Prediction of Response to Deep TMS Depression Therapy Using Machine Learning Methods on EEG Signals

B.Sc. Final Project

Yonathan Lerner
Biomedical Engineering Department
Ben Gurion University of the Negev
Beer Sheva, Israel
lerney@post.bgu.ac.il

***Abstract*— Background: Deep Transcranial Magnetic Stimulation (dTMS) has emerged as a highly effective treatment option for major depressive disorder (MDD), targeting the medial prefrontal cortex (H7 coil) or the lateral prefrontal cortex (H1 coil). However, the lack of knowledge regarding specific biomarkers for predicting treatment response poses a challenge. Given the intensive nature and high costs associated with dTMS treatment, the ability to differentiate responders and non-responders in advance is of paramount importance.**

Methods: This study conducted extended data analysis on cleaned EEG records to predict treatment response in patients undergoing dTMS with the H7 and H1 coils. The analysis encompassed three steps: data abstraction, feature extraction, and treatment response prediction using machine learning models.

Results: The research included 58 patients treated with the H1 coil and 75 patients treated with the H7 coil. Promising results were obtained for both coils on SVM model using frequency power abstraction method, The accuracy score for H7 and H1 coil in above median improvement of HDRS21 scores was 84% and 83%, respectively.

Conclusion: This study contributes to the prediction of treatment response in both the H7 and H1 coils of dTMS and identification of relevant biomarkers for treatment success. The findings demonstrate the potential to optimize dTMS treatment outcomes by differentiating responders and non-responders in advance. Future research should explore larger datasets to further enhance prediction accuracy and ultimately improve patient care.

***Index Terms*— major depressive disorder (MDD), Deep Transcranial Magnetic Stimulation (dTMS), electroencephalography (EEG), Machine Learning, KarmaLego algorithm.**

I. Introduction

MAJOR depressive disorder (MDD) is a prevalent mental health condition characterized by persistent feelings of sadness, hopelessness, and a diminished interest or pleasure in once-enjoyable activities. MDD is a leading cause of disability worldwide, affecting more than 260 million people and is a major contributor to the overall global burden of disease with genetic, biological, and environmental risk factors [1]. Unfortunately, initial treatments often fail to lead to remission and recovery for a significant number of patients. The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study [2] revealed that less than 50% of depressed patients respond to the first antidepressant they try, and merely 30% achieve full remission. Moreover, the likelihood of achieving remission decreases with each subsequent treatment failure [2].

Deep Transcranial Magnetic Stimulation (dTMS) is a non-invasive neurostimulation technique that has emerged as a promising therapeutic intervention for major depressive disorder (MDD) [3], [4]. It has demonstrated a remission rate of around 30% and response rate of around 60% in patients who have previously shown no response to several antidepressants [3], highlighting its potential as an alternative treatment option for medication-resistant depression. The dTMS treatment is typically administered using specialized coils, such as the H7 coil or the H1 coil, targeting the medial and lateral prefrontal cortex, respectively. By precisely modulating neural activity in these specific brain regions associated with mood regulation, dTMS aims to restore the dysfunctional neurocircuits implicated in depression. This targeted approach sets dTMS apart from traditional Transcranial Magnetic Stimulation (TMS), as it allows for deeper brain structure stimulation, leading to more effective outcomes [4], [5].

Despite the promise of dTMS as an innovative treatment for MDD, it remains an expensive and resource-intensive option. The need for specialized equipment and skilled professionals contributes to the high cost and limits its accessibility in many healthcare settings. Moreover, the treatment process is time-consuming, requiring multiple sessions over several weeks to achieve optimal results. Consequently, given the negative impact of treatment non-response on patients' ability to achieve remission and the resource-intensive nature of dTMS, there is a growing need to develop methods for predicting treatment response of patients.

Analyzing EEG signals for potential biomarkers holds promise as a means of predicting individual responses to dTMS depression therapy. Numerous studies [6]-[7] have explored the prediction of treatment response to MDD using EEG based biomarkers. By exploring biomarkers related to specific frequency waves and different time-related patterns, researchers have investigated the relationship between clinical outcomes and changes in EEG frequency bands, particularly in the alpha and theta bands. One biomarker found was alpha band hemisphere asymmetry, where greater alpha activity in left hemisphere relative to right hemisphere was correlated to responders whereas greater alpha activity in right hemisphere relative to left hemisphere was correlated to non-responders [6].

Another potential biomarker for treatment response prediction is brain microstates [8]-[10]. Brain microstates are rapid and brief periods of stable neural activity that occur in the brain during certain cognitive processes. These microstates are captured through electroencephalography (EEG) and represent distinct patterns of synchronized neural firing across different brain regions. These states last only a

few tens of milliseconds but are crucial for understanding the brain's complex information processing. Researchers believe that brain microstates may correspond to fundamental building blocks of cognition, representing the brain's functional architecture during specific tasks such as perception, attention, and memory. Studying these microstates could provide valuable insights into the temporal dynamics and organization of brain function, shedding light on the mechanisms underlying various cognitive processes.

Building upon these findings, this study aims to predict treatment response in MDD patients undergoing dTMS therapy by analyzing these EEG biomarkers. I will investigate various methods for detecting time-related patterns and changes in different frequency bands to identify potential predictors of treatment outcomes.

In conclusion, this research seeks to explore the potential of EEG signal analysis as a means of predicting treatment response to dTMS therapy in MDD patients. The findings from this study may offer valuable contributions to personalized treatment approaches and improve the overall effectiveness of dTMS interventions for those with MDD.

II. Dataset

The study will involve 75 subjects treated with the H7 coil, inducing magnetic stimulation in the medial prefrontal cortex, and 58 subjects treated with the H1 coil, producing magnetic stimulation in the lateral prefrontal cortex.

TABLE I. Dataset Infromation

Info	H1 Coil	H7 Coil
Count	58	75
Age	46.1 ± 12.4	44.4 ± 12.2
HDRS21 baseline	24 ± 3.6	23.9 ± 3.4
HDRS21 Improvement	12.7 ± 6.8	13.3 ± 6.6
response (%)	62.9	67.5
remission (%)	51.8	59.7
Impr. centered by site	0.1 ± 6.7	0 ± 5.83

III. Metrics

The ability of the models proposed in this study was evaluated using 6 conventional metrics: Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) and F1-score. In the following equations (1) to (6) TP represents true positives, TN represents true negatives, FP represents false positive, and FN represents false negative.

Accuracy (AC): The ratio of correct predictions to the total predictions.

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Sensitivity (SN): The ratio of true positives to the total positives.

$$SN = \frac{TP}{TP + FN} \quad (2)$$

Specificity (SP): The ratio of true negatives to the total negatives.

$$SP = \frac{TN}{TN + FP} \quad (3)$$

Positive Prediction Value (PPV): The ratio of true positive predictions to the total positive predictions.

$$PPV = \frac{TP}{TP + FP} \quad (4)$$

Negative Prediction Value (NPV): The ratio of true negative predictions to the total negative predictions.

$$NPV = \frac{TN}{TN + FN} \quad (5)$$

F1 score (F1): The harmonic mean of PPV and Sensitivity

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (5)$$

IV. Methods

A. Pre-processing EEG data

The initial EEG data was already cleaned from noises. The resting EEG signals of the patients were thoroughly examined to ensure that all 31 electrodes were in the same locations. Additionally, the signals were resampled from 2048 Hz to 1024 Hz, reducing the data size while retaining relevant frequency information.

B. Abstracting the data into labelled time intervals

The abstraction of EEG data was a crucial step in simplifying continuous data into discrete data with matching time stamps. This process was carried out using four abstraction methods.

1) Microstates:

Using the microstate toolbox in EEGLAB, each EEG record was transformed into a time series vector with labels representing seven distinct microstates. These microstates were generated through K-means clustering of voltage maps from 283 records, including subjects' baseline and post-treatment data, to represent the most common states of the voltage maps (see Fig. 1).

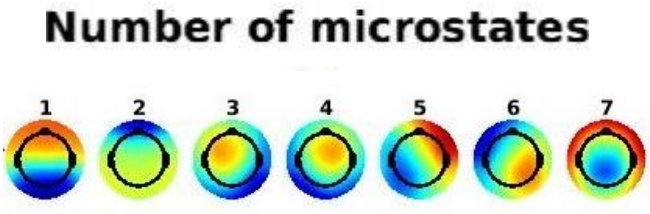

Fig. 1. Brain activity voltage maps clustered to 7 microstates

2) State:

The purpose of this abstraction method is to simplify the representation of voltage values in each electrode. Credit for the code implementation goes to Ori Kintzlinger, and the method is based on the principles outlined in article [11].

The state abstraction method involves four main steps:

a. Allocating bin values: To represent each bin based on the data distribution, an equal-width discretization method (EWD) is utilized, meaning the bin values are divided equally in width between the maximum and minimum values of the data. In this study, I used three bins.

b. Discretising the data: The data is discretized based on the distance to the bin values, finding the closest bin, and assigning the corresponding bin label

c. .Merging time intervals: In order to simplify the data small gaps where there are changes in labels are smoothens.

d. Saving time intervals Saving each time interval and corresponding label.

3) Trend

Like the State abstraction, I also used the code implementation made by Ori Kintzlinger for this method. The difference here was that this method of discretization tested the changes in the signal over time using derivatives. Specifically, the method involved calculating the angle of the gradient using arctan on a data window before the sample. To detect small changes effectively, I chose a window length of 10 samples (1024 Hz sampling rate) for this analysis. The discretization was done using an equal-width discretization method with three bins.

4) Frequency power

In this method, I developed a four-step process to analyze the EEG data in the frequency domain. This analysis was performed on all records at once, and the steps were as follows:

a. STFT (Short-Time Fourier Transform): The signals were divided into time intervals (I chose to use 200ms intervals) to apply STFT, which is a technique used to transform signals from the time domain to the frequency domain. This allows capturing the frequency information within each interval effectively.

b. Calculating frequency bands absolute powers: For the analysis, I focused on specific frequency bands such as alpha (8 to 12 Hz), beta (12 to 30 Hz), theta (3.5 to 7.5 Hz) and gamma (30 to 80 Hz) I calculated the mean absulute power for each interval and each given frequency band.

c. Calculating labels values: I analysed the distribution of mean powers for each frequency band across all intervals and records. To calculate label values to each combination of frequency band and channel, I used percentiles of the normal distribution based on the calculated mean of means (second mean refers to absolute power of specific frequency band and channel, first mean refers to the mean value of second means across all intervals and records) and standard deviation of means (same here).

d. Assigning Labels: For each combination of time interval and frequency band, I assigned a label based on the closest label value for that specific frequency band. I chose to investigate the data using four bins and focused on the theta, alpha, beta, and gamma frequency bands.

It's essential to maintain a balance between the length of time intervals and frequency resolution as shorter time intervals impact frequency resolution.

C. Feature extraction

Two feature extraction methods were employed: the simple way and the KarmaLego algorithm.

1) Simple way: This method is basically label counting for each subject, which I used only on the frequency power abstraction method due to time constraints. While this method is relatively simple, it does not detect any temporal patterns or time-related patterns.

*2) KarmaLego algorithm:*The KarmaLego algorithm [12], a data mining technique, was used to analyze sequential patterns within event sequences. It employs a divide-and-conquer strategy, recursively breaking down the original sequence into smaller segments. The algorithm constructs a pattern tree, with each node representing a frequent sequential pattern found in the data. To identify relevant patterns, a user-defined support threshold is used, ensuring only significant patterns (above the threshold) are included in the tree. By pruning infrequent patterns, KarmaLego reduces the complexity of the pattern tree, focusing on the most meaningful and frequent patterns. It excels at handling sequences of different lengths and can capture long-term, complex dependencies between events.

I applied the KarmaLego algorithm to the State, Trend, and Microstates abstractions. The final features used in models were the number of appearances and the meantime length of patterns per subject.

D. Modelling

In this stage, I developed different classification models to predict the response to treatment. The classification tasks were based on the above median improvement in HDRS21 scores as it provided the most effective label balancing approach. The improvement was centered by site to remove the subjective noise made by different doctors evaluating the patients in different sites.

The models were separated for different treatments (H1 models and H7 models) and were based on baseline EEG records only. In addition, a method of removing part of the central percentile of improvement in HDRS21 score from the training data was obtained. Furthermore, another method of thickening the data by dividing the records into parts with or without overlapping was applied, converting 5-minute records into 19 records of 1 minute with 30 seconds of overlap. To prevent data leakage and maintain the integrity of the subject records groups, I created a train-test split function that separates the data while maintaining label ratio and subject records groups.

The models tried in this study included SVM classifier, random forest, gradient descent trees (such as XGBoost and Catboost), K-means and KNN. Evaluations were performed using the leave-one-out method and the leave-one-group-out method for divided records.

In this study, a novel feature important calculation method is proposed, for Support Vector Machine (SVM) model with a polynomial kernel. The primary objective of this method is to evaluate the significance of each feature in the SVM model's decision-making process. To achieve this, we extract the support vectors and dual coefficients of the SVM model and utilize them to compute the importance score for each feature. For each feature, we perform a weighted sum across the support vectors, with the weight being the absolute value of the corresponding dual coefficient. This sum is then raised to the power of the polynomial kernel's degree, representing the importance of the feature. The feature important scores are subsequently normalized to obtain a comparable ranking.

The decision to use this custom feature importance calculation method for the high-performing SVM model with a polynomial kernel was primarily driven by the model's significant improvement in accuracy (at least 10 percent higher) compared to other tested models on the dataset. Although this method is a custom and new method the relevant of this method results should be considered, as later be elaborated in the results section.

ALGORITHM 1: EXTRACTION OF FEATURE IMPORTANCE FROM POLY KERNEL SVM MODEL

1 ***Initialize*** feature_importance dictionary
2 ***Get*** support vectors and dual coefficients from the model
3 ***Calculate*** feature importance for the model
  A ***For*** each support vector (sv_idx) and dual coefficient (dual_coef)
    I ***For*** each feature (feature_idx) in X_train columns
    - ***Calculate*** the importance for the current feature importance = abs(dual_coef) * (np.dot (sv [sv_idx, feature_idx], X_train.iloc[sv_idx, feature_idx]) + 1) ** model. Degree
    - ***Accumulate*** the importance for the feature: feature_importance[feature_name] += importance
4 ***Sort*** features based on importance in descending order
5 ***Return*** the sorted_importance list

V. RESULTS

A. Model results:

In this study, two different treatments and four abstraction methods, including divided records and full records, were examined. The KarmaLego algorithm was employed to extract features from microstate, state, and trend data. For the frequency power abstraction method, due to time constraints, only the simple way (mentioned in methods) is used for feature extraction. To optimize the models' performance, the training data was filtered by dropping the 32nd to 68th percentiles of HDRS21 improvement centered by site distribution. The machine learning models were then evaluated, and it was observed that the frequency power abstraction combined with full records and SVM using a polynomial kernel yielded the most promising results. The classification focused on HDRS21 score improvement, where instances with improvement above the median were labeled as 1, and those with improvement below the median were labeled as 0. Hyperparameter adjustments for the SVM classifiers, particularly C (ranging from 0.025 to 0.12) and Degree (either 2 or 3) for the polynomial kernel, were essential in obtaining optimal outcomes.

Model results are presented in the tables below:

TABLE II. H7 COIL BEST MODELS

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
SVC 1	0.84	0.78	0.9	0.88	0.81	0.82
SVC 2	0.84	0.72	0.95	0.93	0.79	0.81
SVC 3	0.81	0.83	0.79	0.79	0.84	0.81
SVC 4	0.83	0.69	0.95	0.93	0.77	0.79
SVC 5	0.79	0.94	0.64	0.71	0.93	0.81

TABLE III. H1 COIL BEST MODELS

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
SVC 6	0.83	0.64	1	1	0.75	0.78
SVC 7	0.81	0.89	0.73	0.76	0.88	0.82

TABLE IV. BEST MODELS HYPERPARAMETERS

Model	Hyperparameters
SVC 1	C=0.1, kernel=poly, degree=3
SVC 2	C=0.12, kernel=poly, degree=3
SVC 3	C=0.06, kernel=poly
SVC 4	C=0.11, kernel=poly, degree=3
SVC 5	C=0.025, kernel=poly, degree=3
SVC 6	C=0.1, kernel = poly, degree = 2
SVC 7	C=0.06, kernel=poly, degree=3

Only two models are presented for H1 coil because numerous changes in hyperparameters within H1 models yielded identical results. As a result, two models were arbitrarily chosen from among them.

To evaluate the performance of the proposed models thoroughly, I present Receiver Operating Characteristic (ROC) plots for the top-performing models in this section.

The ROC plot graphically illustrates the performance of a binary classifier by displaying the true positive rate (sensitivity) against the false positive rate (1 – specificity). The knee value on the ROC plot represents the optimal threshold that achieves a balanced trade-off between sensitivity and specificity, which is determined based on the probability outcome of the model's predictions. In other words, this threshold represents the probability level at which the model classifies instances into positive or negative outcomes, ensuring a desirable balance between correctly identifying positive cases (sensitivity) and accurately identifying negative cases (specificity).

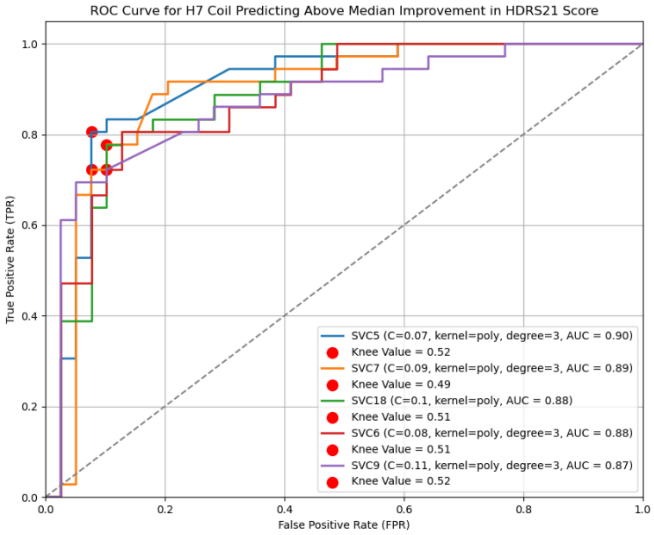

Fig. 2. ROC plot for top-performing models on H7 coil above median improvement in HDRS21 scores (centered by site)

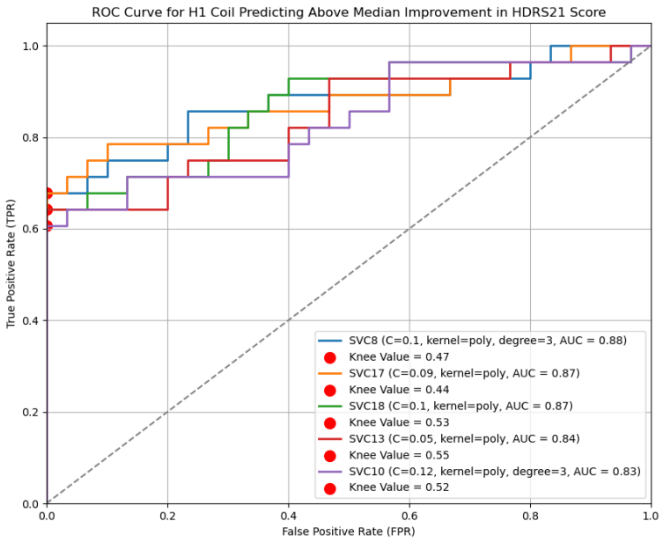

Fig. 3. ROC plot for top-performing models on H1 coil above median improvement in HDRS21 scores (centered by site)

As shown above, in the models of both H1 coil and H7 coil the AUC scores are all above 83%. Highest score of H1 models is AUC of 88% and highest score of H7 models is AUC of 90%.

B. Feature Importance:

Based on the feature importance calculation method for the poly kernel SVM model described in the methods section, I present the following tables and figures. These figure and tables show the feature importance based on electrodes, frequency bands, and power levels (as explained earlier in the frequency power abstraction method). Each of these importance values has been calculated by summing the absolute importance of features within their rrespective category (electrode, frequency band, or power level) and then scaling the values from 0 to 100.

For the feature importance calculations, the model's hyperparameters were set as follows: C = 0.1, kernel = poly, and degree = 2 for H1 and degree = 3 for H7. These models were chosen due to their high accuracy scores.

TABLE V. FEATURE IMPORTANCE - FREQUENCY BAND

Freq. Band	Scaled Importance 0-100	
	H1	H7
Theta	92	1
Alpha	8	0
Beta	0	1
Gamma	100	100

TABLE VI. FEATURE IMPORTANCE - POWER LEVELS

Levels	Scaled Importance 0-100	
	H1	H7
L3	8	100
L2	1	0
L1	100	73
L0	0	7

As shown above in tables V and VI in the model on H1 HDRS21 total score improvement L1 was most significant along with theta and gamma frequency bands. In the model on H7 HDRS21 total score improvement both L3 and L1 was significant and from the frequency bands only gamma frequency band has shown importance.

The following two plots display the combined feature importance from the model, grouped by EEG channels (electrodes). The graphical representation is based on the locations of the electrodes in the X, Y, and Z coordinates. The left-to-right orientation corresponds to the left and right sides of the brain, while the vertical direction indicates the front and back perspectives, all from a view above

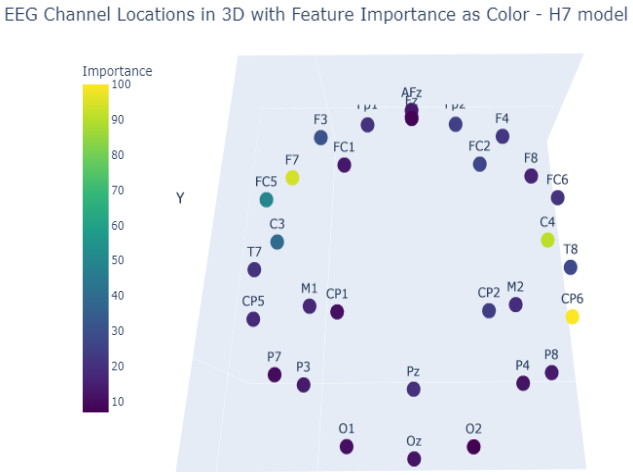

Fig. 4. The electrode locations and their importance in the H7 coil model.

As shown in the figure above, the most important channels to the the model on H7 coil were CP6, F7, FC5 and C4.

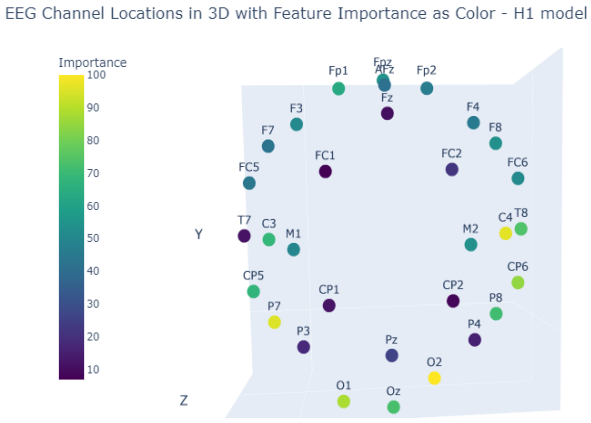

Fig. 5. The electrode locations and their importance in the H1 coil model.

In contrast to model on H7 coil, In the model on H1 coil there are more electrodes with importance to the model. In Both models C4 channel scaled importance is over 80.

After performing the feature importance calculation, an assessment of the method's reliability was conducted by evaluating the models with a subset of features. Specifically, features with scaled importance scores of less than 5 were removed from the original set, which initially comprised 430 features.

For the H7 model, the reduced feature set contained only 20% of the original features. In addition, the C hyperparameter was adjusted to a smaller value, approximately 10 in ratio. The model yielded an accuracy of 80% and an F1 score of 83%, very close metrics to original model that yielded an accuracy of 84% and an F1 score of 82%

For the H1 model, using the same approach, 45% of the original features were retained, and no adjustments to the C hyperparameter were made (given that the model's results in H1 were achieved across a wide range of C values). Remarkably, The H1 model achieved identical performance results to the original model, with an accuracy of 83% and an F1 score of 78%.

TABLE VII. REDUCED FEATURES' MODEL RESULTS

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
H1	0.83	0.64	1	1	0.75	0.78
H7	0.80	1.00	0.62	0.71	1.00	0.83

VI. DISCUSSION

In this study, several methods were proposed to predict treatment response for clinical depression (MDD) patients undergoing deep Transcranial Magnetic Stimulation (dTMS) therapy. The models' predictions were based on the improvement in HDRS21 scores, as the label distribution was evenly balanced. To enhance the model's performance, the middle percentile was removed during training. An important note is that with minor adjustments, the model could predict more known metrics such as response (a decrease of 50% or more in HDRS21 score) or remission (final score below 9).

Regarding the top-performing model, the best algorithm utilized the frequency power abstraction method, with a simple counting of labels as feature extraction and an SVM with a polynomial kernel as the classification model. Feature importance analysis revealed that removing 55% of total features in the H1 coil model and 80% in the H7 coil model resulted in minimal changes in model performance. This suggests the possibility of significant feature reduction, which could be combined with temporal patterns using the KarmaLego algorithm.

On the other hand, some models exhibited lesser performance. However, this does not necessarily indicate that the proposed algorithm is unsuitable. There might have been errors in reading EEG data correctly for state and trend algorithms due to different data formats. Additionally, exploring a larger number of microstates might lead to improved results, as the choice of 7 was made based on convention.

The analysis of feature importance revealed that not all electrodes were equally important to the models. Different electrodes played crucial roles in treatment response prediction for H1 and H7 coil treatments. Moreover, different levels of frequency power and various frequency bands contributed to the models' performance. However, the exact relationship between each feature and treatment outcome remains unclear and warrants further investigation.

VII. CONCLUSIONS

Upon reflecting on the study's findings, it becomes evident that the proposed method can contribute in predicting treatment response for MDD patients undergoing dTMS therapy. The ability to leverage EEG-based predictions could contribute significantly to time and cost savings by avoiding ineffective treatments and optimizing patient care.

An intriguing aspect of the study is the possibility of using the predictions for both H1 coil and H7 coil treatments to guide the selection of the most suitable treatment method in advance. This has the potential to positively impact treatment success rates by aligning patients with the treatment that is most likely to work for them, an idea that genuinely excites me.

However, I am aware that there is room for further improvement. To enhance the stability and robustness of the predictive model, it is clear that a much larger dataset is needed. This would allow for more comprehensive training and validation, potentially leading to even more accurate predictions. Furthermore, continued research and data collection efforts in the field of mental health and EEG-based predictive modelling are crucial to advancing our understanding and improving patient outcomes.

Moreover, the methods proposed in this research extend beyond their application to dTMS therapy for MDD. The insights gained from this study can potentially be adapted for various EEG-based predictions, offering opportunities to make early treatment response predictions in other therapeutic approaches. The prospect of applying these techniques in broader clinical contexts is both fascinating and challenging.

In conclusion, this study highlights the potential of machine learning and EEG analysis in predicting treatment response for clinical depression patients undergoing dTMS therapy. I am grateful for the opportunity to be part of this research, and I hope that the findings contribute in some way to the advancement of personalized medicine and the well-being of patients.

ACKNOWLEDGMENT

I would like to thank Professors Yuval Shahar, Professor Abraham Zangen and Dr. Uri Alyagon for their patient guidance, and to Ori Kintzlinger for his contribution to the project. Thank you all.

REFERENCES

[1] S. L. James *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 392, no. 10159, pp. 1789–1858, Nov. 2018, doi: 10.1016/S0140-6736(18)32279-7.

[2] M. Sinyor, A. Schaffer, and A. Levitt, "The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Trial: A Review," *Can. J. Psychiatry*, vol. 55, no. 3, pp. 126–135, Mar. 2010, doi: 10.1177/070674371005500303.

[3] K. K. Kedzior, H. M. Gellersen, A. K. Brachetti, and M. T. Berlim, "Deep transcranial magnetic stimulation (DTMS) in the treatment of major depression: An exploratory systematic review and meta-analysis," *J. Affect. Disord.*, vol. 187, pp. 73–83, Nov. 2015, doi: 10.1016/j.jad.2015.08.033.

[4] Y. Levkovitz *et al.*, "Efficacy and safety of deep transcranial magnetic stimulation for major depression: a prospective multicenter randomized controlled trial," *World Psychiatry*, vol. 14, no. 1, pp. 64–73, 2015, doi: 10.1002/wps.20199.

[5] Y. Levkovitz *et al.*, "Deep transcranial magnetic stimulation over the prefrontal cortex: Evaluation of antidepressant and cognitive effects in depressive patients," *Brain Stimulat.*, vol. 2, no. 4, pp. 188–200, Oct. 2009, doi: 10.1016/j.brs.2009.08.002.

[6] A. Baskaran, R. Milev, and R. S. McIntyre, "The neurobiology of the EEG biomarker as a predictor of treatment response in depression," *Neuropharmacology*, vol. 63, no. 4, pp. 507–513, Sep. 2012, doi: 10.1016/j.neuropharm.2012.04.021.

[7] F. S. de Aguiar Neto and J. L. G. Rosa, "Depression biomarkers using non-invasive EEG: A review," *Neurosci. Biobehav. Rev.*, vol. 105, pp. 83–93, Oct. 2019, doi: 10.1016/j.neubiorev.2019.07.021.

[8] L. Tait *et al.*, "EEG microstate complexity for aiding early diagnosis of Alzheimer's disease," *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41598-020-74790-7.

[9] V. Férat *et al.*, "Electroencephalographic Microstates as Novel Functional Biomarkers for Adult Attention-Deficit/Hyperactivity Disorder," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 7, no. 8, pp. 814–823, Aug. 2022, doi: 10.1016/j.bpsc.2021.11.006.

[10] Y. Luo, Q. Tian, C. Wang, K. Zhang, C. Wang, and J. Zhang, "Biomarkers for Prediction of Schizophrenia: Insights From Resting-State EEG Microstates," *IEEE Access*, vol. 8, pp. 213078–213093, 2020, doi: 10.1109/ACCESS.2020.3037658.

[11] E. Sheetrit, N. Nissim, D. Klimov, and Y. Shahar, "Temporal Probabilistic Profiles for Sepsis Prediction in the ICU," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, in KDD '19. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2961–2969. doi: 10.1145/3292500.3330747.

[12] R. Moskovitch and Y. Shahar, "KarmaLego - Fast Time Intervals Mining".

Supervisor: _____ Prof. Yuval Shahar
signature

Consultant: _____ Prof. Abraham Zangen
signature

Author: _____ Yonathan Lerner
signature