Accepted Manuscript

Accepted Manuscript (Uncorrected Proof)

Title: Brain Activity Flow and Machine Learning for Predicting Drug Response in Patients with Major Depressive Disorder

Authors: Seyed Morteza Mirjebreili[1], Reza Shalbaf[1], Ahmad Shalbaf[2,*]

1. *Institute for Cognitive Science Studies, Tehran, Iran.*
2. *Department of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*

***Corresponding Author**: Ahmad Shalbaf, Department of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: shalbaf@sbmu.ac.ir

To appear in: **Basic and Clinical Neuroscience**

Received date: 2023/06/08

Revised date: 2023/12/27

Accepted date: 2024/01/21

Please cite this article as:

Mirjebreili, S.M., Shalbaf, R., Ahmad Shalbaf, A. (In Press). Brain Activity Flow and Machine Learning for Predicting Drug Response in Patients with Major Depressive Disorder. Basic and Clinical Neuroscience. Just Accepted publication Jul. 10, 2024. Doi: http://dx.doi.org/10.32598/bcn.2024.2034.6

DOI: http://dx.doi.org/10.32598/bcn.2024.2034.6

Abstract

A major challenge today is personalizing the treatment for Major Depressive Disorder (MDD) patients in order to make it more efficient. In order to address this issue, we have proposed a novel approach based on machine learning models that utilize neural activity flow prior to treatment with selective serotonin reuptake inhibitor (SSRI) medication. The electroencephalogram (EEG) signals of 30 patients were used to calculate the neural activity flow of each patient based on the direct Directed Transfer Function (dDTF). Then, based on the area under the curve (AUC) values, 30 important connections were identified for delta, theta, alpha, beta, and gamma bands. In order to select the most important neural activity flow, these neural activity flows are combined, and forward features, mRMR, and Relief-F methods are applied. Lastly, Support vector machines (SVMs), decision tree, and random forest models are trained using selected neural activity flows. Results showed that the most connections came from F8, Pz, T5, and P4, which are mostly from the frontal and parietal lobes. In addition, the SVM model showed 98% accuracy in classification using forward feature selection, with most of the neural activity flows selected from alpha and beta. Finally, results indicate that patients who responded to treatment differed in their patterns of frontoparietal neural activity flows, which implies the Frontoparietal Network is primarily involved in treatment response at alpha and beta frequencies. Therefore, the proposed method is capable of accurately detecting responders in MDD patients, which can reduce costs for both patients and medical facilities.

Keywords: Electroencephalogram (EEG), Effective Connectivity, Major Depressive Disorder (MDD), Machine learning.

Introduction

Major Depressive Disorder (MDD) is the most commonly diagnosed psychiatric disorder worldwide, affecting more than 300 million individuals [1]. Symptoms of MDD include changes in mood, interests, pleasure, cognitive functions, and vegetative symptoms. Furthermore, MDD increases the risk of developing conditions such as diabetes mellitus, heart disease, and stroke. In addition, MDD has also been associated with suicide, representing approximately half of the 800,000 suicides worldwide [2].

In the present day, several antidepressants that act on neurotransmitter receptors are being used to treat depression. Almost all drugs act on two or more neurotransmitter receptors, i.e., two serotonin receptors, two noradrenergic receptors, or both. Also, several treatments are being investigated, including estrogen replacement therapy, mifepristone (RU-486 or C-1073), as well as antagonists, such as corticotropin-releasing factor, neurokinins, and injectable pentapeptides [3], [4]. MDD is a highly heterogeneous disorder, which may mean that only a few people find antidepressants effective. Several pre-treatment variables have been found to moderate the treatment response, including depression severity and neuroticism, older age, less impairment in cognitive control, and employment[5], [6].

In general, 40% of people suffering from MDD have treatment-resistant depression (TRD) since treatment of MDD requires a trial-and-error sequential treatment strategy, and first-line therapies do not meet their needs [7]–[9]. Often, MDD patients suffer from delayed treatment response, functional impairment, increased suicide risk, and high medical costs due to the inability to predict which treatment will work. Consequently, more effective treatment strategies for patients with MDD are urgently needed [10]–[12].

Due to advances in neuroimaging techniques, biomarkers from neuroimaging studies are important for achieving precision medicine for many psychiatric disorders [13]. In recent years, neuroimaging studies have been published utilizing a variety of methods, including electroencephalogram (EEG), brain volumetric magnetic resonance imaging, functional magnetic resonance imaging, and diffusion tensor imaging to identify biomarkers for treatment response to antidepressants [14]. Using the EEGs can be an effective and relatively inexpensive method for studying developmental changes in brain-behavior relationships, and its high temporal resolution makes it particularly useful for examining neural activity flow in the nervous system [15], [16].

Recently, many studies have focused on the use of EEG to predict how a patient will respond to antidepressant medication in order to overcome this problem [17]–[20]. As an example, patients who responded to treatment demonstrated improved absolute alpha power at baseline, which can be used

as a biomarker to predict treatment response [21]. Also, the interhemispheric neural activity flow in the temporal lobe exhibits 99.61% classification capability using only four EEG channels [22]. A study conducted by Mumtaz et al. involved extracting time-frequency features from different frequency bands of EEG signals and classifying them by using three time-frequency decomposition techniques, including wavelet transforms, short-time Fourier transforms, and empirical modes of decompositions, in order to predict treatment-outcome for MDD patients. Combining the best features from the decomposition methods described above provided a classification accuracy of 91.6% [23]. Likewise, Jaworska et al. utilized demographical features in conjunction with EEG data in order to improve the classification results [19]. As demonstrated by Salle et al., changes in theta cordance of the prefrontal and midline right frontal in the first week of treatment can provide a predictive indicator of the response to antidepressants [20]. Additionally, Kautzky et al. used a random forest approach to correctly identify 25% of patients with treatment-resistant depression based on clinical variables and three polymorphisms [24]. Moreover, Patel et al. predicted an 89% treatment response using various biometrics, including demographic information and structural and functional imaging features [25].

This paper makes significant contributions to the field of predicting treatment outcomes in MDD through the innovative utilization of neural activity flow based on the direct directed transfer function (dDTF). Firstly, we demonstrate that neural activity flow features, particularly those derived from the dDTF, can serve as accurate predictors of antidepressant response in MDD patients, providing insight into differentiating between individuals who positively respond to selective serotonin reuptake inhibitors (SSRI) and those who do not. Secondly, our work achieves new state-of-the-art accuracy in EEG-based prediction for MDD treatment by incorporating neural activity flow as a feature in machine learning models, including Support Vector Machines (SVMs), Linear Discriminant Analysis (LDA), Decision Trees (DT), and Random Forests (RF), surpassing existing benchmarks and enhancing the potential clinical applicability of our findings. Lastly, our analysis identifies specific brain regions and networks that are indicative of treatment failure in MDD, contributing to our understanding of the neural underpinnings of treatment outcomes and offering critical insights for the development of targeted interventions. In summary, our paper presents a novel and comprehensive approach to predicting treatment outcomes in MDD, leveraging neural activity flow and machine learning models. Our contributions include the accurate prediction of antidepressant response, achieving state-of-the-art accuracy in EEG-based prediction, and identifying specific neural correlates of treatment failure, collectively representing a significant step forward in the field and providing valuable insights and tools for advancing personalized treatment strategies for individuals with MDD.

Materials and Methods

Dataset

The EEG signal dataset used in this study was provided by [23]. The data sets included 34 MDD patients and 30 healthy individuals. Among the 34 MDD patients, 17 men and 17 women have an average age of 40.3 ± 12.9. For the eyes closed condition, only 30 EEG segments, of 19 channels each, were available, which are used in this study. MDD patients were diagnosed using DSM-IV criteria [26]. An MDD patient was treated for four weeks with Selective Serotonin Reuptake Inhibitors (SSRIs) antidepressants. If there is a 50% improvement from pre- to post-treatment, the MDD patient is considered a responder; otherwise, the subject is considered a non-responder. As shown in Table 1, based on the Beck Depression Inventory, 12 patients responded to treatment, while 18 patients showed no significant improvement. The above study has been approved by the Human Ethics Committee of the Hospital Universitii Sains Malaysia, Kelantan, Malaysia.

According to the 10-20 electrode placement system, the EEG is recorded for five minutes using a 19-electrode EEG cap with linked-ear references. Five different brain regions are represented by electrodes: the frontal lobe containing Fp1, F3, F7, Fz, Fp2, F4, and F8, the parietal lobe containing P3, Pz, and P4, the occipital lobe containing O1 and O2, the left and right temporal lobe containing T3, T4, T5, T6 electrodes, and finally the central lobe with C3, C4, and Cz electrodes.

EEG preprocessing

In order to prevent erroneous subsequent analysis and ensure that the underlying neuronal activity is accurately reflected in the data, the pre-processing steps have been carried out using the EEGLAB open-source toolbox. To remove baseline drift, a 1 Hz high-pass filter is first applied. Then, the CleanLine open-source plugin is used to remove line noise. Lastly, 3 minutes of data are used for further analysis.

Effective Connectivity

The concept of effective connectivity or neural activity flow refers to the influence a node has over another based on a model of neuronal integration, which identifies neuronal coupling mechanisms [27]. Among the first models used to establish causality between two time series is Granger causality, which was introduced in economics. As explained by Granger causality, a time series of X_1 causes a time series X_2, if knowledge of X_1 helps to makes predictions of X_2 more accurate [28]. A measure of brain activity associated with Granger causality is the directed transfer function (DTF). The DTF represents a linear combination of causal influences along all causal pathways, direct and indirect, beginning at one site and ending at another [29]. In order to distinguish direct from indirect flows, a

dDTF is proposed [30]. With the dDTF method, the strength and direction of direct flow of neural activity is determined using DTF combined with partial coherence. For the purpose of calculating the dDTF, the Source Information Flow Toolbox (SIFT) is used [31], [32]. Through SIFT, each subject's EEG data was divided into 18 segments, each lasting 10 seconds. A Multivariate Autoregressive (MVAR) model of order 20 was then fitted to the data, satisfying two criteria of stability and consistency. This indicates that the model produces data with the same correlation structure as the actual EEG data and is stable/stationary. This step is crucial in ensuring the accuracy and reliability of subsequent analyses. Following that, dDTF values are calculated for each segment of data across all frequency ranges, and since we have 19 electrodes, a matrix with the shape of 19 * 19 * frequency is obtained. At the end of the process, the delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-45 Hz) bands are extracted.

To provide an overview of the dDTF method, the EEG signal is first fitted with a multivariate autoregressive model (MVAR). Then, to model a k-channel process, X(t) is modeled as follows:

$$X(t) = \left(X_1(t), X_2(t), \dots, X_k(t)\right) \quad (1)$$

This would lead to the following expression for the MVAR model:

$$X(t) = \sum_{j=1}^{p} A(j)X(t-j) + E(t) \quad (2)$$

In the above equation, $X(t)$ represents the data vector in time t, $E(t)$ represents the white noise vector, $A(i)$ represents the model coefficients, and p represents the order of the model. After that, as a result of converting the model equation into a frequency domain, we obtain:

$$X(f) = A^{-1}(f)E(f) = H(f)E(f) \quad (3)$$

In the above equation, X(f) is the input signal, E(f) is white noise, and H(f) matrix is referred to as the transfer matrix of the system which f denotes the frequency of the input signal. The DTF can be defined as follows in accordance with the transfer function of MVAR:

$$DTF_{j\to i}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum_{m=1}^{k} |H_{im}(f)|^2} \quad (4)$$

For the dDTF formula, the DTF must be modified with partial coherence, as follows:

$$S(f) = H(f)VH^*(f) \quad (5)$$

$$pCoh_{ij}^2(f) = \frac{M_{ij}^2(f)}{M_{jj}(f)M_{ii}(f)} \quad (6)$$

where $S(f)$ is power spectra, V is the variance of the noise $E(f)$ and M_{ij} is spectral matrix S by removing i-th row and j-th column. Finally, dDTF is defined by the given formula:

$$dDTF_{ij}(f) = DTF_{ij}(f)pCoh_{ij}(f) \quad (7)$$

Feature selection

To select the best features for discriminating between responder and non-responder groups, first one-seventh of the data was set aside for testing and then LDA was used for calculating the area under a curve (AUC) of every neural activity flow in each band. In this study, the AUC has been used since the area under the receiver operating characteristic curve (AUC-ROC) is equivalent to the Mann–Whitney U-statistic [33]. ROC curves are calculated by comparing the model's false-positive rate against its true-positive rate across a range of thresholds. AUC-ROC value was obtained based on the mean values of all cross-validation sets. The mean AUCs are a valid measure of the model's performance in a generalized setting in which the model was trained, given that each of the analyzed learners received a unique training set and a unique model-external validation dataset during training. Next, the top 30 connections from each band with the highest AUC are selected, then the feature selection algorithms are applied.

This paper uses three feature selection methods. The first is based on the area AUC-ROC. In this method, a subset of features is assessed empirically by measuring the prediction accuracy of the feature subset selected by our method. In other words, the forward selection is an iterative process in which we start without any features in the model. We continue to add new features to our model in each iteration, and then we select the subset of features with the highest accuracy out of all the others[34]. Second, the minimum-redundancy maximum-relevance (mRMR) algorithm has been used to rank features to minimize redundancy while maximizing relevance. The mRMR algorithm uses mutual information to compute similarity scores between features and labels of a subset, aiming to minimize the average mutual information between two features and to maximize the average mutual information between each feature and the specific label [35]–[37]. As for the last method, Relief-F is used. In a similar way to k-nearest neighbors, Relief-F assigns weights to each feature based on its ability to separate class labels. If the squared Euclidean distance between a feature and its nearest instances of the same class is greater than the distance between the two instances of the other class, the weight of the feature decreases. Based on Manhattan distance, Relief-F calculates both negative and positive weights for each feature [38], [39].

Classification

In artificial intelligence, supervised learning refers to a subcategory of ML, which uses labeled datasets to train algorithms that are capable of classifying data or predicting outcomes. Support vector machines (SVMs) are supervised learning algorithms used to classify two groups of data. The

algorithms draw lines (hyperplanes) to separate groups based on their patterns. An SVM builds a learning model that assigns new examples to one group or another. As a result of these functions, SVMs are called non-probabilistic binary linear classifiers. This paper also uses a random forest, which is composed of many individual decision trees. Trees in the forest generate a class prediction, and the class with the most votes determines the class prediction for our model. In addition, the LDA classifier is being used to classify two groups using a linear combination of features.

Statistical analysis

For the purpose of evaluating each neural activity flow's importance, the AUC value has been calculated for each neural activity flow. Afterward, the 30 top connections with the highest value were selected. Also, for the evaluation of any machine learning model's performance, we need to test it on some unseen data. Based on the model's performance on unseen data, we can say whether our model is Under-fitting/Over-fitting/Well generalized. A cross-validation (CV) procedure is used to assess the effectiveness of machine learning models; it can also be used to evaluate a model if we have insufficient data. For a CV to be performed, a portion of the training data must be kept aside for evaluation later. In this paper, for CV, the k-fold method has been used. During k-fold cross-validation, the original sample is divided into k subsamples of equal size. A single subsample of the k subsamples is retained as the validation data for testing the model, while the remaining k-1 subsamples are used as training data for training the model. In the process of trial and error, it has been determined that 7 is the optimal value for k. Further analysis was conducted based on the results of the 7-fold CV.

Overview of the proposed method

Figure 1 illustrates the proposed method. In the first step, raw EEG data is preprocessed using EEG-Lab, an open-source toolbox. As part of the preprocessing steps, a high-pass filter with a 1 Hz frequency was applied as well as CleanLine noise. Afterward, the signals are then divided into 18 segments, each lasting 10 seconds. The neural activity flow is calculated from each segment, and a matrix of 19-channels * 19-channels * 45-frequency steps is obtained. Then, the delta, theta, alpha, beta, and gamma bands are extracted by averaging over the frequency ranges of 1-4 Hz, 4-8 Hz, 8-13 Hz, 13-30 Hz, and 30-45 Hz. Following this, AUC-ROC forward feature selection, mRMR, and Relief-F algorithms are used to find the best features from all frequency bands. Lastly, the selected features are used to train SVM, LDA, RF, and DT classifiers.

Results

Topographic maps

In Figure 2, topographic maps are shown for MDD patients in the five frequency ranges of delta, theta, alpha, beta, and gamma. It is clear from this figure that the responder group shows lower delta power compared to the non-responder group, whereas the responders show higher beta power. In the theta band, the most significant difference is observed in the left temporoparietal lobe, whereas in the alpha band, the most significant difference is observed in the central areas. As can be observed in the beta band, respondents showed higher power in general, particularly in the left temporal lobe, which is also observed in gamma.

Regional Differences in Responders vs. Non-Responders

Following the calculation of the neural activity flow matrix for each segment, the average in responders and non-responder groups is calculated (Figure 3). Then, based on the dDTF values corresponding to each directed connection, the AUC values are calculated for each of them independently. Afterward, they're ranked according to their top 30 of AUC values (Table 2 and Figure 4). To identify the most important regions, Table 4 summarizes each region in terms of frequency. According to this table, the frontal lobe has the highest number of neural activity flows, followed by the temporal and parietal lobes. As a result, the frontal and parietal lobes are the dominant regions for the majority of connections. Moreover, Table 3 which is an overview of all electrodes from the top 30 bands, shows that most of the connections end in specific regions, particularly in the electrodes F8, Pz, T5, and P4.

Classification Responder based on the neural activity flow

Table 6 shows the classification results of ML models for each frequency band, as well as a combination of features from all bands, consisting of 150 connections (30 connections from each band). Clearly, the top 150 features have the highest accuracy, specificity, sensitivity, and F1-measure in every model. Following that, the beta and alpha bands yield the highest results. As can be seen, most models have a higher specificity than sensitivity, which indicates that the models are more capable of correctly identifying patients who will respond to the treatment.

It is important to note that one of the main problems with ML models is the curse of dimensionality, and it basically means that the error becomes larger as the number of features increases. To overcome this problem, different feature selection algorithms have been used, including mRMR, Relief-F, and forward feature selection algorithms. Based on Table 7, it can be observed that forward feature selection achieved the highest accuracy. Figure 5 shows how the highest performance can be obtained using only 23 features and that accuracy decreases afterward, and the selected features are shown in

Table 5. Thus, forward feature selection based on the ROC-AUC algorithm improved the accuracy of classification when using the best subset of features.

Discussion

In this study, we demonstrate a ML approach using EEG-derived neural activity flows that accurately predicts antidepressant response and provides neuroscientific insights into mechanisms of treatment outcomes. Specifically, we extract dDTF effective connectivity biomarkers in MDD patients to capture differences between treatment responders and non-responders across brain regions and frequencies. Our findings indicate frontoparietal network connectivity at alpha and beta bands underlies response failures, aligning with cognitive theories implicating this circuitry. By combining dDTF neural activity flow features with SVM classifiers, our model significantly improves predictive performance over previous state-of-the-art EEG methods, achieving over 98% accuracy. This establishes functional connectivity as an informative biomarker for guiding antidepressant selection while elucidating network deficits linked to treatment resistance.

Based on the findings in Table 3, F8, Pz, T5, and P4 are the most important regions that differ between respondents and non-responders. Further, neural activity flows were used to predict treatment outcomes, as well as different feature selection algorithms, were applied to improve classification results. As a result, by using forward feature selection across all frequency bands, the best accuracy of 98.52% was achieved by SMV (Table 7). In spite of the fact that SVMs perform better in very high-dimensional spaces and SVM models have generalizability in practice, the risk of overfitting is lower in SVMs. Also, for its parameters, a unique global optimum can be easily determined [40], [41]. Additionally, SVM is based on its kernel, and by selecting the appropriate kernel function, any complex problem can be resolved, and in this study, the best result is achieved by using the Radial Basic Function (RBF). Also, there have recently been some studies that have criticized the reproducibility of AI methods because the evaluation methods may be incorrect, and many of them may suffer from data leakage or overfitting[42]. As discussed, we have used CV to train the model and evaluate it on the entire dataset to overcome these problems. By fitting the model in every step and estimating its performance independently in each fold of the CV procedure, we can identify problems such as overfitting or selection bias and learn how the model will generalize to an independent dataset since they give an almost unbiased performance. Overfitting was also overcome by using different feature selection methods.

As shown in Table 3, which outlines the starting and ending points of the top 30 activity flows across all frequency bands, only a few areas have the greatest impact on treatment outcomes. First and foremost, the frontal lobe region, especially the F8, is a dominant region for most of the neural activity flows. Following that, the parietal (Pz and P4) and temporal (T5) lobes are dominant regions. Additionally, the findings indicate that this pattern represents a valuable brain biomarker that could be

used to assess the treatment response of MDD patients before they begin their treatments, thereby reducing costs and reducing the amount of time spent on patients and medical centers. For example, F8 has also been shown to have a high classification capability [43]. Also, it may be possible to predict the efficacy of SSRIs by analyzing frontal EEG recordings collected during the first week of treatment [44].

Table 5 illustrates the SVM-selected features for obtaining the best accuracy. With only 23 features, 98.5% accuracy has been achieved, and most of the features are selected from beta and alpha bands. Further, when each frequency was used separately to predict treatment outcome, beta, and alpha bands produced the best results. Whether we use the neural activity flows of each frequency band separately or combine all features in all frequency bands, we conclude that the beta and alpha bands are better discriminators for predicting a person's response to treatment.

A flexible and coordinated modulation of cognitive and emotional processes is enabled by the frontoparietal network (FPN), composed of lateral prefrontal and posterior parietal cortices [45]. It has been shown that the FPN is activated during externally focused attention and goal-oriented task performance. A defining characteristic of MDD is the deficiency in concentration, and cognitive theories suggest that impaired top-down regulation of aberrant emotional processing perpetuates a bias toward negative effects [46]. It has been shown that depression symptoms are associated with decreased neural activity flows between the FPN and other parts of the brain [47]. Based on the results of this study, it appears that the most significant difference between these two groups can be traced to the frontal and parietal lobes (Table 4). Study results supporting our findings suggest that the beta frequency of default mode network-FPN might serve as a neural marker for reoccurring illness[48], [49]. Further studies have shown that EEG beta power correlates with cortisol secretion and attentional processing, as can be seen in Figure 2, where higher beta power is observed in the responding group mainly in regions related to FPN. Moreover, Baskarana et al. showed that changes in beta asymmetry observed at 2 weeks post-treatment in the responding group may reflect differences in arousal induced by antidepressants [50]. Finally, this paper's findings suggest that treatment failure results from alpha and beta-frequency frontoparietal networks at the network level.

Although this simple approach to dDTF analysis enabled the classification of the treatment responses with 98% accuracy, it may not allow inference about the neural activity flow of particular frequency bands (delta, theta, alpha, beta, gamma, etc.) within the neural system. To ensure a good representation of low-amplitude (higher-frequency) rhythms in MVAR modeling, the analysis should be done step by step by filtering out high-amplitude (low-frequency) rhythm(s) (by high-pass filter with a gentle slope, not steep), and by fitting a separate MVAR model to the filtered signals.

The proposed model achieved higher accuracy than other studies in predicting antidepressant treatment outcomes based on EEG signals by combining neural activity flow and forward feature

selection, as shown in Table 8. This indicates that, in addition to showing causality and assisting in understanding the main cause of treatment outcomes, it can also improve classification results. Also, several limitations should be taken into consideration, including the small number of patients from one location that may affect the generalizability of our findings. Moreover, this study used only neural activity flow on channels, but future work may calculate neural activity flow on the brain source localization or combine different types of features. Also, our paper has only used machine learning models. However, with the advancements in deep learning models, it would be beneficial to use them in order to improve prediction results.

Conclusions

In this study, we investigated a novel method for classifying treatment responses in MDD patients based on neural activity flows. By using neural activity flows, altered brain activity can be identified that causes TRD. Based on the findings of this study, it was demonstrated that the most important neural flows that differ between responders and non-responders are related to the frontal and parietal lobes at beta frequency, which suggests that the FPN is mostly involved in treatment response. Also, using this kind of neural flow as an input feature in an SVM model and forward feature selection alongside, we were able to accurately classify responders and non-responders with an accuracy of 98%. The results of this study suggest that ML models can be useful in predicting an individual's response to antidepressants at the beginning of a treatment program.

Acknowledgment:

This research is financially supported by “Shahid Beheshti University of Medical Sciences” (Grant No.).

References

[1] W. H. Organization, 2017 “Depression and other common mental disorders: global health estimates,” World Health Organization,

[2] C. Otte et al. 201., Major depressive disorder,” Nat Rev Dis Primers, vol. 2, no. 1, 1–20.

[3] S. M. Stahl and M. M. Grady, 2003 . Differences in mechanism of action between current and future antidepressants, J Clin Psychiatry, vol. 64, 13–17.

[4] A. Sambunaris, J. K. Hesselink, R. Panagides, J. Panagides, and S. M. Stahl, 1997. Development of new antidepressants, Journal of Clinical Psychiatry, vol. 58, no. 6, 40–53.

[5] Z. D. Cohen and R. J. DeRubeis, 2018. Treatment Selection in Depression, Annu Rev Clin Psychol, vol. 14, no. 1, 209–236.

[6] C. A. Webb et al., 2019. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study, Psychol Med, vol. 49, no. 07, 1118–1127.

[7] M. T. Berlim, M. P. Fleck, and G. Turecki, 2008. Current trends in the assessment and somatic treatment of resistant/refractory major depression: An overview, Ann Med, vol. 40, no. 2, 149–159.

[8] G. Arteaga-Henríquez et al., 2019 “Low-Grade Inflammation as a Predictor of Antidepressant and Anti-Inflammatory Therapy Response in MDD Patients: A Systematic Review of the Literature in Combination With an Analysis of Experimental Data Collected in the EU-MOODINFLAME Consortium,” Front Psychiatry, vol. 10.

[9] A. F. Leuchter, I. A. Cook, A. M. Hunter, and A. S. Korb, 2009. “A new paradigm for the prediction of antidepressant treatment response,” Dialogues Clin Neurosci, vol. 11, no. 4, 435–446.

[10] B. Schwartz, Z. D. Cohen, J. A. Rubel, D. Zimmermann, W. W. Wittmann, and W. Lutz, 2021. “Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy,” Psychotherapy Research, vol. 31, no. 1, 33–51.

[11] V. Bremer et al., 2018. “Predicting Therapy Success and Costs for Personalized Treatment Recommendations Using Baseline Characteristics: Data-Driven Analysis,” J Med Internet Res, vol. 20, no. 8, p. e10275.

[12] L. S. Goldman, N. H. Nielsen, and H. C. Champion, 1999. “Awareness, diagnosis, and treatment of depression,” J Gen Intern Med, vol. 14, no. 9, 569–580.

[13] S.-G. Kang and S.-E. Cho, 2020. “Neuroimaging Biomarkers for Predicting Treatment Response and Recurrence of Major Depressive Disorder,” Int J Mol Sci, vol. 21, no. 6, p. 2148.

[14] A. H. Kemp, E. Gordon, A. J. Rush, and L. M. Williams, 2008. “Improving the Prediction of Treatment Response in Depression: Integration of Clinical, Cognitive, Psychophysiological, Neuroimaging, and Genetic Measures,” CNS Spectr, vol. 13, no. 12, pp. 1066–1086.

[15] M. A. Bell and K. Cuevas, 2012. “Using EEG to Study Cognitive Development: Issues and Practices,” Journal of Cognition and Development, vol. 13, no. 3, pp. 281–294.

[16] D. J. A. Smit, C. J. Stam, D. Posthuma, D. I. Boomsma, and E. J. C. de Geus, 2008. "Heritability of 'small-world' networks in the brain: A graph theoretical analysis of resting-state EEG functional connectivity," Hum Brain Mapp, vol. 29, no. 12, 1368–1378.

[17] A. Khodayari-Rostamabad, J. P. Reilly, G. M. Hasey, H. de Bruin, and D. J. MacCrimmon, 2013. "A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder," Clinical Neurophysiology, vol. 124, no. 10, pp. 1975–1985.

[18] MS Shahabi, A Shalbaf. Prediction of Treatment Outcome in Major Depressive Disorder using Ensemble of Hybrid Transfer Learning and Long Short Term Memory based on EEG Signal. IEEE Transactions on Cognitive and Developmental Systems. 15 (3). pp.1279 – 1288. 2023

[19] N. Jaworska, S. de La Salle, M. H. Ibrahim, P. Blier, and V. Knott, 2019. "Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data," Front Psychiatry, vol. 10, no. p. 768.

[20] Sara de la Salle, Natalia Jaworska, Pierre Blier, Dylan Smith, Verner Knott, 2020. "Using prefrontal and midline right frontal EEG-derived theta cordance and depressive symptoms to predict the differential response or remission to antidepressant treatment in major depressive disorder," Psychiatry Res Neuroimaging, vol. 302, p. 111109.

[21] A. Baskaran et al., 2018. "The comparative effectiveness of electroencephalographic indices in predicting response to escitalopram therapy in depression: A pilot study," J Affect Disord, vol. 227, pp. 542–549.

[22] Y. Zhang, K. Wang, Y. Wei, X. Guo, J. Wen, and Y. Luo, 2022. "Minimal EEG channel selection for depression detection with connectivity features during sleep," Comput Biol Med, vol. 147, p. 105690.

[23] W. Mumtaz, L. Xia, M. A. M. Yasin, S. S. A. Ali, and A. S. Malik, "A wavelet-based technique to predict treatment outcome for Major Depressive Disorder," PLoS One, vol. 12, no. 2, Feb. 2017, doi: 10.1371/journal.pone.0171409.

[24] A. Kautzky et al., 2015. "The combined effect of genetic polymorphisms and clinical parameters on treatment outcome in treatment-resistant depression," European Neuropsychopharmacology, vol. 25, no. 4, 441–453.

[25] M. J. Patel, C. Andreescu, J. C. Price, K. L. Edelman, C. F. Reynolds, and H. J. Aizenstein, 2015. "Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction," Int J Geriatr Psychiatry, vol. 30, no. 10, 1056–1067.

[26] S. B. GUZE, 1995. "Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV)," American Journal of Psychiatry, vol. 152, no. 8, 1228–1228,.

[27] Z. Liu et al., 2017. "Effective connectivity analysis of the brain network in drivers during actual driving using near-infrared spectroscopy," Front Behav Neurosci, vol. 11, p. 211.

[28] C. W. J. Granger, 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," Econometrica, vol. 37, no. 3, p. 424.

[29] A. Korzeniewska, C. M. Crainiceanu, R. Kuś, P. J. Franaszczuk, and N. E. Crone, 2008. "Dynamics of event-related causality in brain electrical activity," Hum Brain Mapp, vol. 29, no. 10, 1170–1192.

[30] A. Korzeniewska, M. Mańczak, M. Kamiński, K. J. Blinowska, and S. Kasicki, 2003. "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," J Neurosci Methods, vol. 125, no. 1–2, 195–207.

[31] T. R. Mullen, 2014. The dynamic brain: Modeling neural dynamics and interactions from human electrophysiological recordings. University of California, San Diego.

[32] A. Delorme et al., 2011. "EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing," Comput Intell Neurosci.

[33] S. J. Mason and N. E. Graham, 2002. "Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation," Quarterly Journal of the Royal Meteorological Society, vol. 128, no. 584, 2145–2166.

[34] H. Mamitsuka, 2006. "Selecting features in microarray classification using ROC curves," Pattern Recognit, vol. 39, no. 12, 2393–2404.

[35] B. Şen, M. Peker, A. Çavuşoğlu, and F. v. Çelebi, 2014. "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms," Journal of Medical Systems 2014 38:3, vol. 38, no. 3, 1–21.

[36] N Amini, M Mahdavi, H Choubdar, A Abedini, A Shalbaf, R Lashgari. Automated prediction of COVID-19 mortality outcome using clinical and laboratory data based on hierarchical feature selection and random forest classifier. Computer methods in biomechanics and biomedical engineering 26 (2), 160-173

[37] C. Ding and H. Peng, 2005. "Minimum redundancy feature selection from microarray gene expression data," J Bioinform Comput Biol, vol. 3, no. 2, pp. 185–205.

[38] M. Peker, A. Arslan, B. Sen, F. v. Celebi, and A. But, 2015. "A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF)," INISTA 2015 - 2015 International Symposium on Innovations in Intelligent SysTems and Applications, Proceedings.

[39] A. Al-Nafjan, 2022. "Feature selection of EEG signals in neuromarketing," PeerJ Comput Sci, vol. 8, p. e944.

[40] G. N. Garcia, T. Ebrahimi, and J. M. Vesin, 2003. "Support vector EEG classification in the Fourier and time-frequency correlation domains," International IEEE/EMBS Conference on Neural Engineering, NER, 591–594.

[41] A. Subasi and M. I. Gursoy, 2010. "EEG signal classification using PCA, ICA, LDA and support vector machines," Expert Syst Appl, vol. 37, no. 12, 8659–8666.

[42] E. Gibney, 2022. "Could machine learning fuel a reproducibility crisis in science?," Nature 2022 608:7922.

[43] F. Hasanzadeh, M. Mohebbi, and R. Rostami, 2021. "Single Channel EEG Classification: A Case Study on Prediction of Major Depressive Disorder Treatment Outcome," IEEE Access, vol. 9, 3417–3427.

[44] D. v. Iosifescu et al., 2009. "Frontal EEG predictors of treatment outcome in major depressive disorder," European Neuropsychopharmacology, vol. 19, no. 11, pp. 772–777.

[45] M. A. G. Martens, N. Filippini, C. J. Harmer, and B. R. Godlewska, 2021. "Resting state functional connectivity patterns as biomarkers of treatment response to escitalopram in patients with major depressive disorder," Psychopharmacology (Berl).

[46] A. S. Fischer, C. J. Keller, and A. Etkin, 2016. "The Clinical Applicability of Functional Connectivity in Depression: Pathways Toward More Targeted Intervention," Biol Psychiatry Cogn Neurosci Neuroimaging, vol. 1, no. 3, 262–270.

[47] D. H. Schultz et al., 2018. "Global connectivity of the fronto-parietal cognitive control network is related to depression symptoms in the general population," Network Neuroscience, vol. 3, no. 1, 107–123.

[48] A. E. Whitton, S. Deccy, M. L. Ironside, P. Kumar, M. Beltzer, and D. A. Pizzagalli, 2018. "Electroencephalography Source Functional Connectivity Reveals Abnormal High-Frequency Communication Among Large-Scale Functional Networks in Depression," Biol Psychiatry Cogn Neurosci Neuroimaging, vol. 3, no. 1, 50–58.

[49] D. A. Pizzagalli, 2011. "Frontocingulate dysfunction in depression: toward biomarkers of treatment response," Neuropsychopharmacology, vol. 36, no. 1, 183–206.

[50] A. Baskaran et al., 2018. "The comparative effectiveness of electroencephalographic indices in predicting response to escitalopram therapy in depression: A pilot study," J Affect Disord, vol. 227, pp. 542–549.

[51] A. Zhdanov et al., 2020. "Use of Machine Learning for Predicting Escitalopram Treatment Outcome From Electroencephalography Recordings in Adult Patients With Depression," JAMA Netw Open, vol. 3, no. 1, e1918377–e1918377.

[52] N. van der Vinne, M. A. Vollebregt, N. N. Boutros, K. Fallahpour, M. J. A. M. van Putten, and M. Arns, 2019. "Normalization of EEG in depression after antidepressant treatment with sertraline? A preliminary report," J Affect Disord, vol. 259, pp. 67–72.

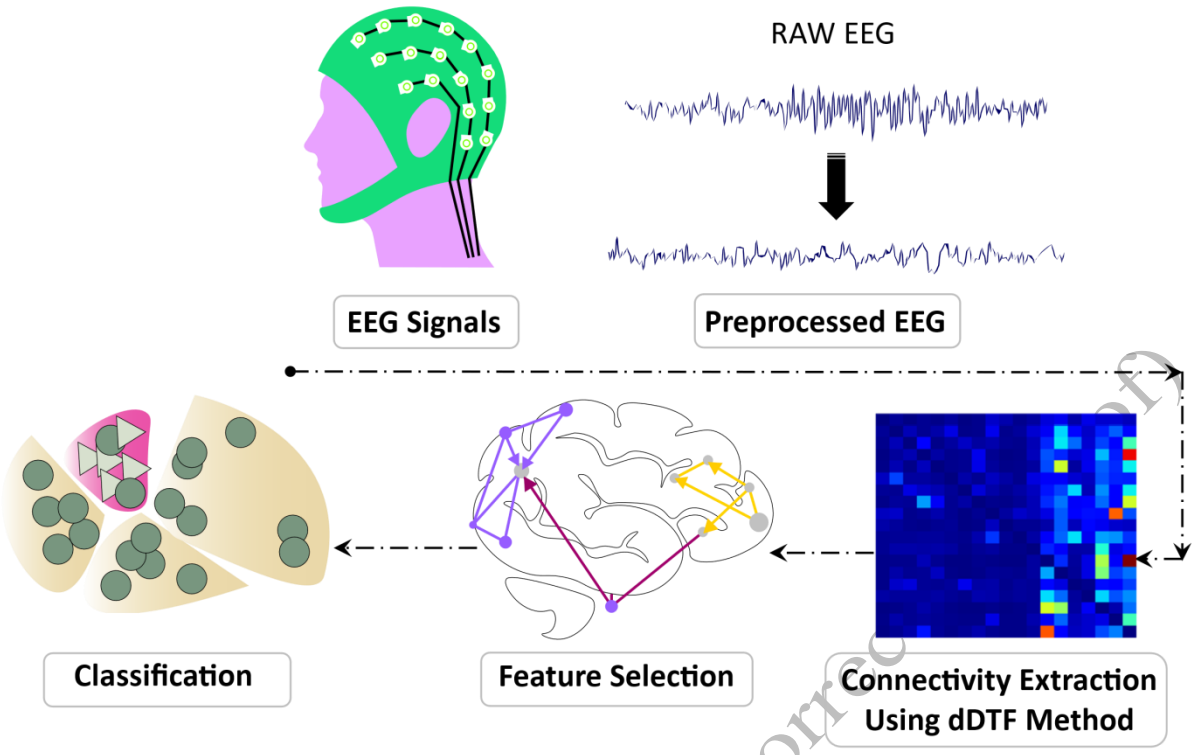

Figure 1 - An EEG signal is recorded in a 10-20 system and then preprocessed. In the following step, the dDTF matrix for the delta, theta, alpha, beta, and gamma frequency bands is calculated. Afterward, several feature selection methods are used to select the best neural activity flows. In the final step, the selected features are used to perform classification.

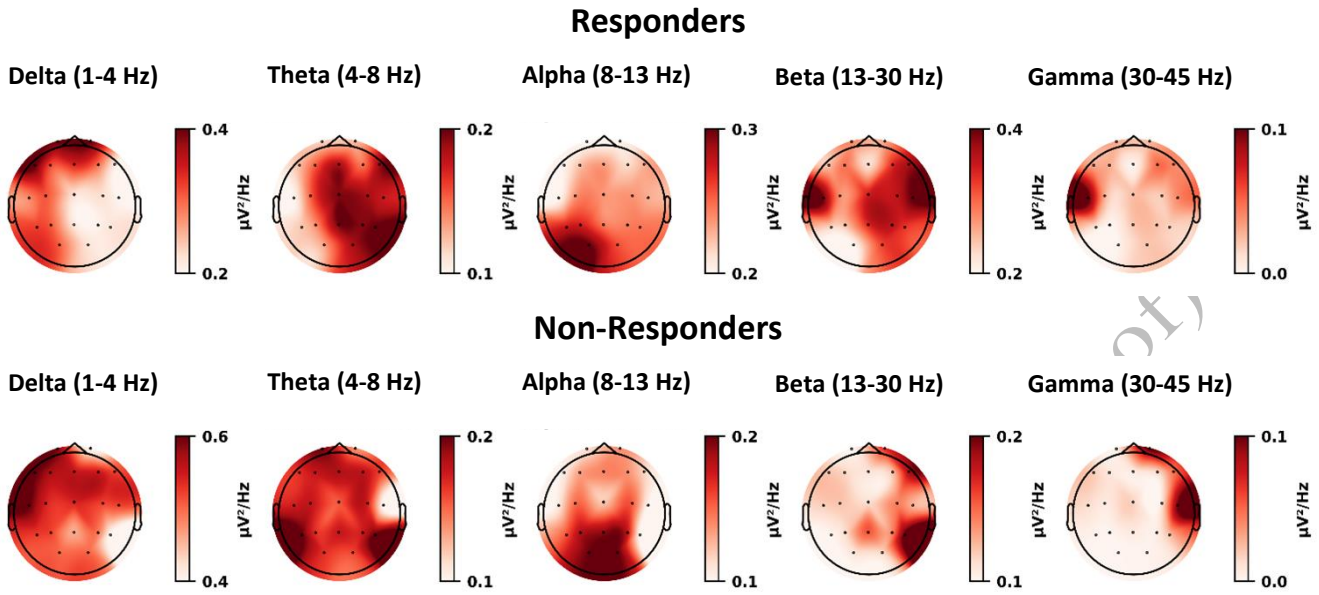

Figure 2 - The mean PSD of responders and non-responders in the delta, theta, alpha, beta and gamma bands.

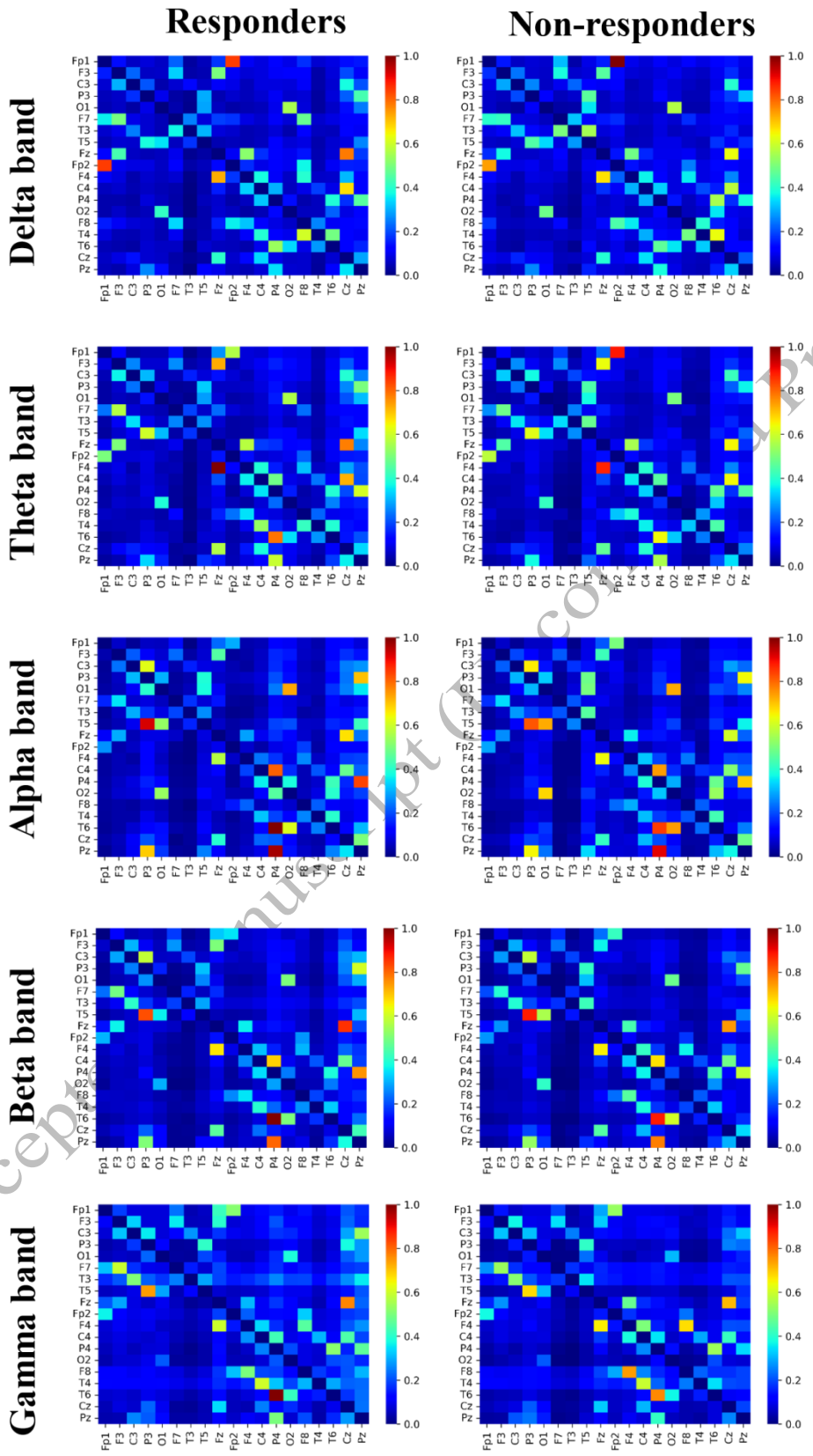

Figure 3. The normalized mean neural activity flow values using the dDTF method between MDD patients who responded to SSRIs and those who did not respond to the medications in delta, theta, alpha, beta, and gamma bands.

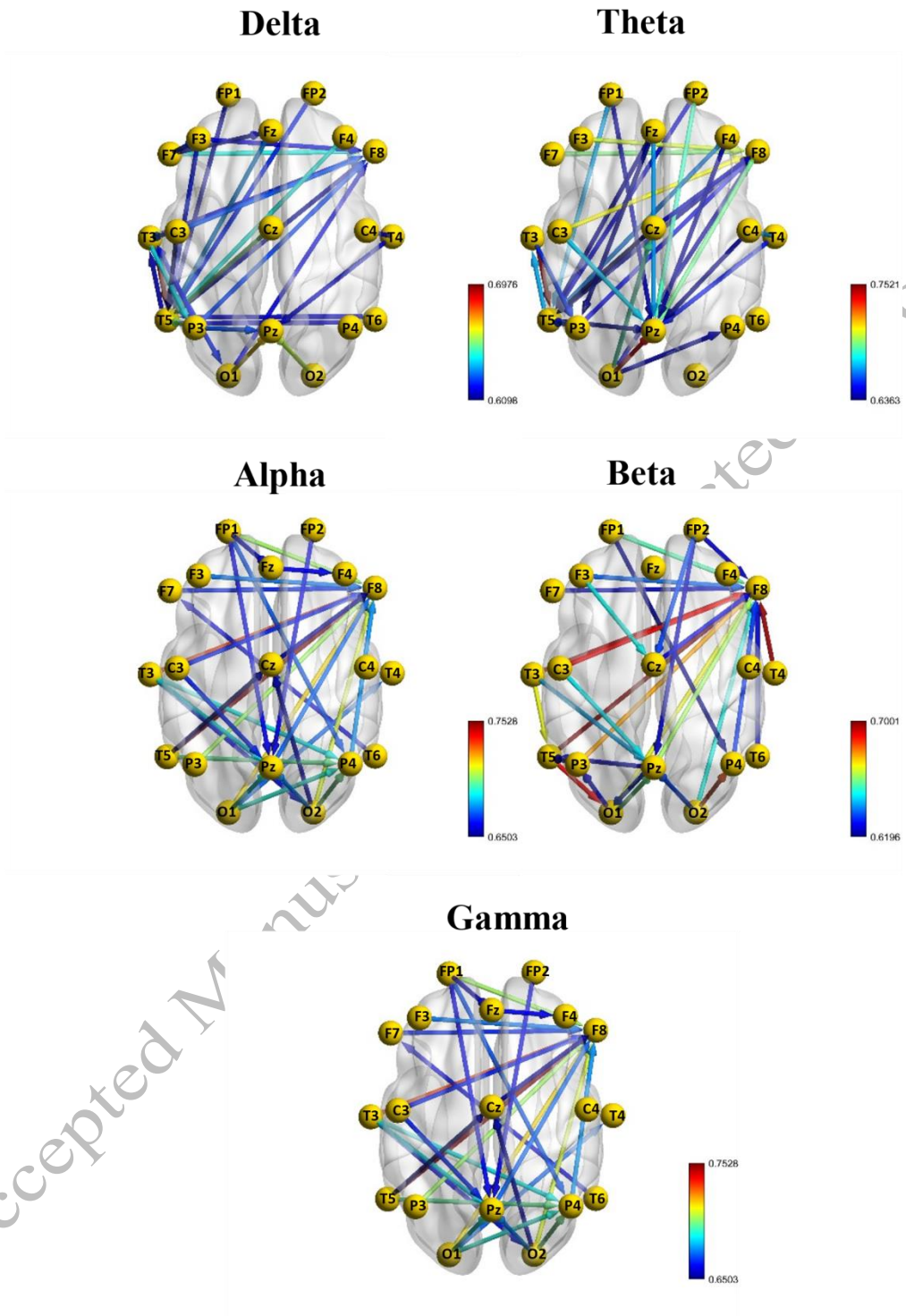

Figure 4. Based on AUC values, the top 30 neural activity flows that show differences in propagation between responders and nonresponders are illustrated. Nodes represent electors in the 10-20 system, and edges represent AUC values.

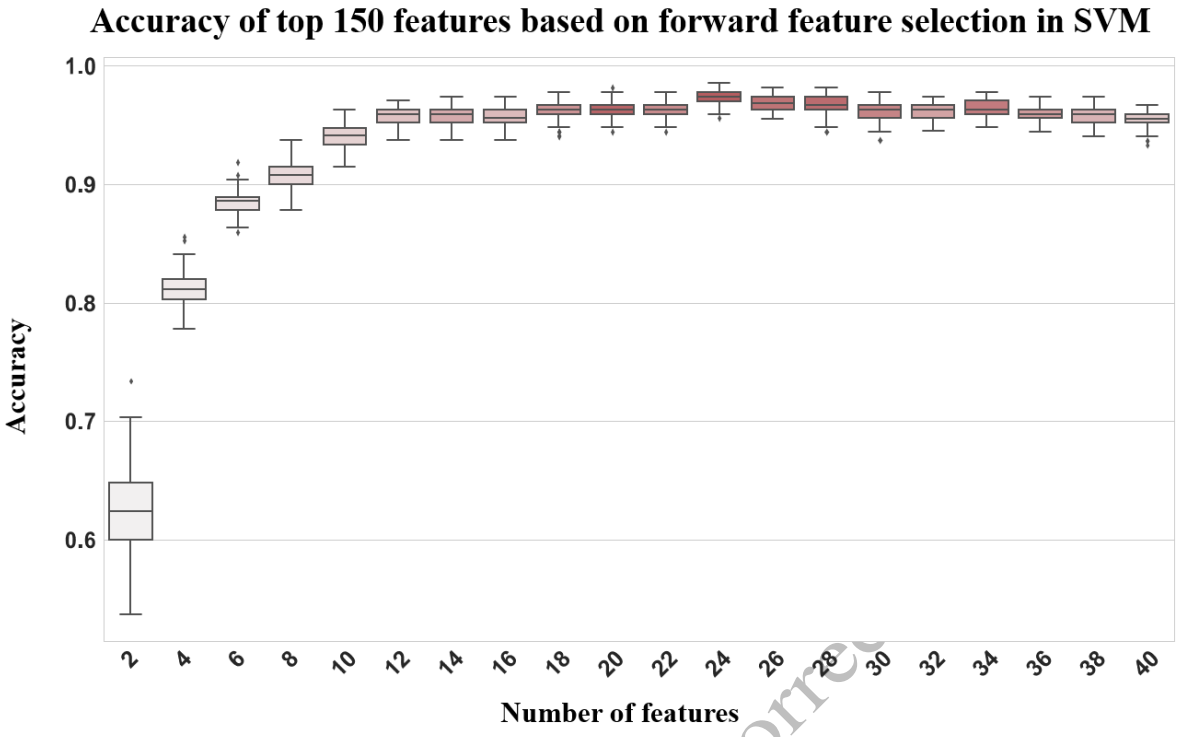

Figure 5. SVM forward feature selection based. The x-axis indicates the number of features used for classification, and the y-axis indicates the model's accuracy. With 23 features, the best accuracy was obtained and since there was no improvement in accuracy, the x-axis is restricted to 40 features.

Table 1 – Summary of MDD patient clinical characteristics

Information	Responder	Non-Responder	Total
Age [years]	40.7 (±13.0)	41.1 (±12.5)	40.3 (±12.9)
Gender (female/male)	8/8	9/9	17/17
Pretreatment BDI-II	18.4 (±7. 4)	22.8 (±12.5)	20.6 (±8.6)
Post-treatment BDI-II	9.1 (±6.3)	22.1 (±3.3)	15.6 (±4.5)

Table 2 - According to the AUC values of responding and non-responding groups, the following are the 30 most significant neural activity flows in the delta, theta, alpha, beta and gamma bands.

Delta			Theta			Alpha			Beta			Gamma		
From	To	AUC	From	To	AUC	From	To	AUC	From	To	AUC	From	To	AUC
T3	T5	0.698	T3	T5	0.752	T3	T5	0.763	T5	F8	0.700	T5	F8	0.753
O1	Pz	0.668	O1	Pz	0.739	T5	T3	0.680	T4	F8	0.698	T3	F8	0.733
O2	Pz	0.659	C3	F8	0.708	T6	F8	0.654	T3	F8	0.697	O1	F8	0.716
P3	T5	0.653	F3	F8	0.703	F8	C4	0.648	C3	F8	0.692	O2	F8	0.711
Cz	T5	0.652	F7	F8	0.694	Pz	F8	0.648	T5	O1	0.689	P3	F8	0.705
F4	T5	0.644	F8	Pz	0.691	O2	P4	0.647	O2	P4	0.685	Fp1	F8	0.704
F7	F8	0.643	Fp2	Pz	0.687	T3	O1	0.646	P3	F8	0.677	O2	P4	0.699
T3	P3	0.640	O1	Cz	0.686	T5	O1	0.642	T3	T5	0.668	T5	P4	0.697
Fz	T5	0.632	C3	Pz	0.672	F7	F8	0.640	O1	F8	0.668	O1	P4	0.691
C3	F8	0.627	T5	T3	0.670	O1	Cz	0.639	Pz	F8	0.662	O1	Pz	0.691
C3	T3	0.627	Fp1	T5	0.670	F3	F8	0.637	O1	Pz	0.658	T3	P4	0.688
P3	Pz	0.627	Fz	Pz	0.667	Fp1	F8	0.636	Fp1	F8	0.656	T3	Pz	0.686
P3	F8	0.625	T4	C4	0.666	T6	P4	0.636	O2	F8	0.651	P4	F8	0.676
F7	F3	0.624	F4	T5	0.663	O1	T5	0.635	F3	Cz	0.650	O2	Pz	0.674
T3	O1	0.622	T3	P3	0.655	F4	C4	0.633	C3	Pz	0.647	F3	F8	0.672
Fp2	T5	0.622	P3	T5	0.653	Fz	F4	0.630	T3	Pz	0.643	Pz	F8	0.671
P3	T3	0.621	T4	Pz	0.651	O1	P4	0.628	F3	F8	0.637	O1	T4	0.671
F3	T5	0.621	P3	F8	0.648	F8	F4	0.626	O2	Pz	0.633	Fp1	P4	0.669
T4	C4	0.618	Fp2	T5	0.645	C4	F4	0.626	Fp2	Cz	0.633	C3	F8	0.665
O1	F8	0.618	O1	P4	0.645	Cz	C4	0.625	Pz	O1	0.633	F7	F8	0.665
F7	Fz	0.618	Fz	T5	0.643	Fp2	F8	0.625	P4	F8	0.633	C3	Pz	0.665
T5	T6	0.617	O1	F8	0.640	P3	F8	0.625	Pz	T5	0.629	Fp1	Pz	0.662
F3	F8	0.617	T5	Pz	0.640	O2	T5	0.624	O1	P3	0.627	Fz	F4	0.662
T6	T5	0.616	Fp1	Pz	0.639	O2	T6	0.623	Cz	F8	0.626	Fp1	Fz	0.662
T3	F8	0.615	Cz	F8	0.639	P4	Pz	0.622	Fp2	F8	0.626	T3	O2	0.661
T4	Pz	0.614	F8	T5	0.639	T5	Cz	0.622	T6	F8	0.626	T6	F7	0.659
T5	T3	0.612	F4	Pz	0.639	Cz	F4	0.621	F7	F8	0.624	Fp2	Pz	0.659
T5	F8	0.611	Cz	T5	0.637	F7	T3	0.621	P3	T5	0.624	O2	Cz	0.659
Fp1	T5	0.611	Fz	P3	0.637	Fp1	T4	0.621	Fp2	Pz	0.623	T5	Cz	0.655
C3	T5	0.610	C4	Pz	0.636	Cz	T3	0.620	Fp1	P4	0.620	Cz	F8	0.650

Table 3 - An overview of electrodes in all frequency bands. On the left side, there are electrode names and the number of connections that originated from the electrode, and on the right side, there are electrode names and the number of connections from which connections ends.

Electrode's name	Number of connections from	Electrode's name	Number of connections to
O1	16	F8	47
T3	15	Pz	26
T5	13	T5	24
O2	11	P4	11
Fp1	11	Cz	7
P3	10	T3	7
C3	9	C4	5
Cz	8	F4	5
F7	8	O1	5
Fp2	8	P3	4
F3	7	Fz	2
Fz	6	T4	2
Pz	5	T6	2
T4	5	F3	1
T6	5	F7	1
F8	4	O2	1
F4	4		
P4	3		
C4	2		

Table 4- A summary of all frequency bands' connections. Rows represent the number of connections within each region, with C, F, O, P, and T reflecting the central, frontal, occipital, parietal, and temporal lobes of the brain and columns D, T, A, B, and G correspond to delta theta, alpha, beta, and gamma respectively. In this table, the From columns show the number of connections that originate from the specified region, and the To column shows the number of features that end in the specified region.

Region	From						To					
C (C3, C4, Cz)	D	T	A	B	G	Sum	D	T	A	B	G	Sum
	4	5	4	3	3	19	1	2	5	2	2	12
F (Fp1, Fp2, F3, F4, F7, F8, Fz)	D	T	A	B	G	Sum	D	T	A	B	G	Sum
	9	13	10	8	8	48	9	6	11	15	15	56
O (O1, O2)	D	T	A	B	G	Sum	D	T	A	B	G	Sum
	3	4	6	6	8	27	1	0	2	2	1	6
P (P3, P4, Pz)	D	T	A	B	G	Sum	D	T	A	B	G	Sum
	4	2	3	6	3	18	5	13	4	8	11	41
T (T3, T4, T5, T6)	D	T	A	B	G	Sum	D	T	A	B	G	Sum
	10	6	7	7	8	38	14	9	8	3	1	35

Table 5 - The selected features use the forward feature selection algorithm using the SVM model. The features are neural activity flows based on the dDTF method in all frequency bands.

From	To	AUC	Band
T3	T5	0.763	alpha
T6	P4	0.636	alpha
T5	O1	0.689	beta
O2	P4	0.699	gamma
T5	F8	0.700	beta
P3	Pz	0.627	delta
T5	O1	0.642	alpha
F3	Cz	0.650	beta
F4	C4	0.633	alpha
O1	P3	0.627	beta
T3	O1	0.646	alpha
T3	O1	0.622	delta
C4	F4	0.626	alpha
Pz	F8	0.662	beta
F7	T3	0.621	alpha
P4	F8	0.633	beta
T4	C4	0.666	theta
P3	T5	0.624	beta
Fp2	Cz	0.633	beta
T5	F8	0.611	delta
P3	F8	0.677	beta
Fz	F4	0.662	gamma
F7	F8	0.665	gamma

Table 6 - A comparison of SVM, LDA, RF, and DT classification results in the delta, theta, alpha, beta, and gamma bands.

BAND	Model	Accuracy	Specificity	Sensitivity	F1-Measure
DELTA	SVM	84.81% (±7.33)	87.43% (±8.81)	73.25% (±14.16)	79.04% (±10.45)
	LDA	81.46% (±8.65)	81.43% (±6.40)	69.09% (±21.11)	73.26% (±15.38)
	RF	80.37% (±8.00)	79.87% (±8.92)	66.51% (±16.99)	71.84% (±13.49)
	DT	75.55% (±6.26)	75.51% (±9.07)	59.32% (±16.00)	65.08% (±10.36)
THETA	SVM	84.82% (±4.22)	89.34% (±9.30)	71.84% (±8.09)	79.04% (±5.55)
	LDA	81.84% (±3.72)	82.43% (±5.79)	70.87% (±11.33)	75.38% (±5.25)
	RF	82.96% (±2.35)	91.01% (±8.50)	64.88% (±8.84)	74.92% (±4.44)
	DT	74.09% (±4.27)	66.92% (±5.71)	68.19% (±9.83)	67.34% (±6.86)
ALPHA	SVM	86.26% (±4.42)	91.87% (±7.70)	73.57% (±10.17)	80.96% (±5.82)
	LDA	78.85% (±4.22)	81.27% (±6.95)	62.32% (±7.42)	70.09% (±5.09)
	RF	87.06% (±7.68)	97.86% (±3.49)	69.66% (±18.57)	79.88% (±13.38)
	DT	77.00% (±8.98)	74.18% (±13.12)	68.78% (±10.72)	70.69% (±10.05)
BETA	SVM	87.04% (±3.87)	89.85% (±4.77)	77.38% (±8.89)	82.71% (±4.41)
	LDA	82.96% (±6.27)	84.59% (±8.28)	71.00% (±9.57)	76.94% (±8.22)
	RF	87.04% (±6.42)	92.40% (±7.27)	74.50% (±11.03)	82.14% (±8.66)
	DT	76.24% (±4.97)	74.41% (±9.19)	64.27% (±13.12)	67.77% (±7.83)
GAMMA	SVM	77.45% (±7.23)	76.68% (±11.79)	62.40% (±12.63)	68.29% (±11.30)
	LDA	70.40% (±6.47)	67.06% (±6.91)	53.73% (±15.17)	58.21% (±11.00)
	RF	80.81% (±8.03)	80.57% (±13.68)	68.51% (±12.38)	73.49% (±11.95)
	DT	66.64% (±5.90)	60.67% (±9.12)	55.60% (±13.74)	56.38% (±8.19)
TOP150	SVM	93.35% (±5.96)	96.43% (±6.56)	87.84% (±12.68)	91.22% (±8.33)
	LDA	82.63% (±6.83)	79.39% (±12.33)	76.29% (±12.60)	77.11% (±9.88)
	RF	88.56% (±8.40)	95.76% (±6.71)	74.81% (±16.92)	83.05% (±14.11)
	DT	80.35% (±5.54)	83.15% (±6.85)	65.15% (±9.88)	72.55% (±6.79)

Table 7 - A comparison of SVM, LDA, RF, and DT classification results in the delta, theta, alpha, beta, and gamma bands using mRMR, Relief-F and forward feature selection uisng feature from all frequency bands.

Feature Selection Method	Model	Accuracy	Specificity	Sensitivity	F1-Measure
mRMR	SVM	94.10% (±4.67)	95.57% (±6.43)	90.26% (±10.69)	92.25% (±6.50)
	LDA	88.19% (±4.88)	85.65% (±6.78)	85.74% (±9.86)	85.31% (±6.20)
	RF	91.84% (±5.09)	97.04% (±3.45)	82.50% (±12.98)	88.55% (±8.24)
	DT	84.07% (±5.22)	83.16% (±7.25)	76.53% (±12.36)	79.08% (±7.21)
Relief-F	SVM	94.83% (±4.73)	97.47% (±6.17)	90.48% (±11.16)	93.22% (±6.57)
	LDA	85.20% (±5.23)	83.03% (±7.89)	79.98% (±12.06)	80.73% (±7.80)
	RF	88.92% (±7.55)	96.84% (±5.49)	75.83% (±18.15)	83.39% (±13.42)
	DT	84.84% (±3.65)	85.59% (±10.19)	76.22% (±4.95)	80.14% (±4.20)
forward feature selection	SVM	98.52% (±1.29)	99.25% (±1.84)	97.30% (±3.14)	98.22% (±1.57)
	LDA	90.78% (±3.78)	91.94% (±8.07)	85.29% (±9.01)	87.92% (±5.04)
	RF	96.66% (±1.18)	100.00% (±0.00)	91.58% (±3.05)	95.58% (±1.69)
	DT	89.98% (±4.47)	95.02% (±4.97)	79.95% (±10.67)	86.33% (±5.80)

Table 8 - Comparing our work with other existing works on the prediction of antidepressant treatment outcome from EEG signals

Authors	year	feature	Classifier	Accuracy
Mumtaz et al.[23]	2017	Combination of Wavelets + STFT + EMD features	Logistic Regression	91.6%
Jaworska et al.[19]	2019	Demographic data, EEG power features, sourcelocalized current density	RF	88%
Zhdanov et al.[51]	2020	power spectral, spatiotemporal complexity	SVM	82.4%
Salle et al. [20]	2020	Change in Montgomery Åsberg Depression Rating Scale, Change in prefrontal cordance	Logistic Regression	85%
Khodayari-Rostamabad et al.[17]	2013	power spectral density, magnitude squared spectral coherence, mutual information, the log ratio of left-to-right hemisphere powers and anterior/posterior log power ratios	mixture of factor analysis classifier	87.9%
Patel et al.[25]	2015	Demographical, Cognitive ability, Functional Imaging, Structural Imaging	ADTree	89.47%
Van der Vinne et al.[52]	2019	EEG Abnormality (Diffuse Slowing, Focal Slowing, Paroxysmal and NonParoxysmal Activity)	ANOVA	Response Rate = 74%
Our work	2022	Effective connectivity	SVM	98.52%