**A Case Study in a Machine Learning Framework Applied to Epilepsy Localization**

Anthony D. Bowman

Department of Electrical and Computer Engineering

University of Alabama at Birmingham, Birmingham, AL, USA

anbowman@uab.edu

Wesley Conwell

University of Alabama at Birmingham, Birmingham, AL, USA

Leon Jololian, PhD

Department of Electrical and Computer Engineering

University of Alabama at Birmingham, Birmingham, AL, USA

leon@uab.edu

**Abstract**

The rise of machine learning methodologies in recent years has seen great success in a variety of applications. However, this new paradigm is often utilized in limited ways through arbitrary selection of machine learning algorithm(s) and static feature sets, particularly in the medical literature. We have previously published a framework that removes these artificial limiters while laying the groundwork for parallel research and development tracks. To showcase the potential power from this expanded use of the machine learning paradigm, we applied this framework to the complex medical problem of epileptic seizure localization. Resting state EEG/MEG data were simultaneously collected from 22 patients prior to epilepsy surgery and retroactively selected for analysis. Power spectral and coherence features were extracted from all sensor time series data. Sets and subsets of these features were used to train multiple machine learning algorithms for classifying epilepsy in different brain regions. Models generated by a variety of algorithms and trained by delta, theta, beta, and low gamma MEG and EEG features were able to achieve an f-measure > 0.95 when distinguishing between left frontal epilepsy and bilateral extra-frontal epilepsy patients. Results show the artificial neural network also achieved this f-measure, but only when trained on the subset of features including beta and low gamma EEG features. Models generated by training the same algorithms and feature sets only achieved maximally an f-measure of 0.818 when classifying right frontal epilepsy versus bilateral extra-frontal epilepsy. In this study, using parallel applications of the machine learning paradigm, we were able to both improve on results seen in previous studies in classifying epilepsy and showcase the potential for meta-analysis across research tracks. This study provides additional insights into how research can be greatly expedited and expanded in scope through parallel exploration of topics which share overlapping feature or data sets.

**Introduction**

With the emergence of the machine learning paradigm in recent years, researchers in a wide variety of fields have sought out solutions generated by machine learning algorithms. While machine learning techniques continue to evolve, their adoption and usage is often carried out in an ad hoc manner, particularly by those whose expertise lies outside the field of computer science or related disciplines. One method to help bridge the gap between the non-expert and such an evolving methodology would be to introduce new development environments to aid in facilitating adoption of the machine learning paradigm. We have previously published a novel framework designed from the ground up with this specific purpose, first as a framework for a course in machine learning and later as a framework for a development environment (Bowman & Jololian, 2021; Bowman, Prabhakar & Jololian, 2022). We encourage the reader to review these papers for a detailed description of the framework. In this research, we applied this framework and methodology to a problem in the medical domain involving the localization of seizure onset zones in epilepsy patients.

To find an automated solution for localizing the seizure onset zone in patients with focal epilepsy, previous studies have explored applying machine learning to neuro-imaging data with limited success. These include attempting to analyze MEG data with a support vector machine to identify high frequency oscillations thought to correspond to epilepsy activity (Guo et al, 2018). Others have used the support vector machine to analyze graph theoretical features extracted from fMRI data to lateralize and localize seizures to the temporal lobes (Wu et al, 2018). Other studies attempted to train an artificial neural network and support vector machine on frequency domain features extracted from MEG data to discriminate between healthy and controls and epilepsy and between frontal focal epilepsy and generalized epilepsy (Aoe et al, 2019; Soriano et al, 2017). These studies are representative of a large portion of medical literature we have reviewed because of their use of either the support vector machine or artificial neural network. While they often do not discuss how they arrived at the algorithm used, we suspect they chose it because they saw it employed with some success in previous literature. This represents an almost arbitrary decision in their methodology. Our proposed framework seeks to correct this aspect of the literature by providing a more comprehensive approach toward predictive model construction and selection. Early results from our framework presented in this paper include some models with accuracy consistent with these previous studies, but also some models with higher accuracy, sometimes using the same classifier but with different features.

Applying our framework to the case study, we instantiated this three-layer architecture with each layer broadening the scope of our search through the solution space. Beginning from the research layer, we were able to generate multiple research questions that could be explored from the same data set. Data selection and preprocessing was then determined according to the needs of those research questions with one control layer generated to address each question. Within each control layer the feature and classifier sets were then defined, and feature extraction implemented. Of note is the ability for this architecture to allow for the researcher to explore the effect completely different features have in relation to each research question. Model construction and testing was then performed in every instantiation of the composite layer with a confusion matrix generated.

**Methods**

We retrospectively selected patients in our database with medically intractable epilepsy (n = 22). All patients had previously undergone surgical resection and were seizure free for at least six months, thus confirming their epileptic locus was within the resected region. As part of their pre-surgical evaluation, all patients had an MEG study performed using the system described below. The study was approved by the Institutional Review Board at the University of Alabama at Birmingham.

**MEG Acquisition and Preprocessing**

All MEG recordings were performed using a whole-head, 148-channel system housed within a magnetically shielded room (4D Neuroimaging, San Diego, CA). All patients were in a reclined position for the duration of the recordings. Multiple recordings were collected from each patient, each lasting 10 min and collected at a sampling rate of 508.63 Hz. Each data file was then preprocessed with in-house MATLAB scripts using Statistical Parametric Mapping software (SPM12b, [http://www.fil.ion.ucl.ac.uk/spm/](about:blank)).

**EEG Acquisition and Preprocessing**

All EEG recordings were gathered using the International 10-20 system of electrode placement and gathered concurrently with each MEG recording. All EEG data was gathered at a sampling rate of 2 kHz and down-sampled to 600 Hz using a low-pass filter. Only data from 25 EEG leads common across all patients were included for further analysis and feature extraction.

**MEG and EEG Spectral and Coherence Feature Extraction**

Feature extraction from both MEG and EEG recordings was performed using a combination of Brainstorm functions and in-house MATLAB scripting (MATLAB, 2018; Tadel et al, 2011). Mean power spectrum density (Welch method) was computed using the Brainstorm function with the frequency bands slightly adjusted to the following: Delta band from 1 – 3 Hz, theta band from 3 – 8 Hz, alpha band from 8 – 12 Hz, beta band from 15 – 29 Hz, low gamma band from 30 – 59 Hz, high gamma band from 60 – 90 Hz. Spectral coherence features between sensors were computed from both MEG and EEG time series data using the mean square coherence function in MATLAB 2018a.

**Machine Learning Training and Testing**

All classifier training and testing was performed using the Waikato Environment for Knowledge Analysis (WEKA) open-source machine learning software (Witten et al, 1999). Default parameters were used for each classifier unless otherwise noted. Unless otherwise noted, all classifiers were trained and tested using 10-fold leave-one-out cross validation with WEKA reporting the confusion matrix, precision, recall, and weighted f-measure for each trained model. Classification was performed using different combinations of features provided to each classifier to find the combination of classifier and feature set that produced the highest f-measure.

**Results**

In Table 1, we present some results with seizure localization to the left frontal lobe. Using the weighted f-measure reported by WEKA as our metric for this matrix, we were able to see a wide range in results depending on algorithm and feature set. For this and future tables of results, we denote a calculated f-measure as “not a number” (“NaN”) when the calculation involves division by zero. The maximum f-measure for each table is bolded. Maximum f-measure of 0.951 was achieved from models generated by logistic regression, stochastic gradient descent (SGD), simple logistic, support vector machine (SMV), and logistic model tree (LMT) algorithms trained by a feature set containing MEG power in the delta, theta, beta, and low gamma ranges. In the case of logistic regression and SVM, achieved maximum f-measure decreased when features extracted from EEG were included in addition to the aforementioned MEG features. The multilayer perception, otherwise known as the artificial neural network (ANN), was the only algorithm to match the maximum f-measure with fewer, albeit different features.

Table 1: Left frontal vs bilateral extra-frontal

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Delta EEG | Theta EEG | Beta and low gamma EEG | Delta, theta, beta, low gamma EEG | Delta, theta, beta, low gamma MEG | EEG&MEG |
| BayesNet | NaN | 0.777 | NaN | 0.777 | 0.872 | 0.872 |
| NaiveBayes | 0.653 | 0.753 | 0.777 | 0.777 | 0.291 | 0.347 |
| NaiveBayesMultinomial | NaN | NaN | NaN | NaN | NaN | NaN |
| Logistic Regression | 0.727 | 0.727 | 0.818 | 0.753 | **0.951** | 0.836 |
| SGD | 0.777 | 0.777 | **0.909** | 0.777 | **0.951** | **0.951** |
| Multilayer Perceptron | 0.777 | 0.777 | **0.951** | 0.777 | **0.909** | **0.909** |
| SimpleLogistic | 0.753 | 0.777 | **0.889** | 0.777 | **0.951** | **0.951** |
| SMO (SVM) | NaN | 0.777 | NaN | NaN | **0.951** | **0.909** |
| DecisionStump | 0.753 | 0.753 | 0.777 | 0.753 | 0.872 | 0.872 |
| J48 | NaN | 0.753 | **0.889** | 0.786 | NaN | 0.786 |
| LMT (log tree) | 0.753 | 0.777 | 0.889 | 0.777 | **0.951** | **0.951** |
| Random Forest | NaN | 0.777 | NaN | NaN | 0.777 | 0.852 |
| Random Tree | 0.700 | 0.727 | **0.909** | 0.727 | 0.852 | 0.818 |

Table 2 shows the results from the same algorithms trained with the same feature sets, relabeled to data from patients with right frontal epilepsy. In this classification task, the highest weighted f-measure of 0.818 was only achieved by the logistic regression algorithm trained using the combined MEG feature set with power in all stated frequency bands.

Table 2: Right frontal vs bilateral extra-frontal

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Delta EEG | Theta EEG | Beta and low gamma EEG | Delta, theta, beta, low gamma EEG | Delta, theta, beta, low gamma MEG | EEG&MEG |
| BayesNet | NaN | NaN | NaN | NaN | **0.805** | 0.805 |
| NaiveBayes | 0.084 | 0.364 | 0.611 | 0.261 | 0.287 | 0.287 |
| NaiveBayesMultinomial | NaN | NaN | NaN | NaN | NaN | NaN |
| Logistic Regression | 0.570 | 0.613 | 0.398 | 0.600 | **0.818** | 0.636 |
| SGD | 0.590 | 0.590 | 0.745 | 0.636 | 0.778 | 0.600 |
| Multilayer Perceptron | 0.513 | 0.484 | 0.422 | 0.600 | 0.727 | 0.579 |
| SimpleLogistic | NaN | 0.566 | NaN | NaN | 0.566 | NaN |
| SMO (SVM) | NaN | 0.590 | NaN | 0.642 | 0.611 | 0.566 |
| DecisionStump | 0.513 | 0.455 | 0.590 | 0.590 | **0.805** | 0.805 |
| J48 | 0.540 | 0.540 | 0.513 | 0.485 | NaN | 0.485 |
| LMT (log tree) | NaN | 0.513 | NaN | NaN | 0.566 | NaN |
| Random Forest | 0.540 | 0.579 | 0.566 | 0.540 | 0.642 | 0.745 |
| Random Tree | 0.438 | 0.600 | 0.485 | 0.441 | 0.611 | 0.745 |

Table 3 shows the results training the same set of algorithms with coherence features extracted from MEG in the theta, alpha, beta, and low gamma frequency ranges (the far-right column being the combined feature set with all frequency bands included; “OOM” denotes an out-of-memory error when training the model). Here, the models were trained to discriminate between right and left temporal lobe epilepsy, also known as lateralization. In this task, the maximum weighted f-measure of 0.818 was achieved from the SimpleLogistic and LMT algorithms trained using only coherence features in the beta frequency band as well as the random tree algorithm using coherence features in the beta and low gamma bands.

Table 3: Left vs right temporal lobe epilepsy, MEG feature sets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Theta Coherence | Alpha Coherence | Beta Coherence | Low Gamma Coherence | Beta & Low Gamma Coherence | TABG Coherence |
| BayesNet | 0.364 | 0.364 | 0.723 | 0.723 | 0.696 | 0.617 |
| NaiveBayes | 0.696 | 0.617 | 0.538 | 0.545 | 0.364 | 0.538 |
| NaiveBayesMultinomial | 0.091 | 0.636 | 0.364 | 0.455 | 0.538 | 0.445 |
| Logistic Regression | 0.445 | 0.331 | 0.723 | 0.445 | 0.636 | 0.455 |
| SGD | 0.331 | 0.261 | 0.538 | 0.636 | 0.723 | 0.723 |
| Multilayer Perceptron | 0.331 | 0.261 | 0.617 | 0.617 | 0.636 | OOM |
| SimpleLogistic | 0.140 | 0.195 | **0.818** | 0.445 | 0.727 | 0.331 |
| SMO (SVM) | 0.261 | 0.261 | 0.617 | 0.617 | 0.636 | 0.455 |
| DecisionStump | 0.331 | 0.140 | 0.727 | 0.445 | 0.727 | 0.331 |
| J48 | 0.636 | 0.261 | 0.727 | 0.445 | 0.727 | 0.445 |
| LMT (log tree) | 0.140 | 0.195 | **0.818** | 0.445 | 0.727 | 0.331 |
| Random Forest | 0.331 | 0.364 | 0.808 | 0.455 | 0.696 | 0.636 |
| Random Tree | 0.445 | 0.455 | 0.723 | 0.455 | **0.818** | 0.538 |

**Discussion of Technical Results**

From an engineering standpoint, our methodology presents a more thorough exploration through both the solution space associated with each research question as well as the capability to efficiently investigate multiple research questions. Whereas the traditional methodology would have arrived at one or two of these models, our framework allows us a much broader view of the landscape from which we can select the maximum weighted f-measure. Through this broader view, we see some models with tested f-measure approximately equal to those in previous studies, potentially confirming their results. Results also show some models achieve higher f-measure, sometimes using the same algorithm as previous studies but trained with a different feature set. In this case, the maximum was observed by multiple models which presents us the opportunity to consider other metrics as well as generate additional research questions that may explore why different feature sets resulted in high accuracy. Further experimentation may explore the effect changing various hyper-parameters has on improving the performance of some algorithms, such as altering the number of hidden layers of the artificial neural network. Stepping back to take in a more abstract view of these results leads to additional questions relating to model performance relative to feature subset, in some cases decreasing substantially. This comparative meta-analysis is further expanded on by the opportunity to efficiently investigate parallel research questions (right and left frontal lobe epilepsy) and comparing the results from the same algorithms and feature sets. We were also afforded the opportunity to efficiently explore another research track in tandem by reusing the same data set, resulting in the investigation into the possible relationship between coherence features and focal epilepsy.

**Discussion of Domain Specific Medical Results**

From a medical standpoint, our results show an interesting disparity between the metrics achieved to discriminate between focal epilepsy in different brain regions. This is most easily seen when comparing the results in Tables 1 and 2: Investigating mirrored hypotheses, left frontal vs right frontal epilepsy, reveals classification of left frontal epilepsy using the defined power spectral features to be the “easier” of the two tasks. This may suggest the underlying neurophysiological characteristics associated with right frontal lobe epilepsy consist of a more complex pattern than that which can be identified by these algorithms and feature sets for left frontal lobe epilepsy. This may imply that focal epilepsy originating from the right frontal lobe may be better characterized by changes in features beyond power spectra. The nature of these features may provide further insight into the processing and functional structure of the right frontal lobe and how electrical pathologies such as epilepsy disrupt normal function.

Our results also show greater opportunity for success in such classification tasks with power spectral features over coherence features, although this observation may change with focal epilepsy in other cortical regions. Further research is needed with a larger data set to confirm this trend. Spectral coherence was chosen as a feature because of the connectivity exhibited between the temporal lobes and other regions of the brain (Haneef et al, 2014; Spencer, 2002). Previous studies have also explored other methods of quantifying neural or cortical connectivity such as transfer entropy, directed transfer functions, and graph theoretic metrics (Basu et al, 2015; Dai et al, 2012; Ursino et al, 2020; Wu et al, 2018). Mirroring the limited success seen in those studies, our results suggest further research is needed to explore these different metrics for connectivity in combination or refined, perhaps with more narrowly defined coherence metrics than used here. Our results clearly show the need to explore feature subsets to train multiple machine learning models for testing as a subset may lead to higher performance than the complete feature set.

**Conclusions**

In this study, we employed a previously published framework for the development of machine learning solutions conducted in the context of a medical case study. Using this methodology to embrace the machine learning paradigm more fully, we were able to efficiently explore both the problem and solution spaces within the case study’s domain. Implementation of the composite layer allowed for empirical identification of algorithm-feature set pairings with higher performance metrics than seen in previous studies. Implementation of the control layer in combination of the composite layer greatly improved the scope of our work, expanding the potential of the project to explore multiple, related domain-specific questions in rapid succession through the re-use of data and feature sets.

**Acknowledgements**

This work was supported by an institutional grant from the University of Alabama at Birmingham. We would also like to acknowledge the contributions of Dr. Ismail Mohamed and Jeff Killen in data collection, management, and selection.

**References**

Aoe, J., Fukuma, R., Yanagisawa, T., Harada, T., Tanaka, M., Kobayashi, M., ... & Kishima, H. (2019). Automatic diagnosis of neurological diseases using MEG signals with a deep neural network. *Scientific reports*, *9*(1), 1-9.

Basu, I., Kudela, P., Korzeniewska, A., Franaszczuk, P. J., & Anderson, W. S. (2015). A study of the dynamics of seizure propagation across micro domains in the vicinity of the seizure onset zone. *Journal of neural engineering*, *12*(4), 046016.

Bowman, A. D., & Jololian, L. (2021). A conceptual framework for an introductory machine learning course. *Journal of Computing Sciences in Colleges*, *37*(1), 78-83.

Bowman, A. D., Prabhakar, S. P., & Jololian, L. (2022, March). A Framework for an Automated Development Environment to Support the Data-driven Machine Learning Paradigm. In *SoutheastCon 2022* (pp. 329-331). IEEE.

Dai, Y., Zhang, W., Dickens, D. L., & He, B. (2012). Source connectivity analysis from MEG and its application to epilepsy source localization. *Brain topography*, *25*(2), 157-166.

Guo, J., Yang, K., Liu, H., Yin, C., Xiang, J., Li, H., ... & Gao, Y. (2018). A stacked sparse autoencoder-based detector for automatic identification of neuromagnetic high frequency oscillations in epilepsy. *IEEE transactions on medical imaging*, *37*(11), 2474-2482.

Haneef, Z., Lenartowicz, A., Yeh, H. J., Levin, H. S., Engel Jr, J., & Stern, J. M. (2014). Functional connectivity of hippocampal networks in temporal lobe epilepsy. *Epilepsia*, *55*(1), 137-145.

MATLAB. (2018). *9.4.0.813654 (R2018a)*. Natick, Massachusetts: The MathWorks Inc.

Soriano, M. C., Niso, G., Clements, J., Ortín, S., Carrasco, S., Gudín, M., & Pereda, E. (2017). Automated detection of epileptic biomarkers in resting-state interictal MEG data. *Frontiers in neuroinformatics*, *11*, 43.

Spencer, S. S. (2002). Neural networks in human epilepsy: evidence of and implications for treatment. *Epilepsia*, *43*(3), 219-227.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience*, *2011*.

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., & Cunningham, S. J. (1999). Weka: Practical machine learning tools and techniques with Java implementations.

Ursino, M., Ricci, G., & Magosso, E. (2020). Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models. *Frontiers in computational neuroscience*, *14*, 45.

Wu, T., Chen, D., Chen, Q., Zhang, R., Zhang, W., Li, Y., ... & Zhang, J. (2018). Automatic lateralization of temporal lobe epilepsy based on MEG network features using support vector machines. *Complexity*, *2018*.

**Biographies**

Anthony D. Bowman is a doctoral candidate seeking a PhD in Computer Engineering at the University of Alabama at Birmingham. He earned both his MS in biomedical engineering in 2017 and BS in computer and information sciences in 2012 from the University of Alabama at Birmingham. Anthony D. Bowman’s research interests include machine learning, data science, and bioinformatics. Anthony D. Bowman may be reached at [anbowman@uab.edu](mailto:anbowman@uab.edu)

**WESLEY CONWELL** is a PhD candidate of interdisciplinary engineering at the University of Alabama at Birmingham. He earned his BS in electrical engineering from the University of Alabama, MBA from the University of Alabama at Birmingham, and MS in electrical engineering from the University of Alabama at Birmingham. Mr. Conwell’s research interests include smart cities, data analytics, big data, Internet of Things and machine learning. Mr. Conwell may be reached at [wlconwel@uab.edu](mailto:wlconwel@uab.edu).

Leon Jololian is a professor of electrical and computer engineering at the University of Alabama at Birmingham. He earned his BS in electrical engineering from Manhattan College, his MS in electrical engineering from Georgia Institute of Technology, his MS in computer science from Polytechnic University, and his PhD in computer science from the New Jersey Institute of Technology. Dr. Jololian’s research interests include machine learning, mobile computing, and the internet of things. Dr. Jololian may be reached at [leon@uab.edu](mailto:leon@uab.edu)