]. Therefore Dr. Staba had set up experiments to look at gamma event coupling following traumatic brain injury (TBI) and post traumatic epilepsy (PTE). This work included experiments that took electroencephalogram (EEG) recordings of three groups of rats (control group, traumatic brain injury without post traumatic epilepsy, and traumatic brain injury with post traumatic epilepsy. They allowed me to review this work as well and formulate my own hypothesis to test.

There were and are many ways to look at gamma event coupling in the experiment data provided to me. One could look at the total amount of gamma event coupling across groups and compare these values. The spatial location and amount of gamma event coupling may be an important biomarker. The difference in gamma event coupling following traumatic brain injury may be able to determine which rats develop PTE and which do not. This data set provided a large a rich possibility of research. Upon talking with Dr. Staba and Dr. Harris, indicated that there has yet to be much in-depth research into the temporal linkage of gamma event coupling and any of these conditions. Therefore, this paper tries to find a way to classify seizure (ictal) and nominal states of the brain, based on the time series data (EEG data,) taken from a rat in the TBI group that has been confirmed with PTE. The hypothesis is that gamma events will be indicative of an ictal period. By beginning with this investigation of the data it will allow for either the continuation of temporal investigation of gamma event coupling between groups in the study or begin investigating a different path.

To create a way to classify brain states of seizure and non-seizure using gamma event coupling a support vector machine will be used to classify the labeled data. A support vector machine is a supervised machine learning algorithm that maximizes the margin between two labelled classes of data. The margin is defined as a distance metric based on (between) the pair of parallel lines made by the three closest points between the classes (these points are known as the support vectors). The entire support vector machine can be described using these three points. The support vector machine is not contained only to cartesian space and euclidean (2D) space distance metrics either. Non-linear support vector machine classifiers can be created by defining a transformation to take the input space to a different feature space that may be of higher dimensions. This makes the linear operation of defining lines in the feature space equivalent to non-linear operations in the input space and reduces computation costs of creating the support vectors. By using a kernel function and the corresponding eigenfunctions to define the transformation it is easy to create and comprehend the transformation used.

Sklearn’s SVC software package was used to create the support vector machine in the code used during this investigation. The creation of a support vector machine using sklearn’s SVC package is done by creating a model using the function with parameters of the various options, which are mostly pre-set and be kept that way. A few that are important are the kernel that is desired (list of defined kernels provided by sklearn), the regularization parameter, and probability (which allows for estimates using this model). After the model is defined, it is fit to the input data matrix and the labelled class vector defining the state of each output of the observations. This will allow the user to then use the provided functions in the software package to find the scores, decisions made by the model, predictions, support vectors, and various other information about the model and of use in assessing its performance.

Another machine learning algorithm used for investigative purposes is principal component analysis. Principal component analysis (PCA) is an algorithm for reducing the number of variables in a model, thereby reducing the dimensions of the model. It does this by looking for commonalities of the variables using the direction of the maximum variance explained as a base component and repeating the process of finding the direction of the next maximum amount of variance explained (always orthogonal to the previous principal components). This method is used for finding commonalities in the observation space or variable (feature) space and visualization of the data for this project. The sklearn decomposition PCA package is used and works similarly to the above SVC package. First a PCA instance is made indicating the number of principal components desired to be in the model and then the model is fit to the data, however in this case it is only fit to the input data matrix.

By using the above methods, a classification was able to be made that separated the two classes with a cross validated score of 0.958 across six folds and the area under the curve of 1. The mean area of the curve over all folds is equal to 0.72 +/- 0.4. The results of this investigation are still preliminary and only performed on one test subject; however, they have great promise. The data upon inspection does not seem to have a highly defined temporal linkage that is easy to pick out using basic techniques like plotting average values against time or histograms of the GEC values. However, by applying the techniques above, the GEC values can indicate whether the rat is in a state of seizure vs. non seizure, this is a substantiating result for the hypothesis of this paper. The confusion matrix and ROC plots have great visual representations of how well this model performs.

Confusion Matrix: 

ROC Plot:

Chart, line chart

Description automatically generated

The confusion matrix above shows how the model predicts all the true positives and true negatives perfectly and has no false values. The receiver operating characteristic plot shows how once the model is considering enough data (in the folds) it is almost able to predict the behavior perfectly. PCA was used to reduce the dimensions of the data to two so that the decision line, support vectors, and margin can be shown visually. The loadings plot was also plotted to see which variables explain the most about the variance in the data.

Chart, scatter chart

Description automatically generated

Above: Scores plot with the SVC applied to it. Below: loadings plot Chart, scatter chart

Description automatically generated

The scores plot is great at visualizing how the radial basis function kernel encircles the non-seizure data. The loading plot shows how coupling between channels 1 and 2, where the seizure was identified, and channels 5 and 6, cortex to hippocampal area (most theta activity) are the most explanatory variables. The PCA model did not perform better than the model with all the data, therefore it is mainly for visualization purposes that it is provided. Various kernels and cross validation iterations were performed on the model, and the model with the radial basis function kernel and stratified cross folds performed the best under cross validation with the metrics provided in the previous paragraph. These preliminary results tentatively signify a correlation between GEC events and seizures and future work is needed to provide greater evidence. The sample size in this data was extremely small, therefore it is hard to give any strong answers due to the low statistical power. This model also requires a neurologist to mark the seizure episode prior to classification which is not easy to do and could lead to individual bias. Therefore, there is much more research and analysis that needs to be performed on this data, but so far it does provide great indication of the value of GEC events in TBI and epilepsy research.

[1] Vespa PM, Shrestha V, Abend N, Agoston D, Au A, Bell MJ, Bleck TP, Blanco MB, Claassen J, Diaz-Arrastia R, Duncan D, Ellingson B, Foreman B, Gilmore EJ, Hirsch L, Hunn M, Kamnaksh A, McArthur D, Morokoff A, O'Brien T, O'Phelan K, Robertson CL, Rosenthal E, Staba R, Toga A, Willyerd FA, Zimmermann L, Yam E, Martinez S, Real C, Engel J Jr. The epilepsy bioinformatics study for anti-epileptogenic therapy (EpiBioS4Rx) clinical biomarker: Study design and protocol. Neurobiol Dis. 2019 Mar;123:110-114. doi: 10.1016/j.nbd.2018.07.025. Epub 2018 Jul 23. PMID: 30048805; PMCID: PMC6344322.

[2] Bragin, A., Almajano, J., Kheiri, F., & Engel, J., Jr (2014). Functional connectivity in the brain estimated by analysis of gamma events. *PloS one*, *9*(1), e85900. https://doi.org/10.1371/journal.pone.0085900

[3] Li L, He L, Harris N, Zhou Y, Engel J Jr, Bragin A. Topographical reorganization of brain functional connectivity during an early period of epileptogenesis. Epilepsia. 2021 May;62(5):1231-1243. doi: 10.1111/epi.16863. Epub 2021 Mar 15. PMID: 33720411; PMCID: PMC8302261.