Predicting Antidepressant Response with Resting Electroencephalogram

Drew Johnston
Brigham Young University
April, 2020

ABSTRACT. This paper endeavors to establish machine learning methods for predicting whether a depressed patient will have a positive response to antidepressants. Outlined in the paper are efforts to extract predictive features from electroencephalogram (EEG) signals of depressed patients and use them to predict patient response. All data comes from the EMBARC dataset provided by the National Institute of Mental Health Data Archive. Traditional machine learning models proved ineffective at predicting patient response from typical EEG features, but a recently developed machine learning approach called SELSER gave promising results, predicting patient response with an accuracy of 78.9%.

Keywords: Depression, Antidepressants, Electroencephalogram, EMBARC.

1. Problem Statement and Motivation

Major depression is one of the most prevalent and debilitating mental disorders in the United States. In the most recent National Survey on Drug Use and Health, an estimated 17.3 million U.S. adults and 3.2 million adolescents had at least one episode of major depression in the last year; 11 million of the adults and 2.3 million of the adolescents experienced severe impairment which interfered with their ability to carry out major life activities (4). In addition to the debilitating effects of living with depression, many who live with depression experience suicidal ideations and attempt suicide.

In recent decades, great strides have been made to alleviate the suffering of those who experience major depression, particularly in the areas of Cognitive Behavioral Therapy and antidepressant medication. The most prevalent and effective antidepressants in use today are Selective Serotonin Reuptake Inhibitors (SSRIs), such as Prozac or Zoloft (5). For reasons as yet unknown, some people respond well to SSRIs, while others experience no change in their depression symptoms or even an increase in the severity of the symptoms when taking SSRIs. While medical professionals may differ in which SSRI they prescribe a patient based on pharmacokinetics or potential side effects, it is difficult to know a priori how the patient will respond to the SSRI (3).

Antidepressants alleviate symptoms of depression through chemical interactions with neurotransmitters in the brain. Electroencephalograms (EEG) record brain wave patterns (see Figure 1.1) that can inform doctors about the physiology of the brain. In this paper, I strive to extract features from EEG signals that will help solve the problem of antidepressant prescription. I hope to answer the question: can a patient's response to an SSRI be predicted by analyzing their EEG before they take the drug? If this can be accomplished, it may save patients months of increased suffering caused by ineffective medication. This would lead to increased quality of life and may save the lives of those at risk of suicide.

Received by the editors April 15, 2020.

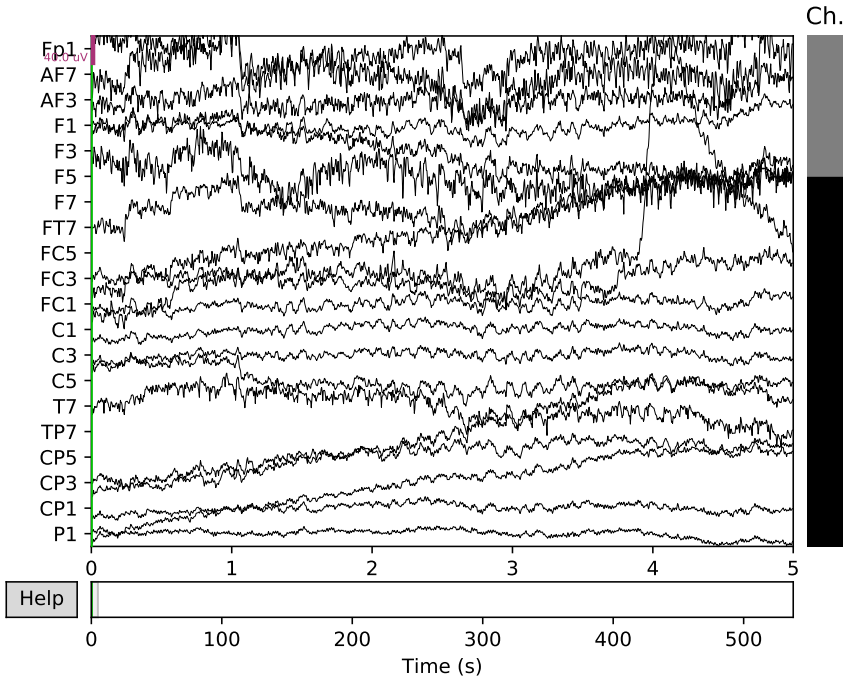

FIGURE 1.1. An example EEG signal with multiple channels

Research has already shown some relationships between EEG signal and patient response. Van der Vinne et al. showed that abnormalities in EEG signals indicated patients were less likely to respond to escitalopram, but not sertraline (6). Jaworska et al. developed eLORETA features to be used with EEG band power measures for response prediction, and Bruder et al. showed that EEG alpha measures could be used to predict SSRI response (1; 2). This paper attempts to develop and verify EEG features that can be extracted from a single EEG taken of a resting patient before any treatment. Performing response prediction with these features will not require domain knowledge, but any results should be verified with a medical professional.

I wish to stress that antidepressants are not the only effective way to get help for those with depression. It has been shown that Cognitive Behavioral Therapy, a strong support system, and appropriately prescribed antidepressants are more effective in conjunction with one another. Please seek out any and all appropriate sources of help if you are suffering.

2. Data

The data used in this paper is provided by the National Institute of Mental Health Data Archive (NDA) under the collection title ”Establishing Moderators/Biosignatures of Antidepressant Response - Clinical Care (EMBARC) MDD Treatment and Controls 2199.” While it is free, to access this data you must be sponsored by an NIH recognized institution with a Federalwide Assurance and have a research-related need.

The data used in the study was gathered by medical professionals at four medical research locations in the United States and great care was taken to ensure the EEG data I employed was usable and accurate. At the beginning of the study, 400 participants with depression were selected. Half of them (200) were randomly selected to receive an SSRI while the other 200 were given a placebo.

If those that were given a placebo did not respond to the placebo, they were switched to the antidepressant after 8 weeks. Those who responded to the placebo did not take any antidepressants during the study and are not included in the models.

As with any study, not all participants remained for the entirety of the study, and some of those who did failed to record a usable EEG. For my purposes, the only usable data came from patients who were given the antidepressant during the study, recorded a valid EEG signal at the beginning of the study, and recorded all features I was interested in. This included Hamilton Depression Rating Scale (HDRS) at weeks 1, 8, and 16, Clinical Global Improvement (CGI) score at week 8, and week 8 patient response, which was a "yes" or "no" value. Week 16 patient response values to be used for those that switched from placebo to the SSRI were determined based on the HDRS values at weeks 8 and 16. Using these filters reduced the 400 original participants to 187 suitable data points to work with across all four research facilities.

3. Ethics

The purpose of this investigation is ultimately to help relieve pain and suffering where possible. However, it's important to recognize that there is still much that is not understood about the complex nature of depression and antidepressant response. Regardless of any conclusions drawn from the results of this paper, patients are encouraged to carefully consider the counsel of their attending physicians in all matters and make educated decisions for themselves. If you have concerns about your medication, please speak with your doctor. Mental health is of a very case-by-case nature and should be treated as such. I would like to add that at the time of writing I have no competing interests with respect to any of the subjects addressed in this paper.

4. Methods

I approached this problem as a binary classification problem, training a variety of machine learning models to classify patient response to the SSRI as positive or negative. Before training models, I used Python's MNE package to extract features from each EEG signal. Features that I looked at included mean, variance, skewness, kurtosis, frequency band power and energy, SVD entropy, Fisher information, Hjorth mobility, Hjorth complexity, and Higuchi Fractal Dimension for each EEG channel. This resulted in 1,345 total features. In an effort to avoid any issues with collinearity and to reduce computation time, Principal Component Analysis (PCA) was used to get 10 principal components from these features, which explained 99.7% of the variance of the data. Models were trained in three different settings: once with all features used, once with the principal components, and once with only powers of the five canonical EEG frequency bands across each channel. Band power was chosen because it produced high feature importance when fitting a Random Forest model to the data and because is has been shown in previous studies to be correlated with SSRI response.

The following methods were tried in each of the settings explained above: Naive Bayes', Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Support Vector Machines with linear, radial, and polynomial kernels, Random Forest, and Gradient Boosting models. Clustering methods such as t-SNE and UMAP were also explored to determine if any patterns might be unveiled through unsupervised methods, but these availed little and were soon abandoned.

The majority of these models performed very poorly with accuracies barely better than randomly guessing. Some nonlinear models outperformed the rest, particularly Random Forest and Support

Vector Machine with a second degree polynomial kernel. For that reason, the remainder of my results with traditional machine learning algorithms will be pertaining to these two models. Models that trained only on the frequency band power instead of the entire dataset performed consistently better as well, so I continued to explore this restricted feature space.

In addition to these two models, I explored the use of Sparse EEG Latent SpacE Regression (SELSER) for response prediction. SELSER was developed in 2019 by a team lead by Wei Wu and copyrighted in February of 2020 by Alto Neuroscience (7). SELSER was developed by Wu for treatment prediction using pre-treatment EEG data. It had not been used to for a classification problem as I have presented in this paper, but it had been shown to robustly predict depression symptom improvements using pre-treatment, resting EEG data, so I expected to get favorable results.

The SELSER algorithm does not accept raw EEG features as were used in training the other models. It requires symmetric spatial covariance matrices, as described in section 4.3. After testing the performance of SELSER with spatial covariance matrices derived from epoched data for frequency band power of the EEGs, I used the entries of these covariance matrices as predictive features for the Random Forest and SVM models as well for comparison. All three models produced much better results when trained on these covariance features instead of the previous raw EEG features.

Because the covariance dataset is composed of symmetric matrices, some elements of image data augmentation can be used to extend the limited dataset and increase model performance. The dataset was augmented by adding random noise to each entry of each matrix. This noise was drawn from a normal distribution with mean zero and variance equal to the variance of the entries of the matrix in question. Great care was taken to ensure that only the data being trained on at each fold of cross validation was being augmented, and not the data used for testing. This data augmentation increased model performance by multiple percentage points for each model.

4.1. Support Vector Machine

In the setting of binary classification, a Support Vector Machine (SVM) is a supervised learning model that algorithmically finds an optimal hyperplane to separate the feature space into two regions based on which class each data point belongs to. Then if a new datapoint shows up in one region, it is assigned to that associated class. Using a second degree polynomial kernel allowed for the model's separating hyperplane to be quadratic in nature, which proved to be an advantage in the feature space.

SVMs are very effective when classes are separable, but can have difficulty when there is significant overlap. While selecting the appropriate kernel can be tricky, the kernel trick allows for favorable computation time for a nonlinear model. The type and degree of the kernel, as well as appropriate regulation terms, were chosen through grid search cross validation with five folds.

4.2. Random Forest

A Random Forest model is built by collecting an ensemble of several decision trees together. Each decision tree takes the observations from a data point and returns which class the tree concludes that it belongs to. All the trees vote on the class and the majority make the final decision. Some strengths of random forests include ability to handle imbalanced or missing data, variance reduction of the data from bootstrap aggregation, and a built in method for exploring feature importance.

To optimize performance, a five-fold grid search cross validation was run using a variety of parameters such as max depth and minimum features per split branch.

4.3. SELSER

SELSER is an extremely new algorithm that became available for public use in February of 2020. I have used it with permission from Alto Neuroscience. Their code was written in MATLAB; I spent some time translating it to Python for this project and plan to share this file with Alto Neuroscience if they are interested in providing the algorithm on multiple platforms.

SELSER was designed specifically for EEG signal data. From the website for Alto Neuroscience: "The algorithm first optimizes a sparse set of spatial filters that map the EEG signals to a latent space, and then relates the band powers of the latent signals to the treatment outcome via a linear regression model. The unknown parameters in the spatial filters and linear regression model are optimized in conjunction under a convex optimization framework, where the accelerated proximal gradient method is employed to efficiently solve the optimization problem" (7).

However, the algorithm was not designed to accept epoched EEG features as I had prepared for the other models. The data, X, required by the model is an array of symmetric spatial covariance matrices. Each X_i is calculated by

$$X_i = \frac{1}{N} x x^T, \tag{4.1}$$

where x is the epoched data of interest and N is the number of epochs included.

Specifically, given a regularization term λ, the nuclear norm $||\cdot||$, the data X, and the labels Y, it solves the optimization problem

$$\min_{W,b} f(W,b) + g(W) \tag{4.2}$$

$$\text{where } f(W,b) = \sum_i (Tr(W^T X_i) + b - Y_i)^2, \tag{4.3}$$

$$g(W) = \lambda ||W||. \tag{4.4}$$

Then W gives the weights and b the intercept for a linear regression model to predict Y given X. The algorithm also returns a matrix of spatial filters and regression coefficients associated with each latent variable.

While frequency band power features are correlated with patient response, it easier to make predictions when these features are mapped to a particular latent space via the SELSER algorithm than when they remain in the usual feature space. SELSER is a powerful tool for EEG analysis, but is very limited in its applications. That being said, it is well-suited to the problem at hand.

5. Results

The models trained on raw EEG features performed poorly, with accuracies of 65.3% for the SVM and only 59.5% for the Random Forest model. However, the spatial covariance data was much more effective. When the augmented covariance data was included in training, all models achieved over 70% accuracy.

The SELSER algorithm outperformed the others with the highest accuracy of 78.9% when trained on the augmented covariance dataset. All accuracy values were attained using 5-fold cross validation. The full results are visualized in Figure 5.1.

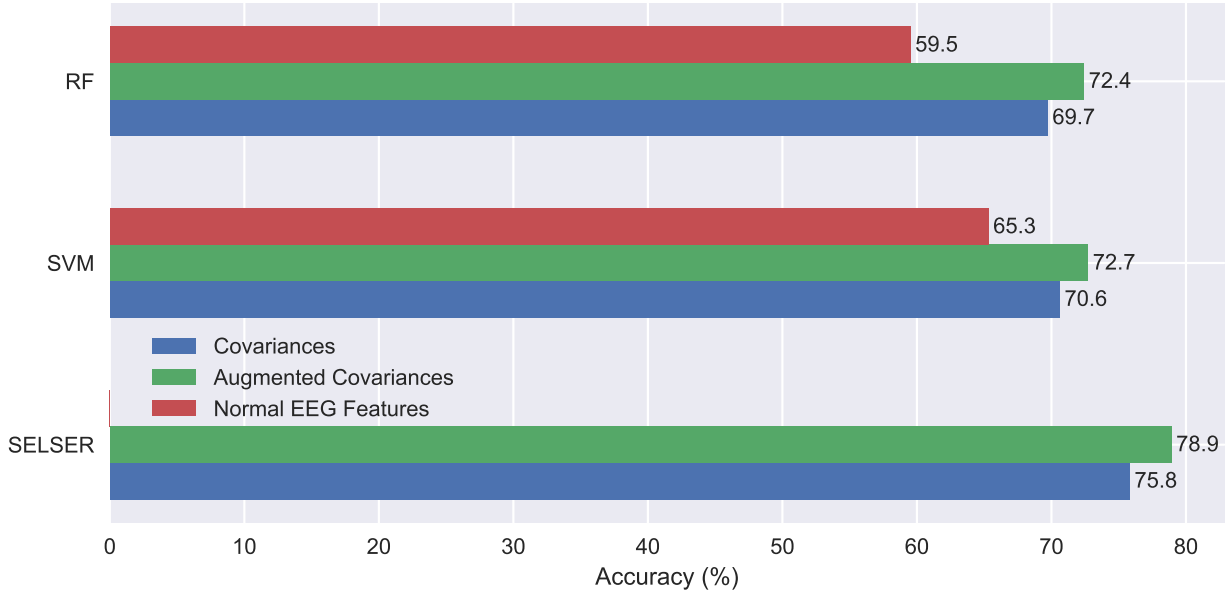

FIGURE 5.1. Model Accuracies by Dataset Used for Training

6. Analysis

One of the intentions of using the Random Forest model was to make use of automated feature importance calculations to determine which features were most important in the decision making process of the Random Forest. That is, which features contribute the most to decreasing the average Gini impurity over all trees. Gini Impurity is a measurement of the likelihood of incorrect classification of new data if the new data is randomly classified according to the distribution of class labels from the data set.

This is normally a valuable feature of Random Forest models, however, the Random Forest model performed poorly when utilizing the raw EEG features. The resulting accuracy was barely better than what would be achieved with random guessing, so feature importances for this model are unlikely to provide helpful information.

The increase in performance that resulted from considering the spatial covariances of only frequency band power of the EEG signal indicate that band power is correlated with patient response and can be used for prediction. The performance of the SELSER algorithm also indicates that this correlation can be better taken advantage of when the covariance data is mapped from the original feature space to a latent space. This difference in performance suggests that SELSER is much better suited than traditional machine learning models for treatment response prediction from resting EEG signals.

While SELSER performed well in comparison to the other models, other methods of utilizing the latent space mapping produced by SELSER should be explored. The algorithm relates the band

powers of the latent signals to the treatment response through linear regression, but other models could be explored that might more effectively predict patient response from these latent signals.

In addition, further research should be made to determine if EEG features beyond band power would benefit from being mapped to this latent space. It is possible that latent signals produced from EEG features such as band energy, kurtosis, or entropy could produce similar or improved results for treatment response prediction.

7. Conclusion

If doctors can know a priori how a patient will respond to an SSRI, they can be more confident in their prescription of antidepressants and potentially save the patient from months of increased suffering and increased risk of suicide, along with a plethora of other uncomfortable side effects. EEG signals are a promising source of features that are correlated with patient response.

While the polynomial SVM and Random Forest models held a slight edge over other tested machine learning models, the raw EEG features extracted did not prove effective in response prediction. However, after shifting to consider the spatial covariance matrices of frequency band power data from the EEG signals, the models produced much more accurate results, particularly when augmenting the data with some Gaussian noise. The SVM and Random Forest models both increased in performance with accuracies over 72%, while the SELSER algorithm outperformed them both with 78.9% accuracy. In practice, it would be preferrable to attain much higher results before using these models to influence such important prescription decisions, but these results are a promising step forward toward effective use of machine learning tools to assist our medical professionals.

8. Future Work

While predicting patient response is an important step, it is important to be able to determine the degree of patient response as well. Furthermore, people respond differently to different SSRI compounds, despite their mechanisms of neurotransmitter interaction being identical in theory. It is not understood why this is the case, but understanding which SSRIs will work and which will not could save months of experimentation when initially prescribing antidepressants. Questions I'd like to explore include: will a patient respond better to an SSRI compared to other antidepressants (such as SNRIs or MAOIs)? And can we predict which SSRI will be most effective for a patient?

References

[1] G. E. Bruder, J. P. Sedoruk, J. W. Stewart, P. J. McGrath, F. M. Quitkin, and C. E. Tenke. Eeg alpha measures predict therapeutic response to an ssri antidepressant: Pre and post treatment findings. *Biological psychiatry*, 63(12):1171, 2008.

[2] N. Jaworska, S. de la Salle, M.-H. Ibrahim, P. Blier, and V. Knott. Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (eeg) and clinical data. *Frontiers in psychiatry*, 9:768, 2019.

[3] P. A. Marken and J. S. Munro. Selecting a selective serotonin reuptake inhibitor: clinically important distinguishing features. *Primary care companion to the Journal of clinical psychiatry*, 2(6):205, 2000.

[4] NIMH. Major depression. `https://www.nimh.nih.gov/health/statistics/suicide.shtml`, April 2019.

[5] N. Schimelpfening. Beyond prozac: The most commonly prescribed antidepressants. `https://www.verywellmind.com/most-common-antidepressants-1066939`, March 2020.

[6] N. van der Vinne, M. Vollebregt, N. Boutros, K. Fallahpour, M. van Putten, and M. Arns. Normalization of eeg in depression after antidepressant treatment with sertraline? a preliminary report. *Journal of affective disorders*, 259:67–72, 2019.

[7] W. Wu, Y. Zhang, J. Jiang, M. V. Lucas, G. A. Fonzo, C. E. Rolle, C. Cooper, C. Chin-Fatt, N. Krepel, C. A. Cornelssen, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nature Biotechnology*, pages 1–9, 2020.

(D. Johnston) BRIGHAM YOUNG UNIVERSITY, PROVO, UTAH 84602, USA
Email address, Corresponding author: `drew.johnston13@gmail.com`