



OPEN ACCESS

EDITED BY

Chella Kamarajan,
Downstate Health Sciences University,
United States

REVIEWED BY

Stavros I. Dimitriadis,
Department of Clinical Psychology and
Psychobiology, Faculty of Psychology,
University of Barcelona, Spain
Andrew F. Leuchter,
University of California, Los Angeles,
United States

*CORRESPONDENCE

Elias Ebrahimzadeh
✉ e_ebrahimzadeh@ut.ac.ir;
elias.ebrahimzadeh@ipm.ir

RECEIVED 14 April 2022

ACCEPTED 13 February 2023

PUBLISHED 09 March 2023

CITATION

Ebrahimzadeh E, Fayaz F, Rajabion L, Seraji M,
Aflaki F, Hammoud A, Taghizadeh Z,
Asgarinejad M and Soltanian-Zadeh H
(2023) Machine learning approaches and
non-linear processing of extracted
components in frontal region to predict rTMS
treatment response in major depressive
disorder.

Front. Syst. Neurosci. 17:919977.

doi: 10.3389/fnsys.2023.919977

COPYRIGHT

© 2023 Ebrahimzadeh, Fayaz, Rajabion, Seraji,
Aflaki, Hammoud, Taghizadeh, Asgarinejad and
Soltanian-Zadeh. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Machine learning approaches and non-linear processing of extracted components in frontal region to predict rTMS treatment response in major depressive disorder

Elias Ebrahimzadeh^{1,2*}, Farahnaz Fayaz³, Lila Rajabion⁴,
Masoud Seraji⁵, Fatemeh Aflaki⁶, Ahmad Hammoud⁷,
Zahra Taghizadeh⁸, Mostafa Asgarinejad⁹ and
Hamid Soltanian-Zadeh^{1,2}

¹School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran, ²School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran, ³Biomedical Engineering Department, School of Electrical Engineering, Payame Noor University of North Tehran, Tehran, Iran, ⁴School of Graduate Studies, SUNY Empire State College, Manhattan, NY, United States, ⁵Department of Psychology, University of Texas at Austin, Austin, TX, United States, ⁶Department of Biomedical Engineering, Islamic Azad University Central Tehran Branch, Tehran, Iran, ⁷Department of Medical and Technical Information Technology, Bauman Moscow State Technical University, Moscow, Russia, ⁸Department of Bioengineering, George Mason University, Fairfax, VA, United States, ⁹Department of Cognitive Neuroscience, Institute for Cognitive Sciences Studies, Tehran, Iran

Predicting the therapeutic result of repetitive transcranial magnetic stimulation (rTMS) treatment could save time and costs as ineffective treatment can be avoided. To this end, we presented a machine-learning-based strategy for classifying patients with major depression disorder (MDD) into responders (R) and nonresponders (NR) to rTMS treatment. Resting state EEG data were recorded using 32 electrodes from 88 MDD patients before treatment. Then, patients underwent 7 weeks of rTMS, and 46 of them responded to treatment. By applying Independent Component Analysis (ICA) on EEG, we identified the relevant brain sources as possible indicators of neural activity in the dorsolateral prefrontal cortex (DLPFC). This was served through estimating the generators of activity in the sensor domain. Subsequently, we added physiological information and placed certain terms and conditions to offer a far more realistic estimation than the classic EEG. Ultimately, those components mapped in accordance with the region of the DLPFC in the sensor domain were chosen. Features extracted from the relevant ICs time series included permutation entropy (PE), fractal dimension (FD), Lempel-Ziv Complexity (LZC), power spectral density, correlation dimension (CD), features based on bispectrum, frontal and prefrontal cordance, and a combination of them. The most relevant features were selected by a Genetic Algorithm (GA). For classifying two groups of R and NR, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) were applied to predict rTMS treatment response. To evaluate the performance of classifiers, a 10-fold cross-validation method was employed. A statistical test was used to assess the capability of features in differentiating R and NR for further research. EEG characteristics that can predict rTMS treatment response were discovered. The strongest discriminative indicators were EEG beta power,

the sum of bispectrum diagonal elements in delta and beta bands, and CD. The Combined feature vector classified R and NR with a high performance of 94.31% accuracy, 92.85% specificity, 95.65% sensitivity, and 92.85% precision using SVM. This result indicates that our proposed method with power and nonlinear and bispectral features from relevant ICs time-series can predict the treatment outcome of rTMS for MDD patients only by one session pretreatment EEG recording. The obtained results show that the proposed method outperforms previous methods.

KEYWORDS

electroencephalography (EEG), repetitive transcranial magnetic stimulation (rTMS), major depressive disorder (MDD), prediction treatment response, independent component analysis (ICA), classification, machine learning approaches, non-linear processing

1. Introduction

Major Depressive Disorder (MDD), more commonly known as clinical depression, is a severe condition with potential morbidity and mortality, affecting and threatening millions of people worldwide. MDD is typically treated with one type of antidepressant, however, 50% to 70% of patients are shown to be categorically unresponsive to medication-based treatments (Chekroud et al., 2017; Ebrahimzadeh et al., 2021a; Leichsenring et al., 2022; Turner et al., 2022). Therefore, there has been an ongoing search for other therapeutic approaches to target patients with resistant depression. One method that has gained increasing attention as a safe alternative or complementary technique to treat MDD is repetitive Transcranial Magnetic Stimulation (rTMS; Ćukić, 2020). rTMS involves a series of short magnetic pulses directed to the brain to stimulate nerve cells. The magnetic pulses stimulate area neurons and change the functioning of the brain circuits involved. This method is a noninvasive treatment that directs magnetic pulses at the left or right dorsolateral prefrontal cortex (DLPFC) at regular intervals to stimulate neurons and trigger action potentials. It may be used as an adjunctive therapy to increase or hasten the efficacy of conventional pharmacotherapy through changing and modulating cortical activity. The overall effectiveness and limited side effects of rTMS have been established in several studies. Nonetheless, clinicians prescribe rTMS after conducting a thorough assessment and a series of trial-and-error tests to improve diagnostic accuracy and treatment outcomes and, importantly, to prevent patient relapse which may be the case if the patient is unresponsive to rTMS. This calls for developing indicators that can help with predicting rTMS response in order for patients to benefit from the merits of this treatment and avoid costly, ineffective procedures. Neurophysiological modalities, including fMRI and EEG, have been used to this end, between which, EEG, being more widely available and cost-effective, makes a more robust biomarker (Bachmann et al., 2013; Patel et al., 2015; Redlich et al., 2016; Wade et al., 2016; Ćukić et al., 2020). This is why an increasing number of studies have applied EEG-based machine learning techniques and statistical methods to distinguish rTMS responders from non-responders (Bares et al., 2007; O'Reardon

et al., 2007; Khodayari-Rostamabad et al., 2011; Arns et al., 2012, 2014; Kito et al., 2012).

The literature also includes similar efforts to determine the responsiveness of other approaches to resistant MDD treatment. A number of studies, such as Bares et al. (2007) have dealt with regarding changes in QEEG prefrontal cordance as a predictor of response to antidepressants. The effectiveness of selective serotonin reuptake inhibitors (SSRI), as another potential predictor of treatment response, has been investigated in the study of Khodayari-Rostamabad et al. (2013), where Khodayari-Rostamabad et al. obtained the patient's initial EEG and used it to extract features and perform a mixture of factor analysis (MFA) model. This led to a classification accuracy of 87.9%. SSRI efficacy was also examined in another study, where logistic regression (LR) was applied to wavelet features of baseline EEG, producing an accuracy of 87.5% (Mumtaz et al., 2017). Transcranial Direct-Current Stimulation (tDCS) is another treatment for mood and cognition improvement in patients with MDD, which was the focus of Al-Kaysi et al. (2017). The authors made use of Linear Support Vector Machine (LSVM), Linear Discriminant Analysis (LDA), and neural networks to classify features extracted from EEG and achieved an accuracy of 76% and 92% in mood and cognition labeling, respectively.

As for rTMS, the authors (Cao et al., 2018a) provided a meta-analysis reporting relatively low rates of 40.9% and 16.4% for response and remission, respectively. Bailey et al. (2018) examined EEG recordings when participants were completing a Working Memory (WM) task and predicted rTMS response with an accuracy of 91%. In another effort, they obtained resting state EEG at baseline and 1 week after the start of the treatment and combined mood and EEG features using an LSVM classifier. This yielded a classification accuracy of 86.6% (Bailey et al., 2019). The extracted EEG features they used included EEG power and weighted phase lag index (wPLI) in alpha and theta (Bailey et al., 2019) and gamma (Bailey et al., 2018) frequency bands, alpha peak frequency (iAPF), and frontal theta cordance (Bailey et al., 2019; Jaworska et al., 2019). The literature also includes other EEG features used for the same purpose such as power spectral features (Khodayari-Rostamabad et al., 2013; Al-Kaysi et al., 2017), coherence (Khodayari-Rostamabad et al., 2013;

Mumtaz et al., 2017), mutual information (MI; Khodayari-Rostamabad et al., 2013), nonlinear features (Hasanzadeh et al., 2019), time-frequency processing (Ebrahimzadeh et al., 2021a), and wavelet coefficients (Mumtaz et al., 2017). Among these, EEG power in different frequency bands and their combinations have received a lot of attention in predicting MDD treatment (Suffin and Emory, 1995; Knott et al., 2000; Cook et al., 2002; Bruder et al., 2008; Spronk et al., 2011; Tenke et al., 2011; Arns et al., 2012; Pellicciari et al., 2013; Wade et al., 2016; Lebiecka et al., 2018). For instance, treatment response was shown to be associated with cordance measures (Leuchter et al., 1994) and, in another study, with an Antidepressant Treatment Response (ATR) index (Iosifescu et al., 2009). Olejarczyk et al. (2020) evaluated the impact of rTMS on functional connectivity in MDD and bipolar disorder by directed transfer function and indices based on graph theory.

A number of studies have brought their attention to the non-Gaussian, higher-order nature of EEG to discover supplementary information that is not detected in the power spectrum. Adopting this view, Hasanzadeh et al. (2019) used non-linear and bispectral features to classify rTMS response.

Furthermore, a large strand of literature employs three prefrontal electrodes, i.e., FP1, FP2, and FPz, to extract predictive features, as the frontal lobe is believed to contain significant changes in MDD. Such studies tend to limit their analyses to the outcome of those electrodes, which could be interpreted as a simplification of the matter: there is little guarantee that the frontal components would not affect the channels from other areas namely the central, parietal, temporal, and occipital. It can then be stated that the components from the frontal lobe are more involved than those from other areas. Identifying these components and extracting features from their time series could lead to more realistic results compared to those of the other EEG channels. In other words, the frontal components form a neural network, which is involved in the rTMS treatment, and the EEG channels reflect their activity. That said, this article tries to shift the focus to the component domain. To do this, we first had the channels decomposed to their components and identified the appropriate components through analyzing their locations and the dipoles related to the MNI model. We then used the time series of the selected components to extract features and performed the classification. To choose predictive features for rTMS treatment response, we investigated a relatively comprehensive set of component's time-series features including spectral, bispectral, and nonlinear features, namely bispectrum features, Lempel-Ziv Complexity (LZC), correlation dimension (CD), fractal dimension (FD), component power in all frequency bands (delta, theta, alpha, and beta), and frontal and prefrontal cordance in the theta band, all extracted from pretreatment resting EEG. In addition to classification, we performed a statistical test to evaluate the differences of features in two groups of responders (R) and non-responders (NR). Ultimately, we employed a Genetic Algorithm (GA) for feature selection. To the extent of our knowledge, this is the first time that the capability of selected bispectral, nonlinear, and spectral features on the components time-series have been investigated simultaneously with the aim of treatment response prediction.

As illustrated in **Figure 1**, we first provide the information of participants, and the procedure of EEG acquisition. After extracting the relative components and features, we perform the classification

and statistical test. Based on our classification study, we will evaluate the prediction ability of feature sets in the Results Section. The features will then be used to categorize the subjects into two groups of R and NR based on our statistical analysis. The results are elaborated in the Discussion Section where the limitations of the current study and suggestions for future work are also presented. The article is concluded in the Conclusion section.

2. Methods and material

2.1. Participants

We recruited 88 patients with MDD in the age range of 18–54 years. The patients were referred to the Neurology Clinical Neuroscience Centre, Tehran, Iran. A psychiatrist diagnosed them with MDD using the Diagnostic and Statistical Manual-IV (DSM-IV) diagnostic criteria (Segal, 2010). The Hamilton Rating Scale for Depression (HRSD) and the Beck Depression Inventory were also used to assess the participants (BDI-II). All participants have an HRSD score ≥ 12 and a BDI-II score ≥ 15 . The demographic and clinical data of the participants are described in **Table 1**. Wilcoxon rank-sum and Friedman tests were used to compare the R and NR groups, with the results displayed in column 4 of **Table 1**.

In terms of age, gender, rTMS treatment features, pretreatment BDI-II and HRSD score, medications, duration of illness, length of current depression episode, and the number of previous drugs, there were no significant differences ($p > 0.05$) between R and NR. Only the posttreatment BDI-II and 17-item HRSD scores differed significantly ($p < 0.05$), with responders scoring significantly lower. The outcome of rTMS cannot be attributed to differences in depression severity of the two groups because the pretreatment BDI-II and HRSD of R and NR are not significantly different. The existence of Axis I or II disorders, substance misuse, suicide risk, unstable medical conditions, implanting devices, cardiac arrhythmia, and pregnancy are all exclusion criteria in this study. Participants having a current or previous head injury, seizures, epilepsy, or neurological diseases were also excluded from the study. In this study, 59 individuals had been on antidepressants, mood stabilizers, and antipsychotics for more than 4 weeks prior to the treatment. The Bioethics Committee of the Iran University of Medical Science authorized all experimental techniques, which followed the principles of the Declaration of Helsinki standards. All participants volunteered and provided written consent after being told about the study procedures and goals. **Table 1** summarizes the demographic and clinical information of the participants.

2.2. Procedure and clinical assessment

A baseline interview was conducted with MDD patients to collect demographic and depression severity data. The 17-item Hamilton Rating Scale for Depression (HRSD; Sharp, 2015), the Montgomery-Asberg Depression Rating Scale (MADRS; Leentjens et al., 2000), and the Beck Depression Inventory-II (BDI-II; Dozois et al., 1998) were used to determine the severity of depression.

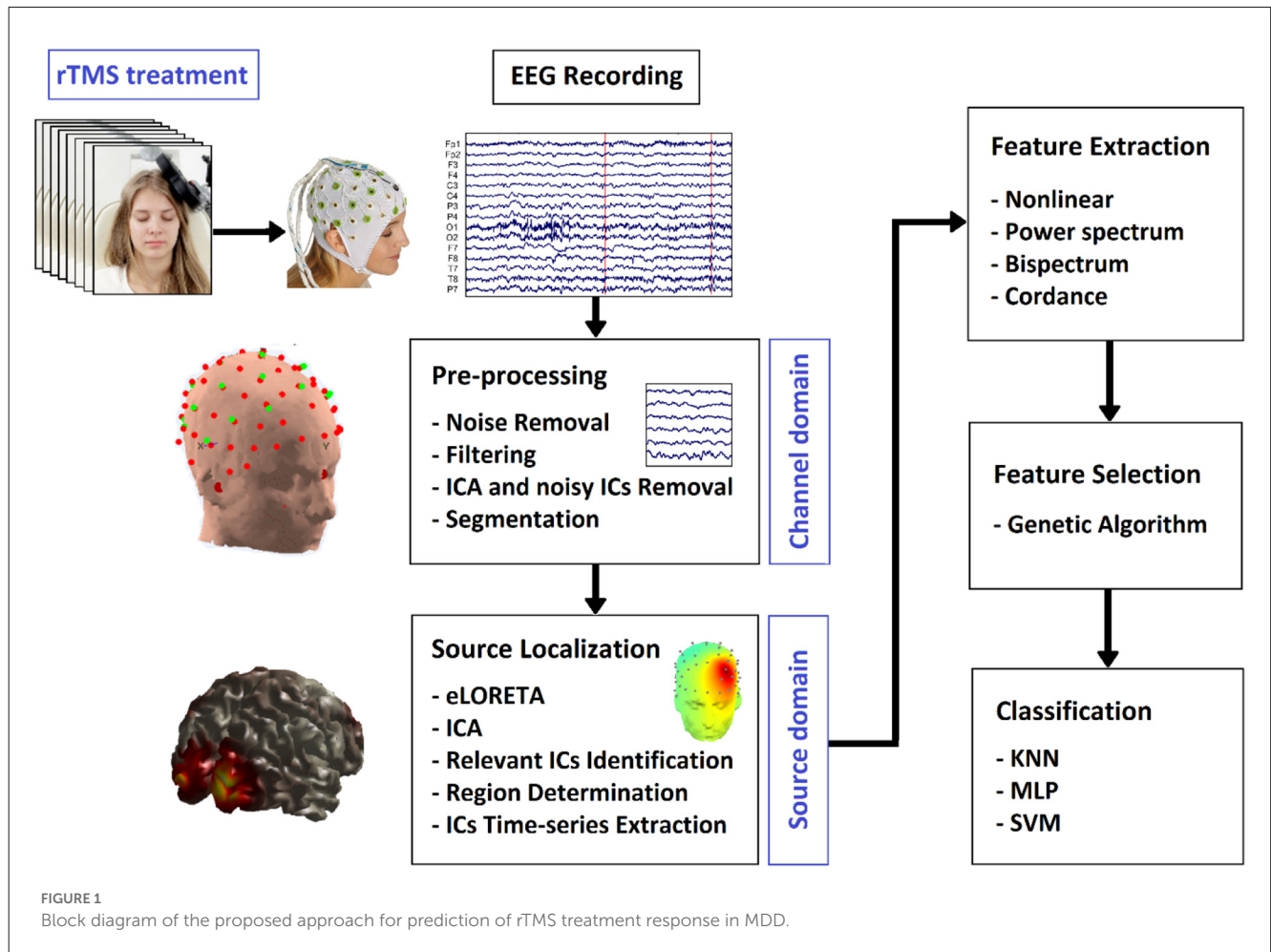


TABLE 1 Clinical and demographic information.

	Responder (n = 46)	Non-Responder (n = 42)	Statistics
Sex (F/M)	27/19	25/17	P = 0.940
Age (SD)	41.2 (15.5)	40.3 (13.8)	P = 0.108
Years of education (SD)	14.2 (2.3)	14.1 (2.6)	P = 0.375
Handedness (R/L/A)	22/18/6	26/15/1	P = 0.101
HDRS	24.5 (4.3)	23.3 (4.4)	P = 0.921
Treatment (HF/LF/bilateral)	14/17/15	11/13/18	P = 0.481
Number of previous medications	2.8 (1.5)	2.2 (1.8)	P = 0.214
Medications (AD/ AD ± MS / AD ± MS ± AP)	16/21/9	19/21/2	P = 0.592
Anxiety (Y/N)	34/12	27/15	P = 0.131
Disease duration (years)	5.3 (7.6)	6.4 (8.3)	P = 0.268
Pre-treatment BDI-II	31.2 (10.1)	29 (9.6)	P = 0.098
Post-treatment BDI-II	9 (4.2)	23 (7.4)	P < 0.001
Pre-treatment HRSD	36 (7.3)	28 (3.1)	P = 0.193
Post-treatment HRSD	7 (5.4)	26 (6.5)	P < 0.001

M, Male; F, Female; SD, Standard Deviation; L, Left; R, Right; A, Ambidextrous; HDRS, 17-item Hamilton Depression Rating Scale. HF, High Frequency (10 Hz); LF, Low Frequency (1 Hz); AD, Antidepressant; MS, Mood Stabilizer; AP, Antipsychotic.

MDD patients received daily (5 days per week) unilateral left 10 Hz rTMS therapy for 3 weeks. Individuals who responded after 3 weeks were given an extra 2 weeks of titrated rTMS treatment (three sessions in week 4, two sessions in week 5), followed by

another EEG session. Nonresponders were randomly assigned to continue with unilateral left 10 Hz rTMS therapy, unilateral right 1 Hz rTMS treatment, or sequential bilateral rTMS treatment consisting of right 1 Hz rTMS followed by left 10 Hz rTMS

treatment for the next 3 weeks. Individuals who did not respond by week 3 but did by week 6 were given another 2 weeks of titrated rTMS treatment (three sessions in week 7, two sessions in week 8). The procedure for rTMS treatment is shown in [Figure 2](#). The stimulation intensity was set at 110% of the resting motor threshold. The left-sided treatment consisted of 40 5-s trains separated by a 25-s break. One train of 1,200 pulses was used for right-sided rTMS. These procedures were combined in bilateral rTMS, but with only 900 right-sided pulses. The coil was held tangentially to the head and its handle facing back and away from the midline at 45°, and the rTMS was applied to the left/right DLPFC and bilateral DLPFC at a point 5 cm anterior in a parasagittal line to the motor threshold location (the left abductor pollicis brevis muscle). A TAMAS (REMEDI, Daejeon, South Korea) with a figure-of-eight-shaped coil (field strength 3 Tesla) delivered the rTMS. At the end of week 1 and week 3, the MADRS and BDI-II tests were redone. Individuals who had not responded by week 3 but had responded by week 6 were assessed using the MADRS, BDI-II, and HRSD at weeks 6 and 8. Anxiolytics and hypnotics were allowed at stable levels before the start of the rTMS therapy, but not 8 h before an EEG recording. During the study period and for at least 5 days before EEG recording and the first rTMS session, no antidepressants, antipsychotics, or anticonvulsants were allowed. It is necessary to mention since it can be quite a serious ethical and medical issue to prohibit the use of antidepressants for patients with depression, whose pre-treatment HRSD scores indicate severity of about 30 points, individuals who had these conditions were excluded from participating in this study. At the beginning and end of the treatment, resting EEG was obtained. The participants' EEGs were recorded for 5 min in each session while they sat in a comfortable chair in a shielded room with their eyes closed ([Ahmadlou et al., 2012](#); [Hasanzadeh et al., 2019](#)). The participants were told not to fall asleep during the experiment. Responding to treatment is defined as more than 50% decrease in BDI-II scores or HRSD or by $BDI \leq 8$ ($HRSD \leq 7$) which indicates remission ([Hasanzadeh et al., 2019](#)).

2.3. EEG recording and pre-processing

All EEG recordings took place at the Neural Clinical Neuroscience Centre, Tehran, Iran. A 32-channel eWave32 amplifier was used for EEG signal recording which followed the International 10–20 System of electrode placement on the scalp. The amplifier, made by ScienceBeam¹, has a sampling rate of 1k samples/second, allowing us to precisely study the temporal dynamics of information processing in the brain (with a referential montage, where the reference electrode was placed in the FCz position; [Raeisi et al., 2020](#); [Seraji et al., 2021](#); [Seraji, 2021](#)). We applied standard pre-processing procedures in EEGLAB (available at <https://sccn.ucsd.edu/eeqlab/> version 2021) to reduce noise and artifacts from the EEG signals ([Ebrahimzadeh et al., 2021b](#)). First, the sampling rate of the signal was reduced to 250 Hz, and

filtered by a Butterworth band-pass filter at 1–60 Hz. Then, all the channels were reviewed, and those with a standard deviation greater than ± 3.1 from the mean standard deviation (across all channels) were excluded as the channels that contain artifact. For eliminating the power-line noise at 50 Hz, the Clean Line algorithm was used ([Ebrahimzadeh et al., 2019a,b](#)). The advantage of this algorithm over the notch filter is that it adaptively estimates and removes sinusoidal artifacts without creating band-holes in the EEG power spectrum ([Ebrahimzadeh et al., 2021b](#)). Next, the ICA algorithm was applied on the EEG signal and the irrelevant components corresponding to eye blink, eye movement, cardiac pulsatile, muscular tension, swallowing, or machine vibration were visually identified using the component's scalp map, spectral power activity, and spectral power distribution ([Ebrahimzadeh et al., 2019c, 2021c](#); [Sadjadi et al., 2021](#)). After identifying all the artifact components, the data were re-composed without them. Finally, the average reference was used to re-reference all of the data. We kept 300 s of each subject's EEG signal to equalize the length of data for all participants, taking into account the parts of the EEG data that were eliminated.

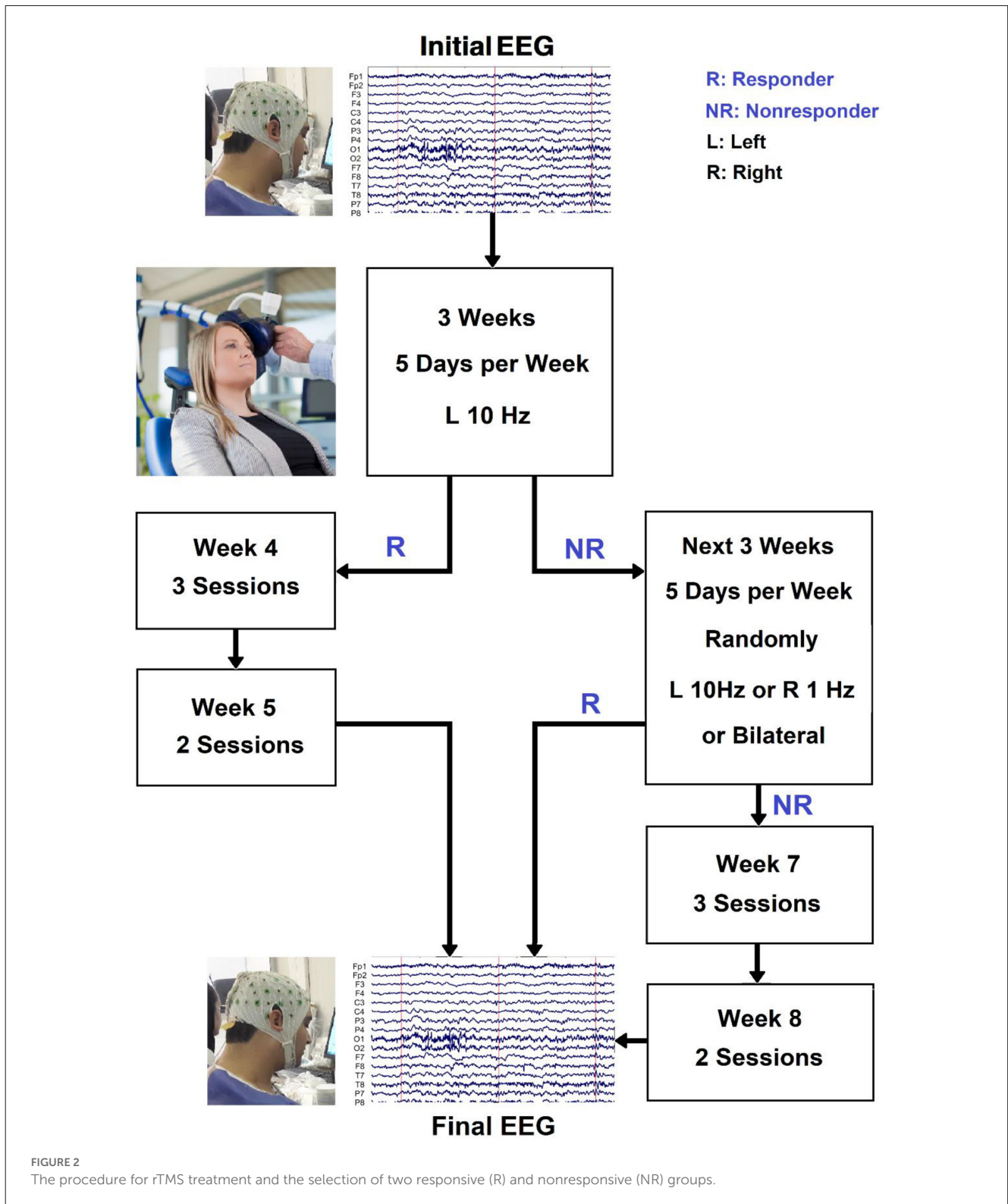
2.4. LORETA

To determine brain electrical sources, researchers use the LORETA method (low resolution electromagnetic tomography). LORETA is a Laplacian-weighted minimum norm algorithm that relies on the patient's prior neuroanatomical and physiological knowledge as well as a mathematical restriction. The method is based on projecting the brain's electric activity onto all of the points in a 3D grid. Unlike dipole source modeling approaches, every site whose activity is reformed is considered a potential source ([Jaworska et al., 2019](#)). As a result, the model does not require a predetermined number of sources. The smoothest spatial distribution is chosen by minimizing the Laplacian of weighted current sources. The idea is that neighboring voxels should have an electrical activity that is as comparable as possible, such as the same orientation and activation. The LORETA approach, being time dependent, allows for an inverse solution by using spatial coefficients as input. It is sufficient to have a one-time sample to generate a combination of sources. That said, we separated EEG activities into time windows for a solution, and investigated the 5-s segments.

2.5. Independent component analysis (ICA)

Independent Component Analysis (ICA) is used for a statistical decomposition of multi-channel EEG signals for source separation. The EEG signal is made up of a variety of contributions, and utilizing this method, independent components can be isolated from the mixed signals. ICA converts a multivariate random signal into a signal with mutually independent components. We extracted the time series of each component of the DLPFC area and then elicited the features. We chose the first three components to study since each participant had at least three components in the DLPFC region, and the first three components are more likely

¹ <http://www.sciencebeam.com/>

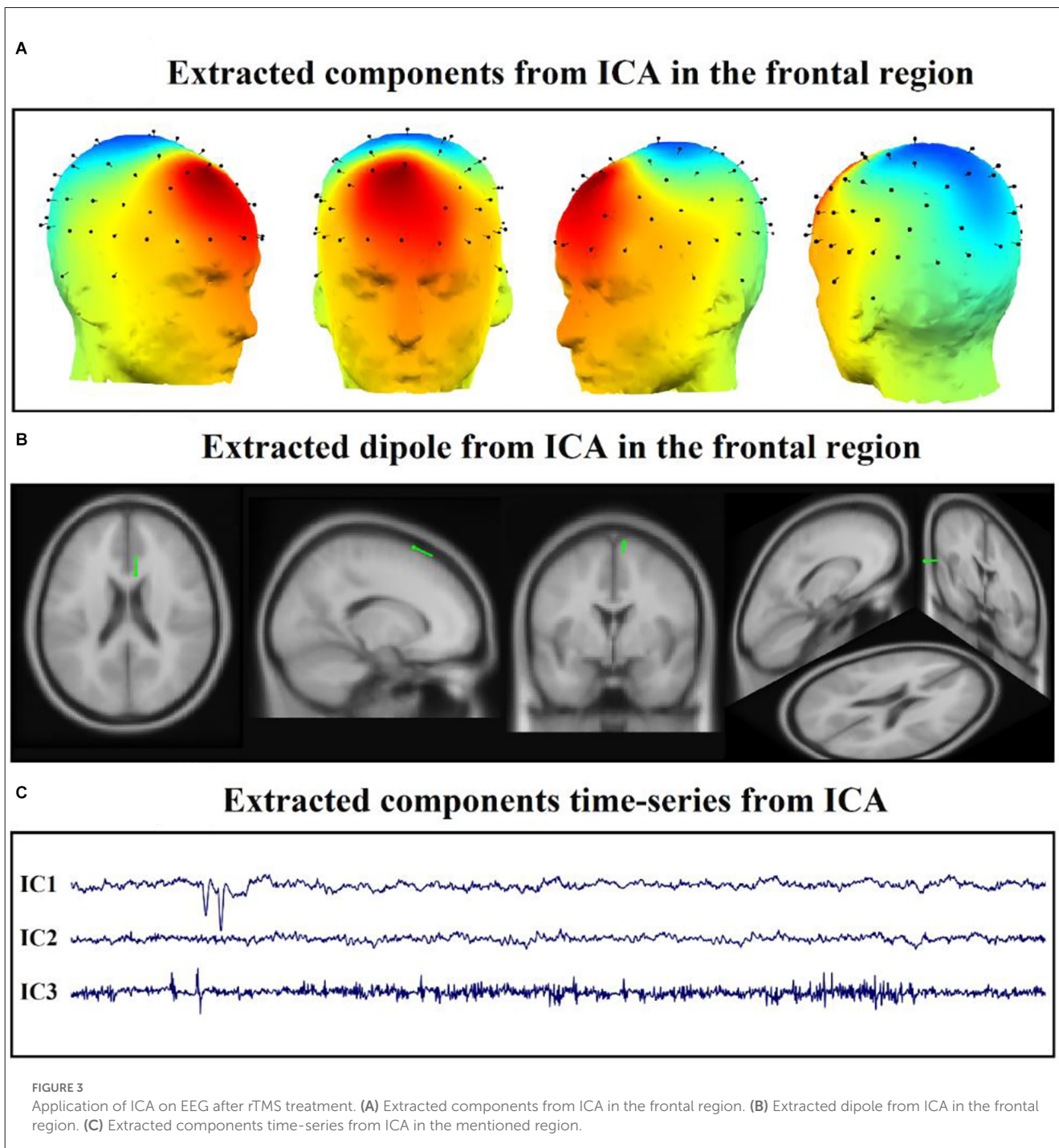


to play a substantial role (Figure 3; Ebrahimzadeh et al., 2019a,c, 2021b,c).

2.5.1. Region of interest (ROI)

Referring to the existing literature, we have defined a region-of-interest method to obtain the volumetric measurements of

DLPFC. In a number of studies, it is shown that the DLPFC site is optimally identified as the midpoint of a line drawn between the F3 and AF3 EEG points (Fitzgerald et al., 2009; Ahdab et al., 2010; Peleman et al., 2010). In addition, authors of Paxinos and Mai (2015), Brodmann (1909), Fischl and Dale (2000), and von Economo et al. (2008) have highlighted the fact that the human cerebral cortex is a highly folded sheet



of neurons with a thickness varying from 1–4.5 mm (overall average being around 2.5 mm). Considering both views, we have determined the ROI as a cylinder that has a circle centered in the middle of F3 and AF3 with a radius of 25 mm and a thickness of 2.5 mm.

2.6. Feature extraction

After extracting the time series of linked components, features extraction is the next stage in predicting treatment

response to rTMS. We evaluated a total of 23 features in four categories: nonlinear, spectral, bispectral, and cordance. Each feature (excluding cordance) was computed for all selected components of both groups of R and NR. Thus, for each measure, we have a feature set containing 23 feature vectors corresponding to three components making a total of 69 features. The next sections describe the measures that were investigated.

2.6.1. Power spectrum

The signal's power in its frequency components is represented by the power spectrum. Delta (1–4 Hz), Theta (4–8 Hz), Alpha

(8–12 Hz), and Beta (12–24 Hz) are common frequency bands. The components time series were used to calculate the average powers. The power spectrum of the components bands was calculated using a Fast Fourier Transform based on the Welch method for each band of each component. The Welch's approach separates the signal of N samples into K data segments of M samples, with D samples possibly overlapping. After that, the overlapping portions are windowed, i.e., multiplied by a symmetric bell-shaped window. The periodogram of each windowed data segment can then be computed using the discrete Fourier transform (DFT). The average adjusted periodogram of all segments is then calculated as the final estimate of the spectrum. In this study, a non-overlapped window with 1,000 samples (1 s length) was chosen for each band of each component.

2.6.2. Cordance

Cordance is a quantitative EEG approach that embodies information from the EEG spectra's absolute and relative powers. Many researchers have used cordance as a way of distinguishing depression and predicting treatment outcomes (Bares et al., 2015; Baskaran, 2016; Bailey et al., 2018, 2019). Age, gender, and the degree of baseline depression have little impact on cordance. The ratio of the delta, theta, alpha, or beta band power over the total power in the full frequency range, which eliminates individual differences, has been termed as the relative EEG power. Some of the previous studies have shown decreased prefrontal cordance of the theta frequency band after the treatment (Cook et al., 2002; Bares et al., 2007, 2015; Cao et al., 2018b; Bailey et al., 2019). Therefore, in MDD, prefrontal theta cordance can be a documented neurophysiological biomarker for predicting response to antidepressants. In this work, theta cordance from extracted component time series in prefrontal was calculated according to the following three steps:

First, the power spectra (P) were averaged for each component. The absolute and the relative power ($P_{s,f}$ and $\bar{P}_{s,f}$ respectively) were expressed as:

$$\bar{P}_{s,f} = \frac{P_{s,f}}{P_{s,f'}} \tag{1}$$

where s is the electrode site and f and f' are the specific frequency band and all bands, respectively.

Second, the absolute and relative EEG powers were normalized by dividing them by the maximum absolute and relative powers in each frequency band (f) as:

$$P_{ANORM(s,f)} = \frac{P_{(s,f)}}{\max(P_s)} \tag{2}$$

$$P_{RNORM(s,f)} = \frac{\bar{P}_{(s,f)}}{\max(\bar{P}_s)} \tag{3}$$

Finally, the cordance values at each electrode site (s) for each frequency band (f), $C_{s,f}$, were calculated as:

$$C_{s,f} = (P_{ANORM(s,f)} - 0.5) + (P_{RNORM(s,f)} - 0.5) \tag{4}$$

In this study, we calculated theta cordance from extracted component time series in prefrontal.

2.6.3. Nonlinear features

Various methods have been proposed over the last two decades for obtaining EEG nonlinear properties to characterize brain activities. The first and most important characteristic of EEG is its dynamic “complexity,” which may be quantified *via* complexity analysis. The degree of unpredictability in time series is mostly represented by the complexity analysis. In this study, we computed four nonlinear features. All nonlinear measures were calculated for epochs of component time series with the length of 3,000 samples and then have been averaged over all epochs. Thus, for every component, we have four values that are corresponding to each nonlinear measure for every subject.

2.6.3.1. Permutation entropy (PE)

Permutation entropy (PE) is a new feature extraction method, which has low computational complexity, robustness, and simplicity (Grova et al., 2008; Berger et al., 2017; Ebrahimzadeh et al., 2018a; Hasanzadeh et al., 2019; Ćukić et al., 2020). It can be used to investigate the local order structure of a dynamic time series and measure the degree of regularity in the EEG data. PE turns an EEG time series into an ordinal pattern sequence. It converts non-stationary time series into a set of ordinal patterns, each of which describes the order relationship between the present and a set of equidistant previous values at a given point in time.

The PE values reported in this work were divided by $\log(m!)$ for normalization. They are dimensionless quantities in the interval [0, 1] (Berger et al., 2017).

For a scalar time series $X(n) = [x(1), x(2), \dots, x(n)]$, the reconstruction time series is:

$$X(i) = [x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)] \quad I = 1, 2, \dots, n(m - 1)\tau \tag{5}$$

where m is “the length of the pattern,” that is the number of sample points included in each pattern and the “time lag” (τ) is the number of samples that spanned each section of the pattern. Then, $x(i)$ is rearranged in increasing order:

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \leq \dots \leq x(i + (j_m - 1)\tau) \tag{6}$$

For m different numbers, there will be $j = m!$ permutations. The vectors $X(i)$ can be mapped to one of the $m!$ permutations. Next, for the time series $X(n)$, the probability of each permutation occurring (p) can be defined as:

$$p_j = \frac{n_j}{\sum_{j=1}^{m!} n_j} \tag{7}$$

where n_j is the number of times the j th permutation is occurring.

The permutation entropy of the time series $x(i) = [x(1), x(2), \dots, x(n)]$ is defined by:

$$H_x(m) = - \sum_{j=1}^{m!} p_j \ln p_j \tag{8}$$

When the time series is random, $H_x(m)$ approaches its maximum value of $\ln(m!)$; when the time series is regular (non-random),

$H_x(m)$ approaches zero. Finally, the corresponding normalized permutation entropy is:

$$PE = \frac{H_x(m)}{\ln j} \tag{9}$$

The smallest value of PE is zero, which means the time series is very regular; and the largest value of PE is one, which means the time series is completely random.

2.6.3.2. Fractal dimension (FD)

Earlier research has shown that EEG complexity analysis using FD may be employed successfully in a variety of clinical settings, implying that FD could be a good indicator of the efficacy of rTMS therapy (Lebiecka et al., 2018). The level of self-similarity of a time series is measured using an FD algorithm, which relates to how many times a pattern in the time series is repeated. FD was calculated using the Higuchi's fractal dimension (HFD) technique (Ahmadlou et al., 2012; Bachmann et al., 2013; Ćukić, 2020). The signal is represented by a sequence $x = [x(1), x(2), \dots, x(n)]$, where n is the total number of samples and $x(i)$ indicates the i th sample of x . From a given sequence, k subsequences X_m^k will be defined as:

$$X_m^k : X(m), x(m+k), \dots, X(m+jk), m = 1, 2, \dots, k, \tag{10}$$

$$j = \text{int}\left(\frac{N-m}{k}\right) \tag{11}$$

where m is the initial time and k is the interval time.

The average length of each subsequence X_m^k can be calculated according to:

$$L_m^k = \frac{1}{k} \sum_{i,j}^k |X(m+ik) - X(m+(i-1)k)| \frac{N-1}{j} \tag{12}$$

where $\frac{N-1}{j}$ is a normalization factor. The total average length for scale k , $L(k)$, is computed as the average of the k values, for $m = 1, 2, \dots, k$, that is:

$$L(k) = \frac{1}{k} \sum_{m=1}^k L_m^k \tag{13}$$

The calculation is repeated for k values ranging from 1 to k_{\max} . k_{\max} is considered to be the number when the slope of the best line fitted to the diagram of $L(k)$ vs. $\frac{1}{k}$ plotted in a log-log plane remains constant ($k_{\max} \geq 2$). The slope of this line is HFD of the time series L . The line is defined by the linear regression coefficient which is determined by the least squares method. $k_{\max} = 16$ is shown to perform the best for this type of signal (Spasic et al., 2005; Lebiecka et al., 2018; Ćukić et al., 2020).

$$\ln L(k) \sim \text{HFD} \ln \frac{1}{k} \tag{14}$$

2.6.3.3. Lempel-Ziv Complexity (LZC)

The rate at which new patterns arise in time series is indicated by a nonlinear dynamic measure. The activities of adding or removing the patterns of the underlying system are examined when employing LZC (Aboy et al., 2006). The LZC method is based on coarse-graining, which converts the signal $s(n)$ into a limited sequence of a few symbols (Li et al., 2008). Traditionally, the coarse-graining procedure turned the signal $s(i)$ into a binary sequence $x(i) = x(1), x(2), \dots, x(i)$. By comparing the signal values with a threshold, the original signal is converted into a binary (0, 1) sequence, with the median of the signal values chosen as the threshold.

$$x(i) = \begin{cases} 0, & \text{if } s(i) < T_d \\ 1, & \text{otherwise} \end{cases} \tag{15}$$

Then by scanning the EEG sequences $\{x(i)\}$ from the beginning, different patterns that appear in the signal are counted by $c(n)$. It has been proven that the upper limit of complexity measure $c(n)$ with a median threshold is:

$$\lim_{n \rightarrow \infty} c(n) = b(n) = \frac{n}{\log_2(n)} \tag{16}$$

The complexity measure $c(n)$ is normalized to $b(n)$, to make the measurement independent of the length of the data, and LZC will be obtained by:

$$LZC = \frac{c(n)}{b(n)} \tag{17}$$

By transforming the signal into a finite sequence $\{x(n)\}$, $c(n)$ will be the complexity of the sequence $\{x(n)\}$ and will denote the number of distinct patterns in the sequence.

2.6.3.4. Correlation dimension (CD)

Based on embedding theory and phase space reconstruction, the correlation dimension of a dynamic system is established. Considering the state points, $x(1), x(2), \dots, x(n)$, new m dimensional vectors can be reconstructed by choosing a time delay t and embedding dimension m .

$$X(i) = [x(i), x(i+t), \dots, x(i+(m-1)t)], \tag{18}$$

$$i = 1, 2, \dots, n - (m-1)t$$

The correlation dimension is obtained by:

$$C(r, n) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \theta(r - |x_i - x_j|) \tag{19}$$

$$D(r, n) = \lim_{r \rightarrow 0} \frac{\ln C(r, n)}{\ln r} \tag{20}$$

where θ is the Heaviside step function which is defined as $\theta(x) = 0$ for $x < 0$ and $\theta(x) = 1$ for $x > 0$; $C(r, n)$ is the correlation integral and $D(r, n)$ is the correlation dimension. The procedure is repeated for increasing m .

2.6.4. Bispectrum features

The spectral band power can show phase variations, but does not reveal interactions between the signal’s frequency components. Note that a signal’s “shape” is determined by its phase, and signals with different wave forms might have the same power spectrum (Acharya et al., 2015).

Bispectrum is a higher-order statistical signal processing technique that examines both the amplitude and the degree of phase coupling of a signal. In the context of EEG signal categorization, bispectrum has demonstrated promising results. The two-dimensional Fourier transform of a signal’s third order correlation is defined as the signal’s bispectrum $Bis(f_1, f_2)$. Bispectrum divides a signal’s skewness (third order moment) throughout all frequencies, characterizing the intensity of interaction between all frequency pairings in the target band.

$$Bis(f_1, f_2) = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \right) E [X(f_1 + f_2) X^*(f_1) X^*(f_2)] \quad (21)$$

where $X(f)$ is the Fourier transform of a time series $x(t)$, * is the complex conjugate, T represents time, and E denotes the expected value. Indeed, bispectrum indicates cross-correlation between frequency components in a two-dimensional frequency plot. The phase coupling information between the frequency components at f_1, f_2 , and $f_1 + f_2$, can be extracted from $Bis(f_1, f_2)$. The degree of phase coupling between frequencies components is obtained by the following normalized bispectrum:

$$Bis_{norm}(f_1, f_2) = \frac{Bis(f_1, f_2)}{\sqrt{P(f_1) P(f_2) P(f_1 + f_2)}} \quad (22)$$

The power spectrum is denoted by $P(f)$, and the magnitude of the normalized bispectrum has a value between 0 and 1. The squared magnitude of the normalized bispectrum will be 1 if the Fourier components at the frequencies f_1, f_2 , and f_1+f_2 are perfectly phase-coupled in every realization, and 0 if they are completely random.

To distinguish the distribution of the bispectral plots of distinct EEG data, quantitative characteristics must be defined. The bispectrum of a genuine signal is uniquely described in the triangle $0 \leq f_2 \leq f_1 \leq (f_1+f_2) \leq 1$, assuming no bispectral aliasing. For characterization of the entire bispectrum, features are defined by integrating along straight lines going across the non-redundant region. Three features were collected from the bispectrum region in this investigation. These features are defined as follows.

Average of magnitude:

$$M_{ave} = \frac{1}{L} \sum_{\Omega} |Bis(f_1, f_2)| \quad (23)$$

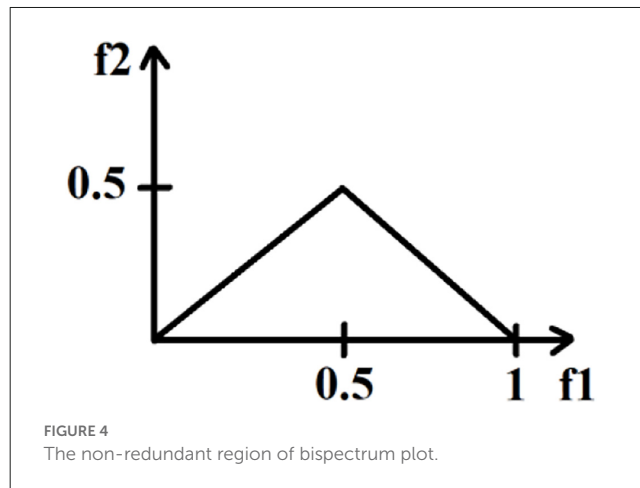
where L is the total number of sample points in the bispectral density array and Ω is the triangle region shown in Figure 4.

Bispectral entropies:

$$P_1 = -\sum_n p_n \log p_n \quad (24)$$

$$p_n = \frac{|Bis(f_1, f_2)|}{\sum_{\Omega} |Bis(f_1, f_2)|} \quad (25)$$

$$P_2 = -\sum_i q_i \log q_i \quad (26)$$



$$q_i = \frac{|Bis(f_1, f_2)|^2}{\sum_{\Omega} |Bis(f_1, f_2)|^2} \quad (27)$$

where P_1 and P_2 are the normalized bispectral entropy and normalized bispectral squared entropy which are also between 0 and 1.

2.7. Feature selection

Before performing classification, we use feature selection for two main reasons: first, to eliminate the irrelevant features which can raise the complexity of classification and lower its accuracy, and second, to address the dimensionality problem. That said, after extracting EEG features, we applied a genetic algorithm (GA) to pick out the most discriminating features. GA is a group of computational models based on natural selection and genetics laws. In a high-dimensional space, it is considered as a stochastic strategy that outperforms deterministic optimization strategies. The GA begins with a population of individuals that represent a potential solution to a particular optimization issue and evolves over generations to a group of more optimal or fit individuals. Then, it replicates the offspring using basic genetic procedures such as selection, crossover, and mutation. A fitness value is assigned to each individual or community, and the fitness of the candidate individual is assessed. Finally, GA selects the best individuals from the existing population (Amoozegar et al., 2019).

2.8. Classification

We suggested a novel strategy to predict the rTMS treatment response based on classification analysis to separate responders from non-responders by the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) classifiers. We ran our analysis on a different set of extracted features from the component time series. First, we assessed the capacity of the researched features to predict treatment response by applying classification to each measure individually. Then, in a second classification study, a mixture of features from each of the four categories (nonlinear, spectral, bispectral, and cordance) were

utilized to classify R and NR. Finally, we assessed a combination of all features for classifying R and NR. For all analyses, the classification process detailed below was comparable. After feature extraction, the feature sets were separately standardized by the z-score normalization to minimize amplitude variations caused to differences in subjects and electrode placements. Afterwards, we employed GA to pick the most informative features.

In all analyses, the number of features was determined by a trial-and-error method to discover a value that maximizes accuracy while avoiding overfitting of the classifier. The square root of the total number of features in the related study was used to match this ideal number of features in our analysis, except for prefrontal cordance analysis, where due to the small number of features, we used all features.

A 10-fold cross-validation method was used to assess the classifiers' performance. Except for two or three groups, the EEG data of 88 subjects were separated into 10 parts, with the number of signals in each part being equal. The classifier was trained using nine parts and the remaining part was used to test it. This process was performed 10 times, and the average and standard deviation of accuracy, sensitivity, and specificity were obtained.

2.8.1. Multilayer perceptron (MLP)

A three-layer MLP with the error back propagation method and variable learning rate was used. For component time series (Ebrahimzadeh et al., 2011, 2018a,b, 2019d; Ebrahimzadeh and Pooyan, 2013; Nikravan et al., 2016), the input layer has the same number of nodes as the input vector length. The output layer, on the other hand, only has one node, which means that only two classes can be classified. The optimal number of neurons in the hidden layer was found by selecting and training all feasible combinations of the selected numbers of neurons in the hidden layer. It is worth noting that the entire training procedure was based only on the training data. We moved on to testing the network using the testing data after the training is completed. The network did not use the testing data while determining the best architecture to maximize the network's generalization.

A linear transfer function and a sigmoid function were used in the output nodes and the hidden layer, respectively (Ebrahimzadeh et al., 2013). The network training continues until the error becomes less than 0.01, or 1,000 training iterations are completed (Ebrahimzadeh et al., 2014).

2.8.2. K-Nearest neighbor (KNN)

After saving labeled feature vectors, this classifier calculates the shortest distance between saved and new feature vectors (Ebrahimzadeh et al., 2018a,c). The KNN algorithm is divided into three steps: (a) calculating distances between all previously classed samples; (b) selecting the K samples with the least distance values; and (c) approving new data. A new sample will be added (classified) to the largest cluster from the K selected samples. We looked at the values of K from 1 to 12 to compare with Hasanzadeh et al. (2019) and found that K = 7 delivers the best results. To make the tables easier to read, we used three K values (3, 7, and 12).

2.8.3. Support vector machine (SVM)

SVM is a machine-learning technique that has proven to be effective in a variety of classification tasks. It focuses on the training examples near the edge of the class descriptors to find the optimal separating hyperplane (the plane with the largest margins) between the two classes of the training samples in the feature space. This method not only fits an ideal hyperplane, but also effectively uses less training samples, resulting in good classification accuracy with small training sets.

SVM is a well-known supervised learning model for classification and regression. The primary principle of SVM is to transfer the input data from the N-dimensional input space to the M-dimensional feature space $M > N$, where the data classes can be separated linearly (Ebrahimzadeh et al., 2018a). In other words, the SVM is a statistical learning theory-based extension of nonlinear models of the generalized portrait algorithm (Nikravan et al., 2016; Ebrahimzadeh et al., 2019d). The purpose of regression is to choose the best model from a group of models (known as estimating functions) in order to accurately anticipate future values. The estimation function for support vector regression is:

$$f(x) = (w \cdot \Phi(x)) + b \quad (28)$$

where $w \in R_n$, $b \in R$ and Φ is a nonlinear function that maps x into a higher dimensional space. W and b are the weight vector and bias, respectively. The weight vector (w) can be written as:

$$w = \sum_{i=1}^L (\alpha_i - \alpha_i^*) \quad (29)$$

By substituting Equation (29) into Equation (28), the generic equation can be rewritten as:

$$f(x) = \sum_{i=1}^L (\alpha_i - \alpha_i^*) (\Phi(x_i) \cdot \Phi(x)) + b. \quad (30)$$

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (31)$$

where the function $k(x_i, x) = (\Phi(x_i) \cdot \Phi(x))$ is known as the kernel function and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_l)$ is the vector of non-negative Lagrange multipliers.

The choice of kernel functions and kernel parameters depends mainly on the application. Among the useful kernel functions are radial basis functions (RBFs) and polynomial kernel functions:

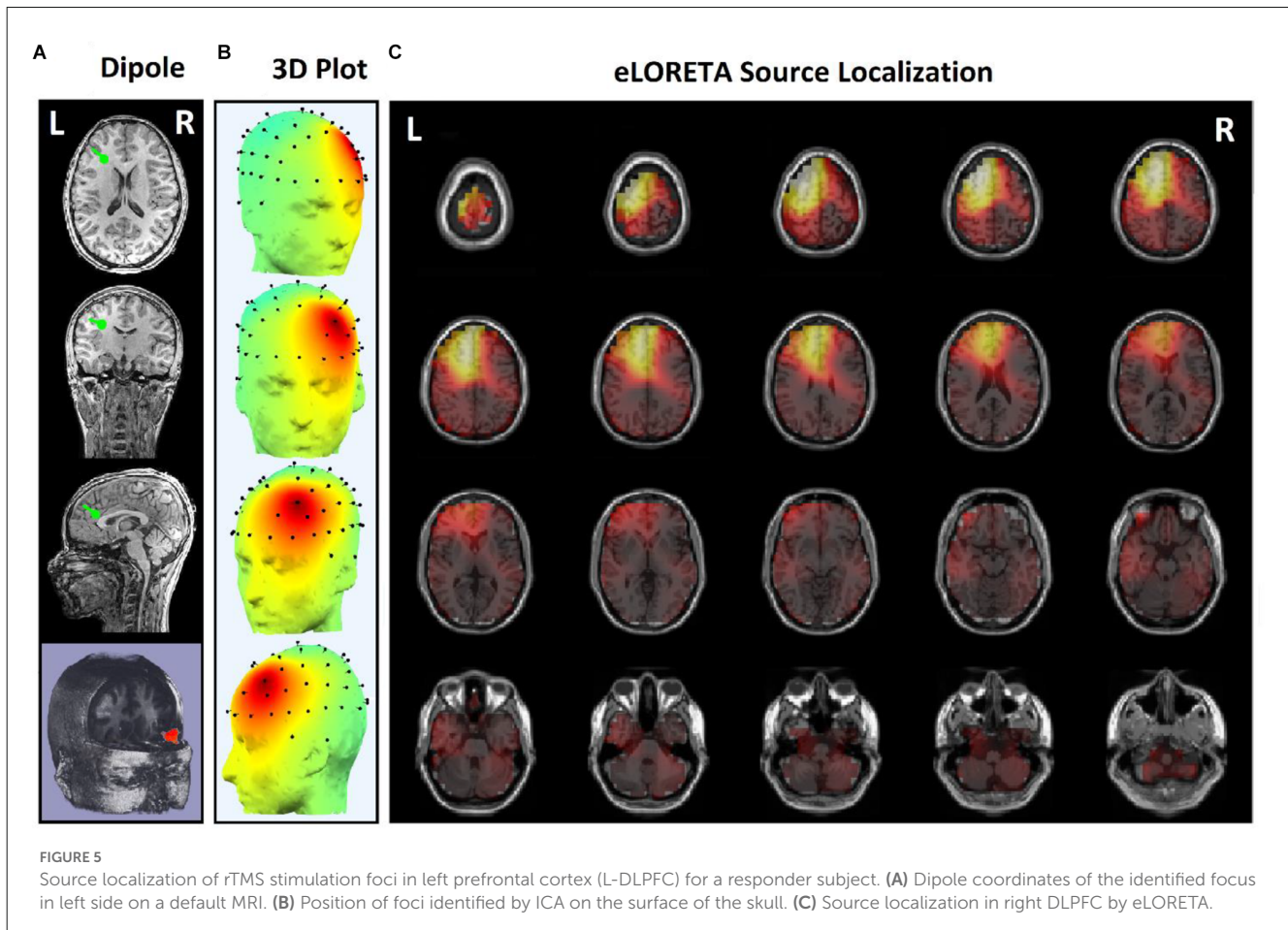
$$\left\{ \frac{-|x - xi|^2}{2\sigma^2} \right\} \quad (32)$$

$$[(x * x_i) + 1]^d \quad (33)$$

where σ and d are kernel width and order, respectively, which were experimentally defined to achieve the best classification result. In this work, RBFs and polynomial kernel functions were used with different sigma values ($\sigma = 0.8, 1, 1.2$) and orders ($d = 1, 2, 3$).

2.9. Statistical comparison

Statistical testing seeks to understand how a system works, whereas machine learning approaches seek to anticipate future



behavior (unobserved consequence; Bzdok et al., 2018). Therefore, they can be used in conjunction with the current investigation to provide a more thorough assessment of the studied features' potential to detect R and NR responses to rTMS treatment. We also ran a statistical test to compare the results of our data to those of earlier studies that used statistical analysis. Since some of the data did not have a normal distribution, we used analysis of variance (ANOVA) to compare every feature between the two groups.

2.10. Evaluation

Accuracy (AC), Sensitivity (SN), Specificity (SP), and Precision (P) are used to assess the proposed method's capacity to predict response to rTMS treatment. TP stands for true positives (correctly predicted R), TN for true negatives (correctly predicted NR), FN for false negatives (incorrectly predicted NR), and FP for false positives (incorrectly predicted R).

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

$$AC = \frac{TP + TN}{TP + TN + FN + FP} \quad (34)$$

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

$$SN = \frac{TP}{FN + TP} \quad (35)$$

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

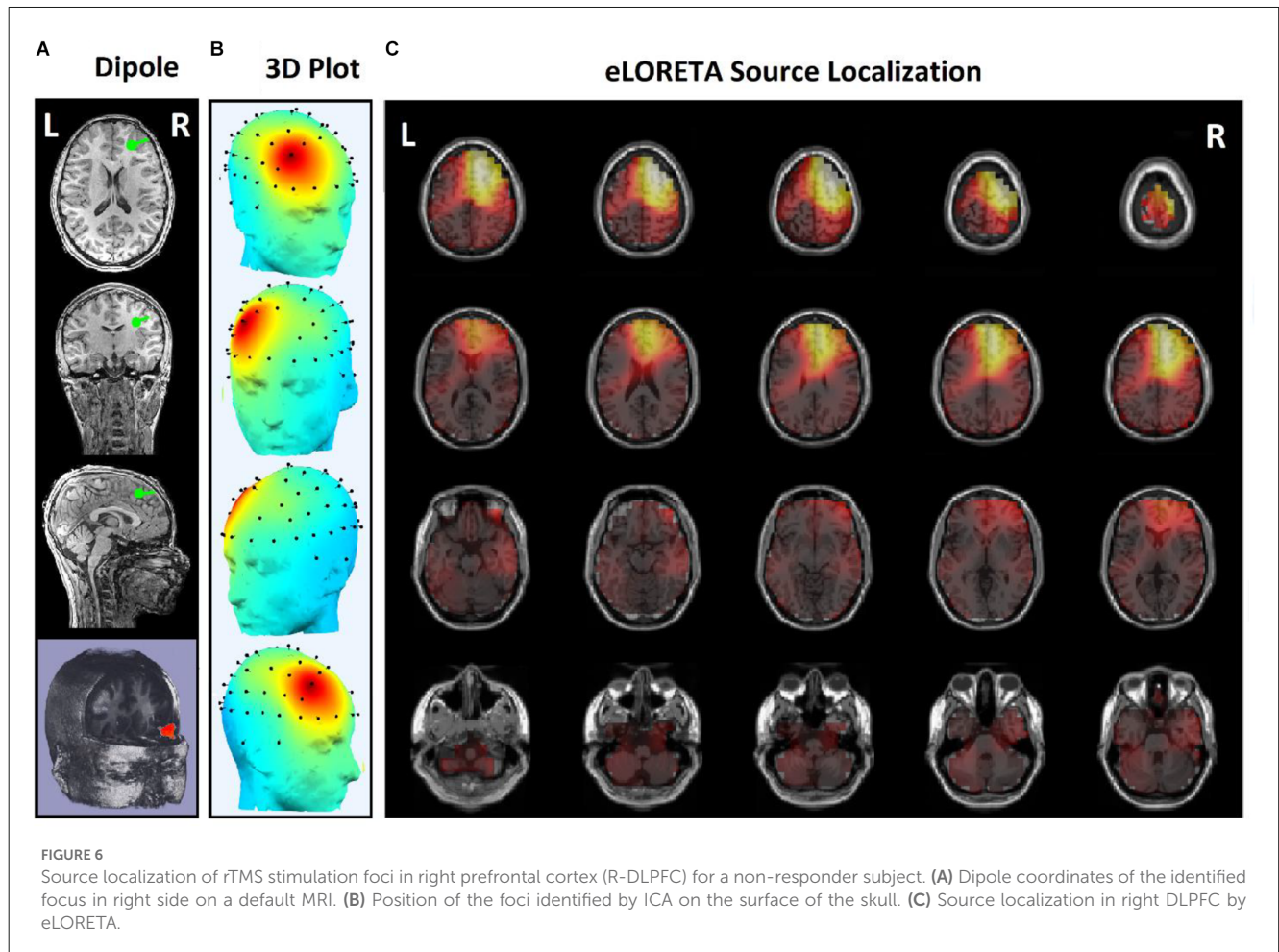
$$SP = \frac{TN}{TN + FP} \quad (36)$$

Precision (P): the ratio of predicted positive cases that were correct

$$P = \frac{TP}{FP + TP} \quad (37)$$

3. Results

Since the effects of stimulation on inactive neural networks in the frontal and pre-frontal areas as brain sources are always greater than EEG channels, we first identified the rTMS-stimulated foci in the mentioned areas and then extracted nonlinear features from the desired foci. In order to identify the relative foci, first the location of the focus in the standard MNI is visually determined and after confirming the dipole coordinates of the representative of that focus in the frontal region, the desired component is determined as the excitable focus. The obtained results indicate the presence of foci in the L-DLPFC (Figure 5), R-DLPFC (Figures 6, 7), and center



of DLPFC (Figure 8) each of which is confirmed by the dipole coordinates and eLORETA results.

Figure 5 shows the position of a focus identified by the ICA algorithm, which is also confirmed by eLORETA algorithm. The focus shown on the left side of DLPFC belongs to a responder subject.

Figures 5 and 6 are the identified foci on the right side of the prefrontal region that belong to a NR subject and a R subject, respectively. It is noteworthy that although the NR subject has a recognizable focus in the mentioned area, the extracted features reveal a significant difference between the NR and R subjects in terms of prediction of the rTMS treatment which will be explained further.

In some subjects undergoing bilateral treatment, the main focus is localized in the center of the frontal region. These patients often have symptoms of depression as well as insomnia and anxiety (Figure 7).

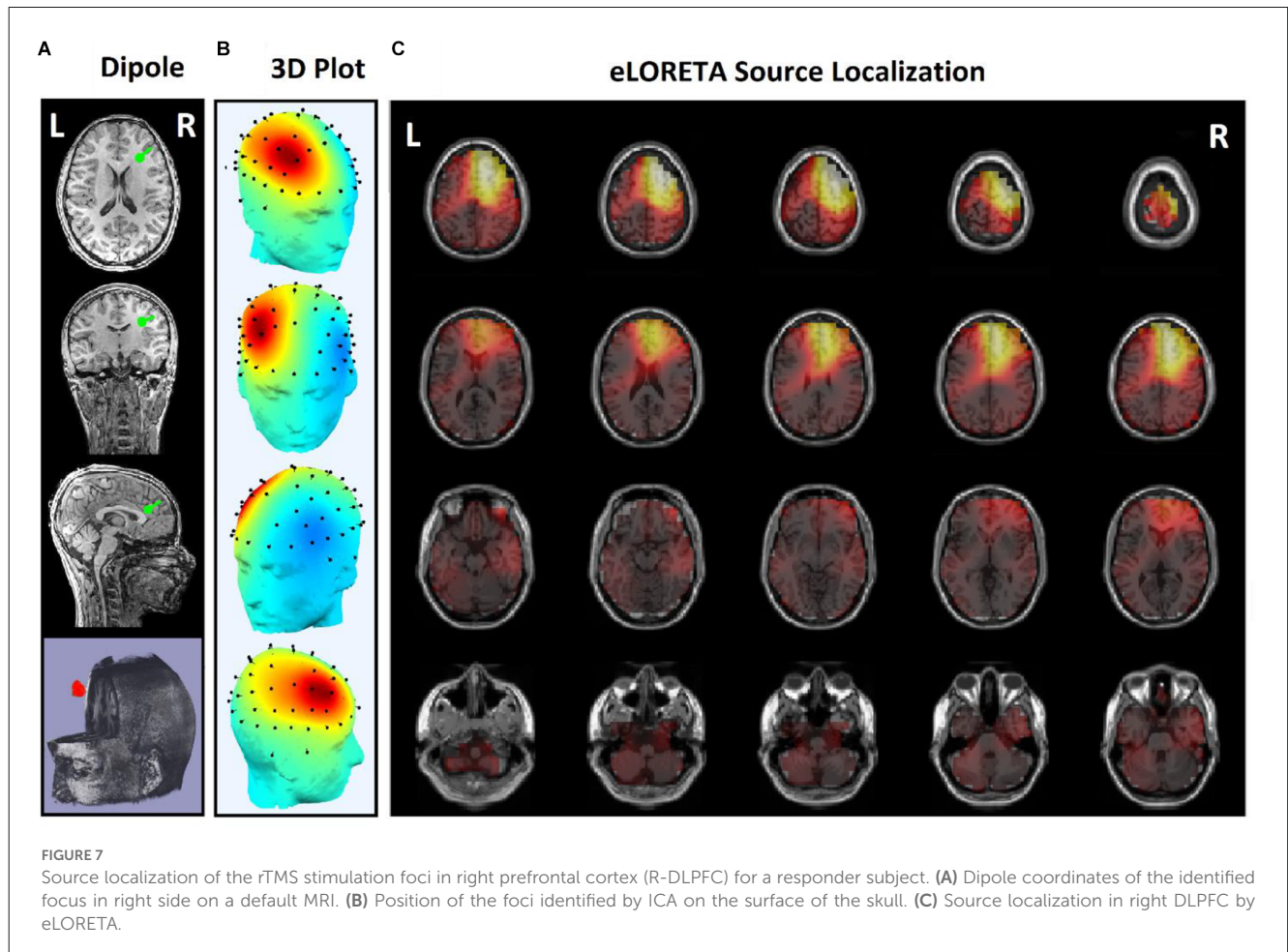
3.1. Analytical results

We extracted 23 features from each component time series: four nonlinear features, including PE, LZC, FD, and CD, four power features, which contain delta, theta, alpha, beta bands, and 12 bispectrum features, i.e., BispSL, Bisp2M, and BispEn in all

bands, and three cordance features, i.e., IC1CORD-T, IC2CORD-T, and IC3CORD-T features in theta band. Then, by applying the GA feature selection, 12 features were selected. Our study shows that these selected features from the relevant components can be used as suitable markers for prediction of rTMS treatment response. In other words, the components identified in DLPFC can be a more suitable platform and tool for extracting effective features.

Table 2 depicts the efficiency of the 23 selected features in predicting rTMS treatment response in MDD, which was evaluated by analysis of variance (ANOVA) on the learning set. According to this table, all of the bispectrum, cordance, and non-linear features had acceptable discriminating power, as implied by *P*-Value. The results of ANOVA test in spectrum features also showed that T, A, and B power differ significantly between R and NR, while due to the high *P*-Values of frequency peaks in the Delta band, these features cannot distinguish the two groups with statistical significance. Therefore, we used 22 features with statistically significant discrimination, excluding the frequency peaks in Delta band from the aforementioned features to form the feature vector.

In this study, our main focus is to use the genetic algorithm as a more advanced method of feature selection. After an initial evaluation of features, their optimal combination, consisting of 12 features, was selected by GA. The SVM classifier parameters and kernel width (σ) must be chosen with attention and caution to optimize the learning cost and prediction performance. To do



so, we compared the classifier's performance by evaluating the error function using an independent validation set and selecting the classifier with the minimum error relative to the validation set. Because this technique can cause overfitting in the validation set, the test dataset was used to confirm the performance of the chosen classifier. In other words, based on a 10-fold cross-validation, the data were divided into 10 equal parts and in each stage, 10% of the data were used as the test dataset and the rest of the data as the training dataset. In fact, nine of the observations were selected as test data and 79 as train data, and this process was repeated 10 times. The optimum values of the parameters, chosen when the error on the validation dataset reached a minimum, were 0.7 for σ and 3 for order (d).

Table 3 illustrates the performance of the two classifiers with variable parameters, i.e., SVM and KNN. The features were used as input of GA to reduce their number and improve the proposed algorithm's performance. For classification, we utilized SVM and KNN. We assessed the performance of classifiers with various kernels and Ks, such as polynomials of order 1, 2, and 3, and RBFs with various sigma values ($\sigma = 0.7, 1, 1.3$) for the SVM classifier. We tested values of k from 1–13 and found that $k = 7$ in the KNN classifier achieved the best results. To reduce the complexity of **Table 3**, we display four k -values (3, 7, 10, and 13). It is highlighted that the input of the classifiers is the components features extracted by GA.

Four measures based on Equations (34) to (37) were used to evaluate the performance of the proposed approach. If for example, a responder (R) is correctly classified as R, it is a TP. On the other hand, if an NR is classified as NR, it is a TN. Any NR that is wrongly classified as an R will result in an FP, whereas any R that is mistakenly classified as an NR will result in a FN. **Table 4** summarizes the results of the test data classification for each class.

The obtained sensitivity, specificity, precision, and accuracy of the proposed method are shown in **Table 5**. Note that the SVM classifier outperformed the other classifiers and achieved 94.31% accuracy, 95.65% sensitivity, 92.85% specificity, and 93.61% precision, respectively.

We applied feature sets extracted from components time-series to classifiers in the following forms to evaluate the capability of the selected features in the prediction of response to rTMS treatment: each feature independently, the combination of features in each group of nonlinear, spectral, bispectral, and cordance features. Our proposed classification approach was evaluated using the following criteria: classification accuracy, specificity, sensitivity, and precision. **Table 6** shows these criteria in relation to each measure which were calculated based on the three components with the highest average λ (weight of extracted independent components) in the frontal area and reduced to 12 features by GA. The evaluation of classification by the combination of studied features is also given in **Table 6**. As can be seen, combinational features are more capable than the

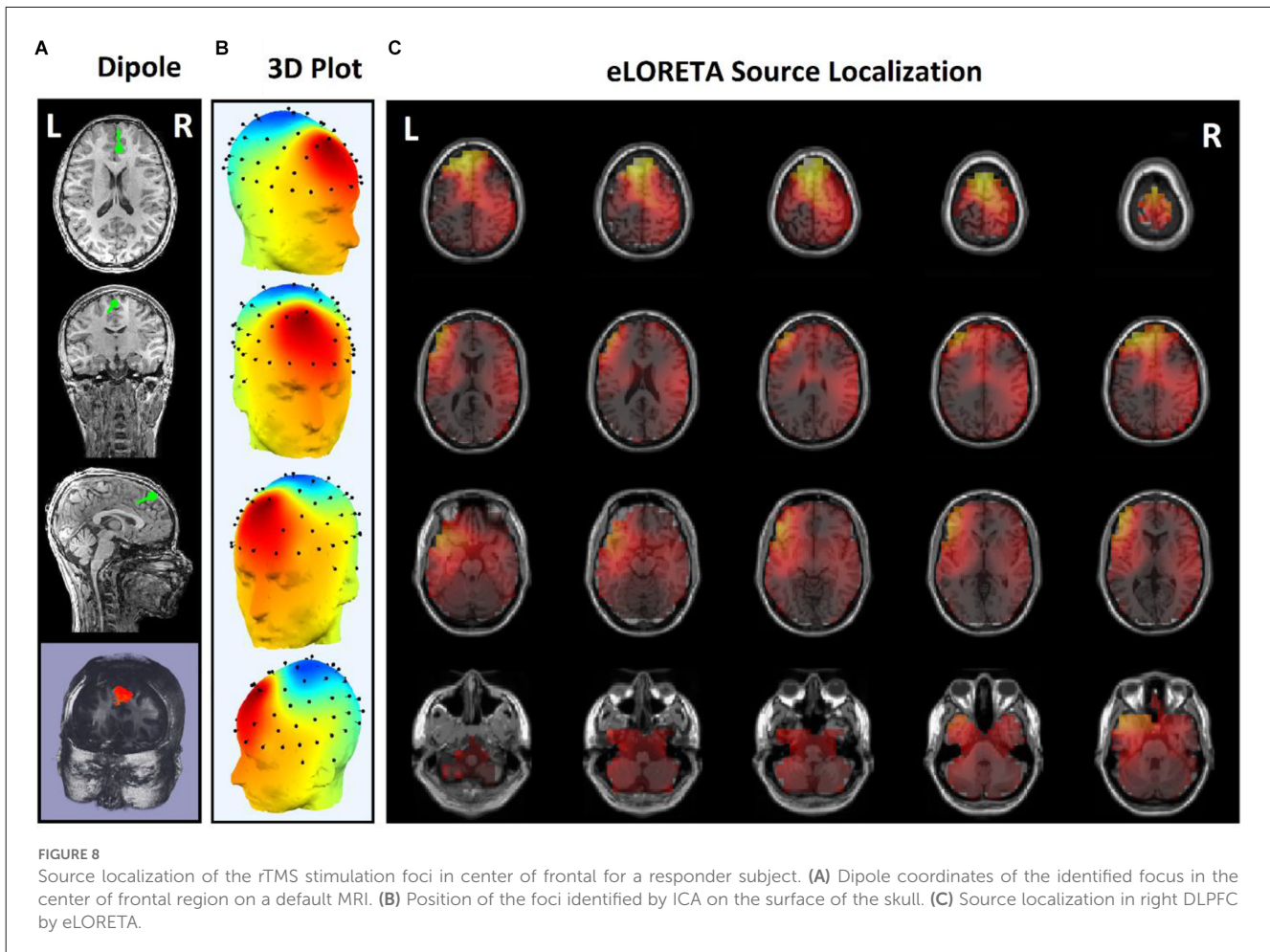


TABLE 2 Results of the ANOVA test.

Extracted features from components	P-Value
Nonlinear (PE, LZC, FD, CD)	$0 \leq 0.0001$
Power (Theta, Alpha, Beta)	$0 \leq 0.0001$
Power (Delta)	0.04
Bispectrum (BispSL, Bisp2M, and BispEn in all bands)	$0 \leq 0.0001$
Cordance (IC1, IC2 and IC3)	$0 \leq 0.0001$

individual features which is why they have been used in this study as the input to predict the rTMS treatment response.

4. Discussion

Identification of biomarkers for predicting the therapeutic outcome of antidepressant treatment by rTMS is the main goal of the current research as it can transform the lengthy process of finding the right treatment for patients with MDD. We have explored the potential of pretreatment cortex activity and extracted features from components as a putative biomarker of treatment response.

Given that EEG channels do not always represent the foci stimulated by rTMS, in this study, we focused on the extraction

of affected foci in the frontal region. The components in the frontal lobe indicate the neural behavior of networks that are activated by electromagnetic stimulation and point to the sources activated during treatment. It seems that one of the factors that can elucidate the improved results of this article, compared to those of previous studies, is the analysis of temporal behavior of the relevant sources rather than the EEG channels. As the classification results illustrated, our proposed method led to high classification accuracy. Table 7 compares our results with those of the previous studies that applied machine learning techniques for prediction of MDD treatment response.

In an attempt to formulate a method for prediction of rTMS clinical effects, we obtained various features from pretreatment resting-state EEG of patients with MDD who had undergone the rTMS procedure, including 34 non-responders and 34 responders. While much of the literature is dedicated to the power of alpha and theta frequency bands (Bailey et al., 2018, 2019), our data analysis results suggest that beta power, when derived from the relevant component time-series, can make for a prognostic biomarker of more significant potential. Similarly, a number of studies such as Lieber and Prichep (1988) and Knott et al. (2000) show that a high rate of beta activity is associated with the depression level. It can then be concluded that the high beta power is linked to lower levels of treatment responsiveness. As for other frequency bands, although theta and alpha bands produced a rather high classification accuracy, only alpha and delta power could be used

TABLE 3 Performance of the KNN and SVM classifiers with different parameters using the combined feature vector.

Classifier	Sensitivity	Specificity	Accuracy
SVM-Poly (d = 1)	81.37%	97.12%	93.52%
SVM-Poly (d = 2)	83.24%	97.40%	92.24%
SVM-Poly (d = 3)	83.63%	97.41%	93.72%
SVM-RBF ($\sigma = 0.7$)	82.61%	98.77%	93.21%
SVM-RBF ($\sigma = 1.0$)	81.17%	98.87%	93.19%
SVM-RBF ($\sigma = 1.3$)	81.44%	97.63%	92.75%
KNN (K = 3)	77.26%	96.37%	92.37%
KNN (K = 7)	80.36%	97.12%	93.18%
KNN (K = 10)	78.12%	96.22%	89.17%
KNN (K = 13)	76.44%	94.22%	87.64%

TABLE 4 Classification results by SVM, KNN, and MLP, using combined feature vectors.

	TP	FP	FN	TN
SVM	44	3	2	39
KNN	41	6	5	36
MLP	42	5	4	37

to statistically differentiate between the two groups (Olejarczyk et al., 2020). Our results confirm that responders exhibit a greater pretreatment alpha power compared to non-responders, which has also been observed in other studies including (Ulrich et al., 1984; Suffin and Emory, 1995; Knott et al., 2000; Bruder et al., 2001, 2008; Lebiecka et al., 2018). Bruder et al. (2008) elucidated that this increased alpha power could point to the correspondence between low arousal and low serotonergic activity. They showed that the fact that 5-HT mediates arousal and serotonergic activity might be suggestive of the activity of the mesencephalic raphe nuclei and cortical afferents (Bruder et al., 2008).

Heller et al. (1995) linked depression to dysfunction of temporoparietal mechanisms which may mediate emotional arousal. They found that delta power was lower in rTMS responders than in non-responders, which was also highlighted in other studies such as Knott et al. (1996) and Knott et al. (2000). Our results depict the significant differences in frontal delta power of responders and non-responders. In comparison, however, the power of theta frequency band has been reported to be of less relevance since both responders (Woźniak-Kwaśniewska et al., 2015) and non-responders (Arns et al., 2012) seemed to have high theta power. We also detected no major differences between the two groups with relation to the theta power. This was also the case in a previous study (Cook et al., 1999). Furthermore, we found the nonlinear features to be of great potential in predicting rTMS treatment response, while they have been relatively overlooked in the literature. The correlation of dimension (CD), for example, was considerably lower in responders, which produced a high classification accuracy of 87%.

We also investigated the predictive abilities of frontal and prefrontal theta cordance which led to classification accuracy of 80.4% and 78.3%, respectively. Our statistical analysis shows that these measures fail to differentiate between the two groups. This is in line with the results of other studies such as Arns et al. (2012) and Bailey et al. (2019).

To evaluate the outcomes, we applied different classifiers and achieved the accuracy of classification using the determined features both individually and in combination (Table 6). Then, we drew a comparison between the obtained results and those of previous works such as (Hasanzadeh et al., 2019; Table 7). As shown, the proposed method has outperformed the previous ones, except for the result of the cognition output in the study of Al-Kaysi et al. (2017), where they aimed to predict mood and cognition output separately. To this end, they designated 0 or 1 values to mood and cognition based on whether their scores from the middle of treatment had decreased or increased compared to the baseline scores. They then used machine learning techniques to predict these values, also known as the outputs, based on the EEG features. It should be highlighted that after the treatment, the cognition output exhibited alterations in the Symbol Digit Modalities Test (SDMT). The results of cognition output are irrelevant to depression therapy prediction because SDMT is not a depression severity rating scale. The mood prediction, on the other hand, highlighted changes in MADRS with a comparatively low accuracy of 76%. A similar study (Bailey et al., 2018) employed both mood and EEG measures and achieved higher accuracy.

The mood measures are generally subject to expert rating and therefore can be different from one to another. Since the proposed method is solely based on the EEG activity, we believe it is more appropriate than that of the mentioned study. In addition, Bailey et al. (2018) separated the two groups of responders and non-responders using two sessions of EEG, one prior to the treatment and one after a week of treatment, whereas we have used only one session of pretreatment EEG and obtained high classification results (accuracy = 91.3%, specificity = 91.3%, sensitivity = 91.3%). The fact that only one session of EEG recording is required, brings to attention the efficiency of the proposed method. It also lifts the financial and mental burden of undergoing a week-long, possibly ineffective treatment.

The major novel aspect of this study was applying predictive analytics and machine learning on the component time-series extracted from scalp EEG to interpret and summarize the neural activity in DLPFC. We were able to identify outstanding features from a big amount of data using machine learning to recognize outcome predictors that were previously undetectable. This capability is greatly enhanced by using time series of foci identified as stimulated neural networks. In fact, the proposed

TABLE 5 Classifier performance in percentage, for 10-fold cross validation and combined feature vector.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
SVM	94.31	95.65	92.85	93.61
KNN	87.50	89.13	85.71	87.23
MLP	89.77	91.30	88.09	89.36

TABLE 6 Accuracy of SVM, KNN, and MLP classifiers with the selected features (individually and in combination) for classifying NR and R.

Features	SVM	KNN	MLP
Non-Linear	84.91%	76.82%	78.24%
Power	91.82%	78.13%	80.66%
Bispectrum	89.23%	81.67%	82.15%
Cordance	76.18%	82.73%	71.63
Combination	94.31%	87.50%	89.77%

TABLE 7 Comparison of classification results of studies that applied machine learning techniques for prediction of MDD treatment response with those of the current study.

Features	Accuracy (%)	Specificity (%)	Sensitivity (%)
Band Power (Metin et al., 2020)	80.0	-	-
Power Spectral, eLORETA, multiscale-entropy-based, microstate-based (Zhdanov et al., 2020)	79.2*	91.0	67.3
MADRS, PSD, iAPF, theta Cordance, wPLI (Bailey et al., 2019)	86.60	89.0	84.0
MADRS, eLORETA, Theta Cordance (Jaworska et al., 2019)	87.6	90.3	83.2
Theta frequency band (Erguzel et al., 2015)	89.1	-	94.4
PSD, Alpha asymmetry (Al-Kaysi et al., 2017)	76 (Mood) 92 (Cognition)	-	-
Power Spectral, cordance, asymmetry (Cao et al., 2018b)	81.3	82.1	91.9
Wavelet (Mumtaz et al., 2017)	87.5	95	80
PSD, PSD ratio, Coherence, Mutual Information (Khodayari-Rostamabad et al., 2013)	87.9	80.9	94.9
MADRS, Working Memory accuracy, PSD, wPLI (Bailey et al., 2018)	91.0	92.0	91.0
Non-Linear, power, Bispectrum, cordance (Hasanzadeh et al., 2019)	91.3	91.3	91.3
Time-Frequency domain analysis (Ebrahimzadeh et al., 2021a)	82.43	75.0	86.0
Our proposed method based on components analysis	94.31	92.85	95.65

*Only the baseline EEG data, MADRS, Montgomery and Asberg Depression Rating Scale; PSD, power spectral density; wPLI, weighted phase lag index.

method indicates the high ability to predict the response to treatment based on neuronal activity.

We obtained these results using EEG and machine learning in a standard clinical setting of rTMS therapy for MDD, indicating that these methods are useful tools for studying cortical networks and possibly guiding TMS treatment. We were able to restrict the data's intrinsic complexity and choose electrophysiological variables relevant to therapy-induced alterations and response prediction using the algorithms we used. This method shows potential in constructing treatment regimens, devices, and measurements to enable screening and personalizing treatment viable in the office setting, given the price and accessibility of EEG and the application of data-driven approaches.

This study aimed to evaluate features of extracted components from EEG as potential predictors of MDD treatment responses in patients who received excitatory rTMS to L-DLPFC. The results denote that EEG decomposition components embody different energies in different patients which can be used to separate the responders from non-responders. We show that performing non-linear analyses on the time-series of the EEG components

can lead to more promising results and reliable predictors of rTMS treatment response, in comparison with the classical linear methods. Furthermore, we applied MLP, KNN, and SVM classifiers to different features extracted from pretreatment EEG, including nonlinear, power spectrum, bispectrum, and cordance extracted from ICs, and also a combination of them. Component's power of beta and power of all frequency bands yield classification accuracy of 94.3%, which explains that component power, particularly in beta band, is a valid biomarker of treatment response. Correlation dimension and some of the features based on bispectrum amplitude also have significant predictive abilities. We believe that the appropriate results of the proposed method are suggestive of its potential for clinical applications.

4.1. Limitations and future research directions

The lack of a sham condition is one of the work's limitations. Previous work points to changes in cortical networks during

placebo responses (Cook et al., 2002; Hunter et al., 2006; Benedetti et al., 2011). Furthermore, brain connectivity on functional imaging may predict placebo responses (Tétreault et al., 2016), which emphasizes the importance of future sham-controlled studies. Although the size of our samples is larger than other studies, it is still relatively small compared to the extracted features. Overfitting, in which the algorithm learns tendencies in the data that are not generalizable to new samples, is a common concern with such studies. We tried to address this by employing regularization (reducing model complexity), cross-validation (training and testing the model on different data), and shuffling (demonstrating that the suggested approach does not learn random patterns in the data). The other limitation that this sample size creates is that the variables setting apart the responders from non-responders can only be identified based on their response to rTMS treatment in general. Therefore, replication with a larger sample size is needed to confirm the validity of our findings.

Another concern is that it is possible for the results to have been affected by unpredictable sampling bias. For one thing, most of the participants were not quite sure about the number of their prior depressive episodes. More detailed clinical features such as the current medication dosage were also not recorded. In addition, it is relatively likely that we have dismissed EEG measures with high predictive potential. The sample size also did not allow for a separate analysis of left and right, and bilateral rTMS treatment response in the randomization phase of the study after the initial 3 weeks of left-sided treatment (which was consistent across all participants). Although, to our knowledge, previous research has not yet detected such differences (Fitzgerald et al., 2018), we still believe that deciding if and to what extent left or right-sided rTMS treatment is effective for a patient seems to be a priority for future research work.

Notwithstanding these limitations, we discovered reliable and therapeutically relevant results. To maximize the benefits from this research, future work can be focused on the following:

(1) Independent replication in a larger sample to confirm the predictive ability of the features chosen in this study, or the most predictive features from the literature. (2) Once the predictive ability of our features is proven to be effective in large samples, the focus then should be laid on simplifying and manualizing the calculational procedures to enable a more practical clinical application. An improved machine learning algorithm (perhaps using deep learning for more sophisticated prediction in larger datasets) could then use anonymous data from patients across clinics for better prediction accuracy and generalizability. (3) Research should be conducted to find out whether the outcomes for the patients categorized as non-responders would be worse if they undergo failed treatment as opposed to not undergoing the treatment altogether. If it is established that non-responders show increased depression severity after undergoing unsuccessful treatment, there will be a strong case that the prediction of non-response should result in referrals to an alternative treatment (rather than undergoing rTMS treatment). However, if outcomes remain relatively the same for participants undergoing an unsuccessful treatment, the treatment could still benefit the patient, because even if response prediction outcomes are accurate, there is still a small chance patients will respond in reality (Bailey et al., 2018). (4) Regardless of the answer to point 3, there is still an

obvious need for alternative treatment methods to be developed for non-responders. With regard to brain stimulation, this could entail predicting whether left or right-sided treatment is most suited for particular individuals, and applying rTMS to the medial prefrontal cortex instead of the DLPFC (Dunlop et al., 2017). (5) Investigating the statistical effect of treatment heterogeneity as a paradigm on their classification matrices. Ideally, the type of treatment should be entered as a covariant in the analyses to determine if it has a significant effect on classification.

5. Conclusion

The aim of this study was to evaluate the ability of components extracted from EEG to predict response of rTMS treatment in MDD patients. As expected, the features extracted from the foci of the frontal region had more capability and capacity to differentiate between the two groups (R and NR) than the EEG channels. We have also shown through the use of an advanced machine learning method that nonlinear and frequency features of the components can predict rTMS treatment response. The obtained results show that the proposed method is more capable than other similar methods. Machine learning successfully predicted lack of response to rTMS with high specificity and identified pre- and post-rTMS status using EEG coherence. This approach may provide mechanistic insights and may also become a clinically useful screening tool for rTMS candidates. In future studies, we intend to obtain defined features from other sources decomposed from EEG to improve the classification accuracy. In addition, examining sources identified in other areas of the brain can also increase the yield of studies in the field of rTMS treatment. Further work can also be done on the application of the proposed method to other neurological diseases such as migraine and obsessive-compulsive disorder.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the local ethics committee of the Iran University of Medical Sciences. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

EE, MA, and HS-Z conceived the presented idea. EE developed the theory and performed the computations. Material preparation, data collection, and analysis were performed by EE, FF, MS, and MA. The first draft of the manuscript was written by EE, FF, and MS

and all authors commented on previous versions of the manuscript. LR and HS-Z verified the analytical methods. The visualization and validation were done by EE. All authors provided critical feedback and helped shape the research, analysis, and manuscript. All authors read and approved the final manuscript. HS-Z supervised the project.

Acknowledgments

We thank all the volunteers for their contribution and would like to show our gratitude to Dr. Melika Akbarimehr (Department of Clinical Psychiatry, Neural Clinical Neurology and Neuroscience Centre, Tehran, Iran) and her group for sharing their pearls of wisdom during the course of this research and re-checking the results of psychological tests.

References

- Aboy, M., Hornero, R., Abásolo, D., and Álvarez, D. (2006). Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans. Biomed. Eng.* 53, 2282–2288. doi: 10.1109/TBME.2006.83696
- Acharya, U. R., Sudarshan, V. K., Adeli, H., Santhosh, J., Koh, J. E. W., Puthankatti, S. D., et al. (2015). A novel depression diagnosis index using nonlinear features in EEG signals. *Eur. Neurol.* 74, 79–83. doi: 10.1159/000438457
- Ahdab, R., Ayache, S. S., Brugière, P., Goujon, C., and Lefaucheur, J.-P. (2010). Comparison of “standard” and “navigated” procedures of TMS coil positioning over motor, premotor and prefrontal targets in patients with chronic pain and depression. *Neurophysiol. Clin.* 40, 27–36. doi: 10.1016/j.neucli.2010.01.001
- Ahmadlou, M., Adeli, H., and Adeli, A. (2012). Fractality analysis of frontal brain in major depressive disorder. *Int. J. Psychophysiol.* 85, 206–211. doi: 10.1016/j.ijpsycho.2012.05.001
- Al-Kaysi, A. M., Al-Ani, A., Loo, C. K., Powell, T. Y., Martin, D. M., Breakspear, M., et al. (2017). Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. *J. Affect. Disord.* 208, 597–603. doi: 10.1016/j.jad.2016.10.021
- Amoozegar, S., Pooyan, M., and Roughani, M. (2019). Toward a closed-loop deep brain stimulation in Parkinson's disease using local field potential in parkinsonian rat model. *Med. Hypotheses* 132:109360. doi: 10.1016/j.mehy.2019.109360
- Arns, M., Cerquera, A., Gutiérrez, R. M., Hasselman, F., and Freund, J. A. (2014). Non-linear EEG analyses predict non-response to rTMS treatment in major depressive disorder. *Clin. Neurophysiol.* 125, 1392–1399. doi: 10.1016/j.clinph.2013.11.022
- Arns, M., Drinkenburg, W. H., Fitzgerald, P. B., and Kenemans, J. L. (2012). Neurophysiological predictors of non-response to rTMS in depression. *Brain Stimul.* 5, 569–576. doi: 10.1016/j.brs.2011.12.003
- Bachmann, M., Lass, J., Suhhova, A., and Hinrikus, H. (2013). Spectral asymmetry and Higuchi's fractal dimension measures of depression electroencephalogram. *Comput. Math. Methods Med.* 5:2013. doi: 10.1155/2013/251638
- Bailey, N. W., Hoy, K. E., Rogasch, N. C., Thomson, R. H., McQueen, S., Elliot, D., et al. (2018). Responders to rTMS for depression show increased fronto-midline theta and theta connectivity compared to non-responders. *Brain Stimul.* 11, 190–203. doi: 10.1016/j.brs.2017.10.015
- Bailey, N. W., Hoy, K. E., Rogasch, N. C., Thomson, R. H., McQueen, S., Elliot, D., et al. (2019). Differentiating responders and non-responders to rTMS treatment for depression after one week using resting EEG connectivity measures. *J. Affect. Disord.* 242, 68–79. doi: 10.1016/j.jad.2018.08.058
- Bares, M., Brunovsky, M., Kopecek, M., Stopkova, P., Novak, T., Kozeny, J., et al. (2007). Changes in QEEG prefrontal cordance as a predictor of response to antidepressants in patients with treatment resistant depressive disorder: a pilot study. *J. Psychiatr. Res.* 41, 319–325. doi: 10.1016/j.jpsychires.2006.06.005
- Bares, M., Brunovsky, M., Novak, T., Kopecek, M., Stopkova, P., Sos, P., et al. (2015). QEEG theta cordance in the prediction of treatment outcome to prefrontal repetitive transcranial magnetic stimulation or venlafaxine ER in patients with major depressive disorder. *Clin. EEG Neurosci.* 46, 73–80. doi: 10.1177/1550059413520442
- Baskaran, A. (2016). The comparative effectiveness of EEG biomarkers in antidepressant response and illness prediction in major depressive disorder. Ph.D. Thesis. Queen's University at Kingston, Ontario, Canada.
- Benedetti, F., Carlino, E., and Pollo, A. (2011). How placebos change the patient's brain. *Neuropsychopharmacology* 36, 339–354. doi: 10.1038/npp.2010.81
- Berger, S., Schneider, G., Kochs, E. F., and Jordan, D. (2017). Permutation entropy: too complex a measure for EEG time series. *Entropy* 19:692. doi: 10.3390/e19120692
- Brodman, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig, Germany: Barth.
- Bruder, G. E., Sedoruk, J. P., Stewart, J. W., McGrath, P. J., Quitkin, F. M., and Tenke, C. E. (2008). Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: pre-and post-treatment findings. *Biol. Psychiatry* 63, 1171–1177. doi: 10.1016/j.biopsych.2007.10.009
- Bruder, G. E., Stewart, J. W., Tenke, C. E., McGrath, P. J., Leite, P., Bhattacharya, N., et al. (2001). Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an SSRI antidepressant. *Biol. Psychiatry* 49, 416–425. doi: 10.1016/s0006-3223(00)01016-7
- Bzdok, D., Altman, N., and Krzywinski, M. (2018). Points of significance: statistics versus machine learning. *Nat. Methods* 15, 233–234. doi: 10.1038/nmeth.4642
- Cao, X., Deng, C., Su, X., and Guo, Y. (2018a). Response and remission rates following high-frequency vs. low-frequency repetitive transcranial magnetic stimulation (rTMS) over right DLPFC for treating major depressive disorder (MDD): a meta-analysis of randomized, double-blind trials. *Front. Psychiatry* 9:413. doi: 10.3389/fpsy.2018.00413
- Cao, Z., Lin, C.-T., Ding, W., Chen, M.-H., Li, C.-T., and Su, T.-P. (2018b). Identifying ketamine responses in treatment-resistant depression using a wearable forehead EEG. *IEEE Trans. Biomed. Eng.* 66, 1668–1679. doi: 10.1109/TBME.2018.2877651
- Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., and McCarthy, G. (2017). Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry* 74, 370–378. doi: 10.1001/jamapsychiatry.2017.0025
- Cook, I. A., Leuchter, A. F., Morgan, M., Witte, E., Stubbeman, W. F., Abrams, M., et al. (2002). Early changes in prefrontal activity characterize clinical responders to antidepressants. *Neuropsychopharmacology* 27, 120–131. doi: 10.1016/S0893-133X(02)00294-4
- Cook, I. A., Leuchter, A. F., Witte, E., Abrams, M., Uijtdehaage, S. H. J., Stubbeman, W., et al. (1999). Neurophysiologic predictors of treatment response to fluoxetine in major depression. *Psychiatry Res.* 85, 263–273. doi: 10.1016/s0165-1781(99)00010-4
- Čukić, M. (2020). The reason why rTMS and tDCS are efficient in treatments of depression. *Front. Psychol.* 10:2923. doi: 10.3389/fpsyg.2019.02923

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Čukić, M., Stokić, M., Simić, S., and Pokrajac, D. (2020). The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. *Cogn. Neurodyn.* 14, 443–455. doi: 10.1007/s11571-020-09581-x
- Dozois, D. J. A., Dobson, K. S., and Ahnberg, J. L. (1998). A psychometric evaluation of the beck depression inventory-II. *Psychol. Assess.* 10, 83–89. doi: 10.1037/1040-3590.10.2.83
- Dunlop, K., Sheen, J., Woodside, B., Colton, P., Olmsted, M., Feffer, K., et al. (2017). A randomized comparison of 1 Hz vs. 20 Hz vs. sham dorsomedial prefrontal rTMS for treatment-resistant depression: preliminary clinical results. *Brain Stimul.* 10:464. doi: 10.1016/j.brs.2017.01.359
- Ebrahimzadeh, E., Alavi, S. M., Bijar, A., and Pakkhesal, A. (2013). A novel approach for detection of deception using smoothed pseudo wigner-ville distribution (SPWVD). *J. Biomed. Sci. Eng.* 6, 8–18. doi: 10.4236/jbise.2013.61002
- Ebrahimzadeh, E., Asgarinejad, M., Saliminia, S., Ashoori, S., and Seraji, M. (2021a). Predicting clinical response to transcranial magnetic stimulation in major depression using time-frequency EEG signal processing. *Biomed. Eng. Appl. Basis Commun.* 14:2150048. doi: 10.4015/s1016237221500484
- Ebrahimzadeh, E., Shams, M., Seraji, M., Sadjadi, S. M., Rajabion, L., and Soltanian-Zadeh, H. (2021b). Localizing epileptic foci using simultaneous EEG-fMRI recording: template component cross-correlation. *Front. Neurol.* 12:695997. doi: 10.3389/fneur.2021.695997
- Ebrahimzadeh, E., Shams, M., Rahimpour Jounghani, A., Fayaz, F., Mirbagheri, M., Hakimi, N., et al. (2021c). Localizing confined epileptic foci in patients with an unclear focus or presumed multifocality using a component-based EEG-fMRI method. *Cogn. Neurodyn.* 15, 207–222. doi: 10.1007/s11571-020-09614-5
- Ebrahimzadeh, E., Kalantari, M., Joulani, M., Shahraki, R. S., Fayaz, F., and Ahmadi, F. (2018a). Prediction of paroxysmal Atrial Fibrillation: a machine learning based approach using combined feature vector and mixture of expert classification on HRV signal. *Comput. Methods Programs Biomed.* 165, 53–67. doi: 10.1016/j.cmpb.2018.07.014
- Ebrahimzadeh, E., Manuchehri, M. S. M. S., Amoozegar, S., Araabi, B. N. B. N., and Soltanian-Zadeh, H. (2018b). A time local subset feature selection for prediction of sudden cardiac death from ECG signal. *Med. Biol. Eng. Comput.* 56, 1253–1270. doi: 10.1007/s11517-017-1764-1
- Ebrahimzadeh, E., Fayaz, F., Nikravan, M., Ahmadi, F., and Dolatabad, M. R. (2018c). Towards an automatic diagnosis system for lumbar disc herniation: the significance of local subset feature selection. *Biomed. Eng. Appl. Basis Commun.* 30:1850044. doi: 10.4015/s1016237218500448
- Ebrahimzadeh, E., Soltanian-Zadeh, H., Araabi, B. N., Fesharaki, S. S. H., and Habibabadi, J. M. (2019a). Component-related BOLD response to localize epileptic focus using simultaneous EEG-fMRI recordings at 3T. *J. Neurosci. Methods* 322, 34–49. doi: 10.1016/j.jneumeth.2019.04.010
- Ebrahimzadeh, E., Soltanian-Zadeh, H., Araabi, B. N., Fesharaki, S. S. H., Habibabadi, J. M., Nadjar Araabi, B., et al. (2019b). Localizing epileptic focus through simultaneous EEG-fMRI recording and automated detection of IED from inside-scanner EEG. *Iran. J. Biomed. Eng.* 13, 135–145. doi: 10.3389/fneur.2021.695997
- Ebrahimzadeh, E., Shams, M., Fayaz, F., Rajabion, L., Mirbagheri, M., Nadjar Araabi, B., et al. (2019c). Quantitative determination of concordance in localizing epileptic focus by component-based EEG-fMRI. *Comput. Methods Programs Biomed.* 177, 231–241. doi: 10.1016/j.cmpb.2019.06.003
- Ebrahimzadeh, E., Foroutan, A., Shams, M., Baradaran, R., Rajabion, L., Joulani, M., et al. (2019d). An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal. *Comput. Methods Programs Biomed.* 169, 19–36. doi: 10.1016/j.cmpb.2018.12.001
- Ebrahimzadeh, E., and Pooyan, M. (2013). Prediction of sudden cardiac death (SCD) using time-frequency analysis of ECG signals. *Comput. Intell. Electr. Eng.* 3, 15–26.
- Ebrahimzadeh, E., Pooyan, M., and Bijar, A. (2014). A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. *PLoS One* 9:e81896. doi: 10.1371/journal.pone.0081896
- Ebrahimzadeh, E., Pooyan, M., and Elias Ebrahimzadeh, M. P. (2011). Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals. *J. Biomed. Sci. Eng.* 04, 699–706. doi: 10.4236/jbise.2011.411087
- Erguzel, T. T., Ozekes, S., Tan, O., and Gultekin, S. (2015). Feature selection and classification of electroencephalographic signals: an artificial neural network and genetic algorithm based approach. *Clin. EEG Neurosci.* 46, 321–326. doi: 10.1177/1550059414523764
- Fischl, B., and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U S A* 97, 11050–11055. doi: 10.1073/pnas.200033797
- Fitzgerald, P. B., Hoy, K. E., Elliot, D., McQueen, S., Wambeek, L. E., and Daskalakis, Z. J. (2018). Exploring alternative rTMS strategies in non-responders to standard high frequency left-sided treatment: a switching study. *J. Affect. Disord.* 232, 79–82. doi: 10.1016/j.jad.2018.02.016
- Fitzgerald, P. B., Maller, J. J., Hoy, K. E., Thomson, R., and Daskalakis, Z. J. (2009). Exploring the optimal site for the localization of dorsolateral prefrontal cortex in brain stimulation experiments. *Brain Stimul.* 2, 234–237. doi: 10.1016/j.brs.2009.03.002
- Grova, C., Daunizeau, J., Kobayashi, E., Bagshaw, A. P., Lina, J. M., Dubeau, F., et al. (2008). Concordance between distributed EEG source localization and simultaneous EEG-fMRI studies of epileptic spikes. *Neuroimage* 39, 755–774. doi: 10.1016/j.neuroimage.2007.08.020
- Hasanzadeh, F., Mohebbi, M., and Rostami, R. (2019). Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal. *J. Affect. Disord.* 256, 132–142. doi: 10.1016/j.jad.2019.05.070
- Heller, W., Etienne, M. A., and Miller, G. A. (1995). Patterns of perceptual asymmetry in depression and anxiety: implications for neuropsychological models of emotion and psychopathology. *J. Abnorm. Psychol.* 104, 327–333. doi: 10.1037//0021-843x.104.2.327
- Hunter, A. M., Leuchter, A. F., Morgan, M. L., and Cook, I. A. (2006). Changes in brain function (quantitative EEG cordance) during placebo lead-in and treatment outcomes in clinical trials for major depression. *Am. J. Psychiatry* 163, 1426–1432. doi: 10.1176/ajp.2006.163.8.1426
- Iosifescu, D. V., Greenwald, S., Devlin, P., Mischoulon, D., Denninger, J. W., Alpert, J. E., et al. (2009). Frontal EEG predictors of treatment outcome in major depressive disorder. *Eur. Neuropsychopharmacol.* 19, 772–777. doi: 10.1016/j.euroneuro.2009.06.001
- Jaworska, N., de la Salle, S., Ibrahim, M.-H., Blier, P., and Knott, V. (2019). Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data. *Front. Psychiatry* 9:768. doi: 10.3389/fpsy.2018.00768
- Khodayari-Rostamabad, A., Reilly, J. P., Hasey, G. M., de Bruin, H., and MacCrimmon, D. (2011). Using pre-treatment electroencephalography data to predict response to transcranial magnetic stimulation therapy for major depression. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2011, 6418–6421. doi: 10.1109/IEMBS.2011.6091584
- Khodayari-Rostamabad, A., Reilly, J. P., Hasey, G. M., de Bruin, H., and MacCrimmon, D. J. (2013). A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin. Neurophysiol.* 124, 1975–1985. doi: 10.1016/j.clinph.2013.04.010
- Kito, S., Hasegawa, T., and Koga, Y. (2012). Cerebral blood flow ratio of the dorsolateral prefrontal cortex to the ventromedial prefrontal cortex as a potential predictor of treatment response to transcranial magnetic stimulation in depression. *Brain Stimul.* 5, 547–553. doi: 10.1016/j.brs.2011.09.004
- Knott, V., Mahoney, C., Kennedy, S., and Evans, K. (2000). Pre-treatment EEG and it's relationship to depression severity and paroxetine treatment outcome. *Pharmacopsychiatry* 33, 201–205. doi: 10.1055/s-2000-8356
- Knott, V. J., Telner, J. I., Lapierre, Y. D., Browne, M., and Horn, E. R. (1996). Quantitative EEG in the prediction of antidepressant response to imipramine. *J. Affect. Disord.* 39, 175–184. doi: 10.1016/0165-0327(96)00003-1
- Lebiecka, K., Zuchowicz, U., Wozniak-Kwasniewska, A., Szekeley, D., Olejarczyk, E., and David, O. (2018). Complexity analysis of EEG data in persons with depression subjected to transcranial magnetic stimulation. *Front. Physiol.* 9:1385. doi: 10.3389/fphys.2018.01385
- Leentjens, A. F. G., Verhey, F. R. J., Lousberg, R., Spitsbergen, H., and Wilmink, F. W. (2000). The validity of the hamilton and montgomery-Åsberg depression rating scales as screening and diagnostic tools for depression in Parkinson's disease. *Int. J. Geriatr. Psychiatry* 15, 644–649. doi: 10.1002/1099-1166(200007)15:7<644::aid-gps167<3.0.co;2-1
- Leichsenring, F., Steinert, C., Rabung, S., and Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry* 21, 133–145. doi: 10.1002/wps.20941
- Leuchter, A. F., Cook, I. A., Lufkin, R. B., Dunkin, J., Newton, T. F., Cummings, J. L., et al. (1994). Cordance: a new method for assessment of cerebral perfusion and metabolism using quantitative electroencephalography. *Neuroimage* 1, 208–219. doi: 10.1006/nimg.1994.1006
- Li, Y., Tong, S., Liu, D., Gai, Y., Wang, X., Wang, J., et al. (2008). Abnormal EEG complexity in patients with schizophrenia and depression. *Clin. Neurophysiol.* 119, 1232–1241. doi: 10.1016/j.clinph.2008.01.104
- Lieber, A. L., and Prichep, L. S. (1988). Diagnosis and subtyping of depressive disorders by quantitative electroencephalography: I. discriminant analysis of selected variables in untreated depressives. *Hillsdale J. Clin. Psychiatry* 10, 71–83.
- Metin, S. Z., Balli Altuglu, T., Metin, B., Erguzel, T. T., Yigit, S., Arkan, M. K., et al. (2020). Use of EEG for predicting treatment response to transcranial magnetic stimulation in obsessive compulsive disorder. *Clin. EEG Neurosci.* 51, 139–145. doi: 10.1177/1550059419879569
- Mumtaz, W., Xia, L., Mohd Yasin, M. A., Azhar Ali, S. S., and Malik, A. S. (2017). A wavelet-based technique to predict treatment outcome for major depressive disorder. *PLoS One* 12:e0171409. doi: 10.1371/journal.pone.0171409

- Nikravan, M., Ebrahimzadeh, E., Izadi, M. R., and Mikaeili, M. (2016). Toward a computer aided diagnosis system for lumbar disc herniation disease based on MR images analysis. *Biomed. Eng. Appl. Basis Commun.* 28:1650042. doi: 10.4015/S1016237216500423
- O'Reardon, J. P., Solvason, H. B., Janicak, P. G., Sampson, S., Isenberg, K. E., Nahas, Z., et al. (2007). Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. *Biol. Psychiatry* 62, 1208–1216. doi: 10.1016/j.biopsych.2007.01.018
- Olejarczyk, E., Zuchowicz, U., Wozniak-Kwasniewska, A., Kaminski, M., Szekely, D., and David, O. (2020). The impact of repetitive transcranial magnetic stimulation on functional connectivity in major depressive disorder and bipolar disorder evaluated by directed transfer function and indices based on graph theory. *Int. J. Neural Syst.* 30:2050015. doi: 10.1142/S012906572050015X
- Patel, M. J., Andreescu, C., Price, J. C., Edelman, K. L., Reynolds III, C. F., and Aizenstein, H. J. (2015). Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. *Int. J. Geriatr. Psychiatry* 30, 1056–1067. doi: 10.1002/gps.4262
- Paxinos, G., and Mai, J. K. (2015). *The Human Nervous System*. San Diego, CA: Elsevier Academic Press.
- Peleman, K., Van Schuerbeek, P., Luypaert, R., Stadnik, T., De Raedt, R., De Mey, J., et al. (2010). Using 3D-MRI to localize the dorsolateral prefrontal cortex in TMS research. *World J. Biol. Psychiatry* 11, 425–430. doi: 10.1080/15622970802669564
- Pellicciari, M. C., Cordone, S., Marzano, C., Bignotti, S., Gazzoli, A., Miniussi, C., et al. (2013). Dorsolateral prefrontal transcranial magnetic stimulation in patients with major depression locally affects alpha power of REM sleep. *Front. Hum. Neurosci.* 7:433. doi: 10.3389/fnhum.2013.00433
- Raeisi, K., Mohebbi, M., Khazaei, M., Seraji, M., and Yoonessi, A. (2020). Phase-synchrony evaluation of EEG signals for multiple sclerosis diagnosis based on bivariate empirical mode decomposition during a visual task. *Comput. Biol. Med.* 117:103596. doi: 10.1016/j.compbiomed.2019.103596
- Redlich, R., Opel, N., Grotegerd, D., Dohm, K., Zaremba, D., Bürger, C., et al. (2016). Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry* 73, 557–564. doi: 10.1001/jamapsychiatry.2016.0316
- Sadjadi, S. M., Ebrahimzadeh, E., Shams, M., Seraji, M., and Soltanian-Zadeh, H. (2021). Localization of epileptic foci based on simultaneous EEG-fMRI data. *Front. Neurol.* 12:645594. doi: 10.3389/fneur.2021.645594
- Segal, D. L. (2010). "Diagnostic and statistical manual of mental disorders (DSM-IV-TR)," in *Corsini Encycl. Psychol.*, 1–3. doi: 10.1002/9780470479216.corps.y0271
- Seraji, M. (2021). *Transcranial Alternating Current Stimulation Affects Resting-state Functional Connectivity*. M.Sc. Thesis. New Jersey, USA: Newark Rutgers, The State University of New Jersey.
- Seraji, M., Mohebbi, M., Safari, A., and Krekelberg, B. (2021). Multiple sclerosis reduces synchrony of the magnocellular pathway. *PLoS One* 16:e0255324. doi: 10.1371/journal.pone.0255324
- Sharp, R. (2015). The Hamilton rating scale for depression. *Occup. Med. (Chic. Ill)* 65, 340. doi: 10.1093/occmed/kqv043
- Spasic, S., Kalauzi, A., Grbic, G., Martac, L., and Culic, M. (2005). Fractal analysis of rat brain activity after injury. *Med. Biol. Eng. Comput.* 43, 345–348. doi: 10.1007/BF02345811
- Spronk, D., Arns, M., Barnett, K. J., Cooper, N. J., and Gordon, E. (2011). An investigation of EEG, genetic and cognitive markers of treatment response to antidepressant medication in patients with major depressive disorder: a pilot study. *J. Affect. Disord.* 128, 41–48. doi: 10.1016/j.jad.2010.06.021
- Suffin, S. C., and Emory, W. H. (1995). Neurometric subgroups in attentional and affective disorders and their association with pharmacotherapeutic outcome. *Clin. Electroencephalogr.* 26, 76–83. doi: 10.1177/155005949502600204
- Tenke, C. E., Kayser, J., Manna, C. G., Fekri, S., Kroppmann, C. J., Schaller, J. D., et al. (2011). Current source density measures of electroencephalographic alpha predict antidepressant treatment response. *Biol. Psychiatry* 70, 388–394. doi: 10.1016/j.biopsych.2011.02.016
- Tétreault, P., Mansour, A., Vachon-Presseau, E., Schnitzer, T. J., Apkarian, A. V., and Baliki, M. N. (2016). Brain connectivity predicts placebo response across chronic pain clinical trials. *PLoS Biol.* 14:e1002570. doi: 10.1371/journal.pbio.1002570
- Turner, E. H., Cipriani, A., Furukawa, T. A., Salanti, G., and de Vries, Y. A. (2022). Selective publication of antidepressant trials and its influence on apparent efficacy: Updated comparisons and meta-analyses of newer versus older trials. *PLoS Med.* 19:e1003886. doi: 10.1371/journal.pmed.1003886
- Ulrich, G., Renfordt, E., Zeller, G., and Frick, K. (1984). Interrelation between changes in the EEG and psychopathology under pharmacotherapy for endogenous depression. *Pharmacopsychiatry* 17, 178–183. doi: 10.1055/s-2007-1017433
- von Economo, C. F., Koskinas, G. N., and Triarhou, L. C. (2008). *Atlas of Cytoarchitectonics of the Adult Human Cerebral Cortex*. New York, NY: Karger.
- Wade, B. S. C., Joshi, S. H., Njau, S., Leaver, A. M., Vasavada, M., Woods, R. P., et al. (2016). Effect of electroconvulsive therapy on striatal morphometry in major depressive disorder. *Neuropsychopharmacology* 41, 2481–2491. doi: 10.1038/npp.2016.48
- Woźniak-Kwaśniewska, A., Szekely, D., Harquel, S., Bougerol, T., and David, O. (2015). Resting electroencephalographic correlates of the clinical response to repetitive transcranial magnetic stimulation: a preliminary comparison between unipolar and bipolar depression. *J. Affect. Disord.* 183, 15–21. doi: 10.1055/a-2034-1911
- Zhdanov, A., Atluri, S., Wong, W., Vaghei, Y., Daskalakis, Z. J., Blumberger, D. M., et al. (2020). Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Netw. Open* 3:e1918377. doi: 10.1001/jamanetworkopen.2019.18377