TOP-EEG: A Robust Software to Predict the Outcomes of Therapies for Depression Using EEG Signals in DGMD Domain

Hesam Akbari[1], Wael Korani[1(✉)], Junhua Ding[1], Reza Rostami[2], and Reza Kazemi[3]

[1] Department of Information Science, University of North Texas, Texas, USA
{Hesam.Akbari,Wael.Korani,junhua.ding}@unt.edu
[2] Department of Psychiatry, University of Tehran, Tehran, Iran
rrostami@ut.ac.ir
[3] Faculty of Entrepreneurship, University of Tehran, Tehran, Iran
rezakazemi@ut.ac.ir

Abstract. Self-mutilation and suicide are negative consequences of depression disorder, if it is left untreated. Two commonly prescribed ways of treatments for depression are Selective serotonin reuptake inhibitors (SSRIs) and repetitive transcranial magnetic stimulation (rTMS) therapies. Although the effectiveness of these two therapies have been approved by food and drug administration (FDA), the successful rate to these therapies is around 50%. In general, a psychiatric prescribes one of the therapies based on his/her experience and waits for a period of time to check the improvement. However, if the therapy fails to reduce depression levels, the risk of self-harm or suicide may increase. This paper proposes a robust software called Therapy Outcomes Predictor by Electroencephalogram (TOP-EEG) to predict the improvement of two different therapies for depressed patients based on pretreatment electroencephalogram (EEG) signals. The TOP-EEG software utilizes a novel recently proposed signal processing technique called dynamic graph mode decomposition (DGMD) to automatically decompose the EEG signals into the intrinsic mode functions (IMFs). Then, five entropy-based features quantize the value of the complexity and randomness of the IMFs. Statistically significant features are selected using the Kruskal-Wallis test and fed into traditional machine learning algorithms and artificial neural network architectures for classification. A 10-fold cross-validation strategy is applied during training and testing to minimize bias in the results. The cascade-forward neural network (CFNN) architecture shows the best performance among the other classification algorithms. The TOP-EEG software is evaluated using two different databases: SSRI and rTMS therapies, including data from 30 and 15 depressed patients, respectively. The results show that the TOP-EEG software achieves classification accuracy levels of 93.16% and 94.59% to predict the outcomes of the SSRI and rTMS therapies, respectively. This is the first time in the literature a software is developed to predict the outcomes of two therapies for depression and recommend the best course of treatment. The software is reliable and can be used in clinics and hospitals to assist neurologists and psychiatrists in prescribing the most effective treatment for depression.

M. Mahmud et al. (Eds.): ICONIP 2024, CCIS 2296, pp. 403–419, 2025.
https://doi.org/10.1007/978-981-96-7033-8_28

Keywords: TOP-EEG · SSRI · rTMS · depression · EEG

1 Introduction

Depression might lead to self-harm and suicide if it is left untreated. The World Health Organization (WHO) reported that around 700,000 people die by suicide every year [15]. There are about twenty suicide attempts for each death, which indicates the importance of this problem [12]. The first recommenced treatments for depression are selective serotonin reuptake inhibitors (SSRI) therapy [16] and repetitive transcranial magnetic stimulation (rTMS) therapy [17]. In SSRI therapy, specific medications are prescribed for depressed patients for a period of four to six weeks. On the other hand, in rTMS therapy, an electromagnetic coil is placed near the scalp to generate magnetic pulses. It produces electrical currents that go though the nervous cells. This process is repeagted five times a week for four to six weeks. Predicting the effectiveness of therapies is a challenge for neurologists and psychiatrists. It is because of the complexity of the human brain. The response rate to SSRI therapy is 49% [13], and the response rate for rTMS is between 40% and 60% [14]. If the therapy fails, the risk of suicide increases for depressed patients.

Electroencephalogram (EEG) signals are a neuroimaging technique that has become popular in the last two decades. The brain activities are recorded using EEG electrodes on the scalp. EEG signal recording is an inexpensive and real-time technique. The EEG was used for detecting mental disorders like depression [6]. It motivates us to develop a new computer-aided decision (CAD) system based on the EEG signals to determine the best course treatment for depression disorder. In TOP-EEG software, pre-treatment EEG signals are classified as either responders (R) or non-responders (NR) to each therapy. There are two therapies, including SSRI and rTMS, the TOP-EEG software performs two binary classification tasks. In this way, the TOP-EEG software predicts the outcomes of SSRI and rTMS therapies based on the pre-treatment EEG signals.

The EEG signals are recorded from the scalp, and different channels of the EEG monitor the activities of a human brain. The EEG signal is a non-linear and non-stationary signal with complex behavior [5]. Decomposing signals into sub-bands is a common technique for analyzing nonlinear signals such as EEG. Several techniques have been developed to decompose nonlinear signals in the literature. The short-time Fourier transform (STFT) is a known signal processing technique. The EEG signals are decomposed into sub-bands by STFT in a straightforward way. However, the STFT cannot decompose signals that do not have sinusoidal and cosine waves, such as impulse functions. To address this limitation, the discrete wavelet transform (DWT) was proposed [4]. In the DWT, various mother wavelets can be used to decompose the input signals. For instance, when the signal has an impales function wave, the Har mother wavelet is a proper mother wavelet. The input signal is decomposed by using high-pass and low-pass filters at each decomposition level of DWT. The signal spectrum is reduced in half at each stage of DWT decomposition. Although DWT has shown

acceptable performance in EEG signal processing applications, it is not adaptive to the input signal. It means that the filter band of DWT is fixed for any input signal.

The empirical mode decomposition (EMD) was proposed to solve the inadaptability nature in the DWT method. The EMD decomposes the input signal into intrinsic mode functions (IMFs) based on the signal's local maxima and minima. Although EMD is an adaptive method that is significant for applications like EEG signal analysis, it is not built on a mathematical proof. Later, variational mode decomposition (VMD) was introduced to overcome the previous limitation [7]. The VMD method provides a solid mathematical proof, which makes it more reliable. However, the main limitation of VMD is that it can not decompose multi-dimensional data such as EEG. It means that it is significant just for one-channel decomposition, not multi-channel decomposition.

The multivariate VMD (MVMD) [11] is a variant of VMD designed for multi-dimensional signals like EEG. MVMD was applied to EEG signals for various applications. However, MVMD has some limitations, such as assuming static connectivity structures and being unable to track dynamic interactions in time-varying graph signals. In addition, several parameters are required to be adapted by the user, such as the number of IMFs. Choosing the number of IMFs is important and also challenging, because it affects the accuracy of decomposition. Furthermore, it is complex and computationally expensive, which makes it a non-significant option for analyzing large-scale real-time series. Also, the MVMD is restricted to a narrow frequency range that fails to capture multiscale dynamics in the signal. It suffers from mode mixing and duplicate mode problems, which cause inaccurate results. The MVMD cannot provide any information about dynamic connectivity structures associated with signal components (i.e., channels in EEG signals).

In 2024, the dynamic graph mode decomposition (DGMD) was developed to solve the weakness of MVMD [10]. The DGMD is a powerful signal processing technique for multi-dimensional signals such as EEG. The DGMD method solves the static connectivity structure by inferring dynamic functional connectivity structures associated with each oscillatory component. It allows DGMD to conduct a comprehensive analysis of network dynamics. The DGMD uses a successive scheme, which is not required for defining parameters by the user. So, it improves the data-driven capability and performance of DGMD compared to MVMD. In DGMD, optimization formulations and a successive decomposition scheme reduce computational resources while simultaneously preserving significant performance and accuracy of decomposition.

The signal is decomposed into narrow-band oscillatory components by DGMD, which gives a detailed analysis of signal variations along with different frequency scales. DGMD uses advanced optimization techniques, a joint optimization formulation, and simultaneously estimating signal modes and their connectivity structures. Thus, the mode mixing problem is solved in DGMD. In addition, it infers dynamic connectivity matrices for each oscillatory component, which leads to a better understanding of the dynamic connections across

networks in the time-varying graph signal. DGMD provides some attributes such as data-driven decomposition, multiscale analysis, dynamic connectivity inference, and simultaneous mode and connectivity estimation, which promote it as a powerful signal processing technique for analyzing complex time-varying graph signals like EEG data. In the current work, the DGMD technique is used to analyze the nonlinear dynamics of multi-EEG signals.

Feature extraction is one of the most challenging parts of any classification task. The features need to be discriminative enough to decode the hidden information within the data. Computing entropy as a system parameter has been used as a feature to describe brain behavior using EEG signals. In the current work, five different types of entropy are used to extract information from the EEG signals. Feature selection is used to reduce computational complexity while maintaining high system performance. In the current work, a statistical-based method named Kruskal-Wallis (KW) test is applied to the extracted features to select statistically significant features. The KW test was applied to EEG signal processing applications before, such as the detection of seizure [1], alcoholism [3] and schizophrenia [2].

Classification is the final step in computer-aided diagnosis systems. It is crucial because it enables the system to automatically and accurately identify and categorize features into a proper class. Choosing the best classification technique ensures that the system is reliable and efficient. A robust classification technique generalizes the structure of data by decoding the connections and relationships between different features. Deep learning-based methods have shown significant performance in EEG signal processing applications. The architecture of deep learning algorithms directly impacts their performance and complexity. Cascade-forward neural network (CFNN) architecture is a variant of the feed-forward neural network (FFNN) architecture. In the paper, we implented different classifier to choose the best classifier for dual tasks software.

The rest of the paper is organized as follows: Sect. 2 covers the materials, describing the databases that are used in the current work. Section 3 presents the proposed TOP-EEG software, detailing the steps including DGMD, entropy feature extraction, and classification. Results and discussions are reported in Sect. 4. Finally, the paper concludes in Sect. 5.

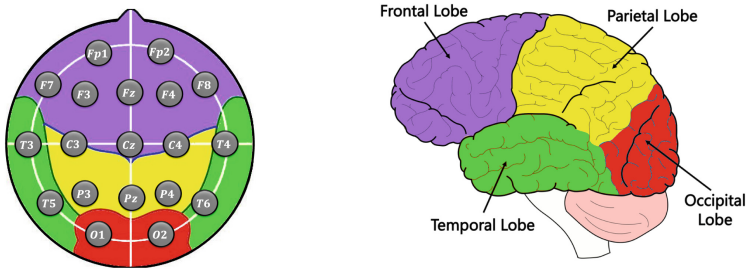
2 Materials

Two different databases are used to evaluate the proposed TOP-EEG software. The database for SSRI therapy is Mumtaz database, which is a public database, and the database for rTMS therapy is Atieh Hospital database, which is a private database.

2.1 Mumtaz Database

The Mumtaz database is a public database that includes the EEG signals of 30 patients diagnosed with major depressive disorder. These patients were diagnosed and prescribed SSRI therapy [8]. The Beck Depression Inventory (BDI)

score was used to identify the level of depression in patients. The EEG signals were recorded for depressed patients before starting the SSRI therapy. The EEG signals were recorded using the 10–20 international system using 19 channels. The locations of EEG channels on the scalp and the brain lobes are shown in Fig. 1. The EEG recordings were sampled at a frequency of 256 Hz. The recommenced medications in Mumtaz experiments were Escitalopram (10–20 mg/day), Fluvoxamine (100–300 mg/day), Sertraline (50–200 mg/day), and Fluoxetine (20–60 mg/day). Patients were on the SSRI therapy for four weeks. After four weeks, the BDI score was evaluated again. The EEG signals of patients who showed more than a 50% improvement based on their BDI scores were labeled as responders (R) (12 patients); otherwise, the patients were labeled as non-responders (NR) (18 patients) to the SSRI therapy. Figure 2 shows a sample of a 10-second EEG signal, which is equal to 2560 samples.



(a) 10-20 standard (b) brain lobes

Fig. 1. Visualization of the location of electrodes on the skin in 10–20 standard and brain lobes.

2.2 Atieh Hospital Database

Atieh hospital database is used to evaluate the performance of the proposed TOP-EEG software in predicting the outcomes of rTMS therapy. This database was recorded at Atieh Hospital, Tehran, Iran. This database has the EEG signals from 15 depressed patients diagnosed. Dr. Reza Rostami and Dr. Reza Kazemi diagnosed the patients with depression disorder and prescribed rTMS therapy as a treatment. The EEG signals were recorded for these patients before starting the rTMS therapy. The EEG signals were recorded using the international 10–20 standard system. The sampling frequency of recorded EEG signals was 512 Hz. The BDI scores were evaluated for the patients. The rTMS therapy was repeated five times a week for four weeks. The BDI scores were evaluated again for all patients after finishing the rTMS therapy. The pre-treatment EEG signals of patients who showed more than 50% improvement based in BDI scores after the rTMS therapy were labeled as R to the rTMS therapy. On the other hand, the pre-treatment EEG signals of patients who did not show this level of improvement were labeled as NR to rTMS therapy. In the Atieh Hospital database, there

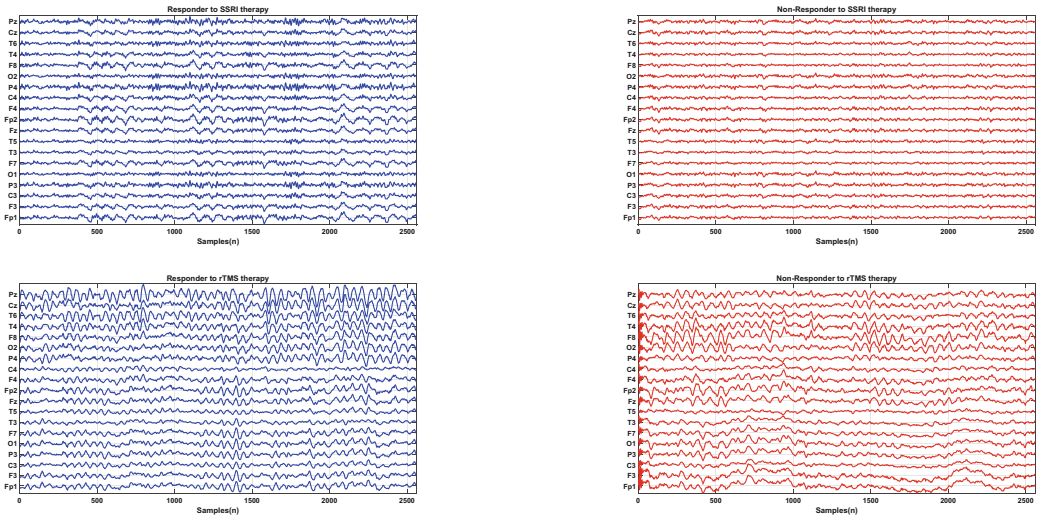


Fig. 2. a sample of R (left) and NR (right) EEG signals to SSRI (up) and rTMS (down), respectively.

are pre-treatment EEG signals of nine R and six NR to rTMS therapy. Figure 2 shows the R and NR EEG signals to rTMS therapy. In the current work, the EEG signals in the rTMS database are divided into five seconds, which is equal to 2560 samples.

3 Proposed TOP-EEG Software

In this section, the steps of the proposed TOP-EEG software are discussed. First, the DGMD method decomposes the EEG signals into IMFs. Then the mean value of five entropy-based features is extracted for IMFs. Fainlly, the KW test selects the statistically significant features, which are fed into a CFNN architecture for classification. Figure 3 shows the steps of the proposed TOP-EEG software.

3.1 Dynamic Graph Mode Decomposition

The dynamic graph mode decomposition (DGMD) technique is a time-frequency method used to decompose a time-varying graph signal into intrinsic mode functions (IMFs) and their associated dynamic connectivity structures [10]. Assume a given dynamic graph signal $\tilde{\mathbf{s}}(t)$ that comprises a set of time-series signals: $\tilde{\mathbf{s}}(t) = [\tilde{s}_1(t), \tilde{s}_2(t), \ldots, \tilde{s}_N(t)]$ where $\tilde{s}_n(t)$ represents the time-series corresponding to the n^{th} vertex. Similarly, $\tilde{\mathbf{s}}(t)$ can be represented by a matrix $\mathbf{S} \in \mathbb{R}^{N \times T}$ where T denotes the number of time samples and N refers to the number of vertices (channels) of a time-varying graph signal. Then, the DGMD technique is defined as the following optimization problem:

$$\underset{\{\bar{g}_n\},\{\bar{r}_n\},\{\omega_g\},\left\{\mathbf{W}^{(t_m)}\right\}}{\text{minimize}} \left(\mathcal{J}_1 + \mathcal{J}_2\right) \tag{1}$$

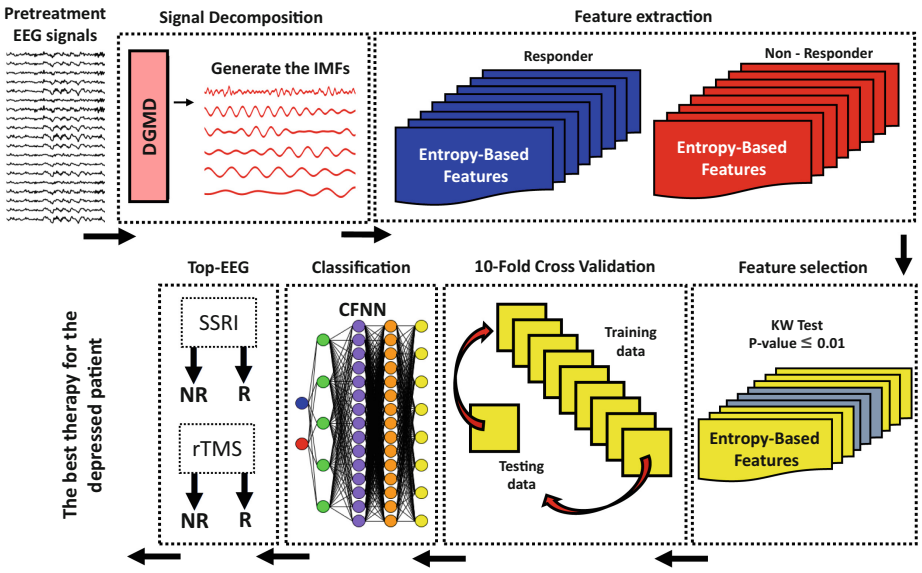


Fig. 3. Step by step of the proposed TOP-EEG software.

$$\textbf{subject to} \begin{cases} \tilde{g}_n(t) + \tilde{r}_n(t) = \tilde{s}_n(t), \quad \forall n, \\ \mathbf{W}^{(t_m)} \geq \mathbf{0}, \quad \mathbf{W}^{(t_m)} = \mathbf{W}^{(t_m)^T}, \\ \text{diag}\left(\mathbf{W}^{(t_m)}\right) = 0, \quad \forall m, \end{cases} \tag{2}$$

where $\tilde{g}_n$ is the graph mode of interest, ω_g represents the center frequency of each graph mode, $\tilde{r}_n$ denotes the residual signal, and $\mathbf{W}$ denotes the graph structure. The subscript g in $\mathbf{W}_g$ indicates that $\mathbf{W}$ is derived using a graph-based method.

$$\mathcal{J}_1 = \left\{ \alpha \sum_{n=1}^{N} \left\| \partial_t [\tilde{g}_n(t) e^{-j\omega_g t} \right] \right\|_2^2 + \sum_{n=1}^{N} \| h_g(t) * \tilde{r}_n(t) \|_2^2 \} \tag{3}$$

where α controls its significance, N denotes the number of channels, $\left\| \partial_t \left[\tilde{g}_n(t) e^{-j\omega_g t} \right] \right\|_2^2$ signifies the derivative of the estimated oscillatory component with respect to time, and $\| h_g(t) * \tilde{r}_n(t) \|_2^2$ represents the complex conjugate of the estimated residual component multiplied by the estimated oscillatory component. The $h_g(t)$ is a filter with a frequency response of $h_g(\omega) = 1/\alpha \left(\omega - \omega_g\right)^2$, where $h_g(\omega) \to \infty$ if $\omega \to \omega_g$.

$$\begin{aligned} \mathcal{J}_2 = \{ & \beta \sum_{m=1}^{M} \| \mathbf{W}_g^{(t_m)} \circ \mathbf{Z}^{(t_m)} \|_{1,1} + \frac{\gamma}{2} \sum_{m=1}^{M} \left\| \mathbf{W}_g^{(t_m)} \right\|_F^2 \\ & - \sum_{m=1}^{M} \mathbf{1}^T \log \left(\mathbf{W}_g^{(t_m)} \mathbf{1} \right) + \mu \sum_{m=2}^{M} f^{(t_m)} (\mathbf{W}_g^{(t_m)}, \mathbf{W}_g^{(t_{m-1})}) \} \end{aligned} \tag{4}$$

where β, γ, and μ control the importance of different terms. M represents the number of time points, $W(t_m)$ denotes the dynamic connectivity matrix at t_m,

$\|\mathbf{W}_g^{(t_m)} \circ \mathbf{Z}^{(t_m)}\|_{1,1}$ represents the element-wise product between the estimated oscillatory component and the dynamic connectivity matrix, $\mathbf{1}^T \log(\mathbf{W}_g^{(t_m)}\mathbf{1})$ signifies the logarithm of the product of the dynamic connectivity matrix and the estimated IMF.

$f^{(t_m)}(\mathbf{W}_g^{(t_m)}, \mathbf{W}_g^{(t_{m-1})})$ represents a function dependent on the dynamic connectivity matrices at consecutive time points. These elements collectively contribute to extracting meaningful oscillatory components and their associated dynamic connectivity structures from time-varying graph signals through the optimization process. The $f^{(t_m)}$ is a function of time that represents a measure of dissimilarity between the adjacency matrix at t_m and the adjacency matrix at t_{m-1}. Different priors for $f^{(t_m)}$ can be chosen based on specific application settings. The $f^{(t_m)}$ is chosen to be a Tikhonov smoothness prior, defined as $\|\mathbf{W}_g^{(t_m)} - \mathbf{W}_g^{(t_{m-1})}\|_F^2$, which promotes smooth changes in graph edges over time. The subscript g in $\mathbf{W}_g$ indicates that $\mathbf{W}$ is derived using a graph-based method. The vector of ones is represented by $\mathbf{1}$. The symbol $\circ$ is used for the Hadamard product, the matrix or vector transposition is shown by $(.)^T$, and $\|\cdot\|_2$ is the L_2-norm of a vector. Additionally, $\mathbf{Z} \in \mathbb{R}_+^{N\times N}$ defines a pairwise distance matrix where $\mathbf{Z}_{t_m,t_{m+1}} = \|\tilde{\mathbf{s}}_{t_m} - \tilde{\mathbf{s}}_{t_{m+1}}\|^2$.

The restrictions ensure that the IMFs and dynamic connectivity structures are orthogonal. By solving this optimization problem, DGMD decomposes the time-varying graph signals into IMFs and their associated dynamic connectivity networks, providing a significant map for the nonlinear dynamics of the signal. In the current study, the multiscale dynamic graph signal analysis Matlab code [9] is used to implement the DGMD. The difficult parameters of the toolbox are used to decompose the EEG signals into IMFs.

3.2 Feature Extraction

Entropy is a measure of the disorder in a complex system. In signal processing, it is defined as the measurement of the randomness in the time series. The entropy is used as a feature to decode the nonlinear behavior of EEG signals, and it has been used for several applications, such as the detection of alcoholism and depression [3]. In the current work, entropy features are computed to decode the complex behaviour of the EEG signals in the R and NR groups. Five entropy features are computed for the IMFs of EEG signals: shannon entropy (Sh), log Energy entropy (Le), threshold entropy (Th), Stein's unbiased risk estimate entropy (Su), and norm entropy (Nr). The summtion of one IMF of EEG signals is representd as $x(n) = [x_1, x_2, \ldots, x_n]$. The five entropy features are computed as follows:

$$Sh = -\sum_{i=1}^{n} p_i \log(p_i) \tag{5}$$

where p_i is the probability of the i^{th} sample of IMF.

$$Le = \sum_{i=1}^{n} \log(x_i^2) \tag{6}$$

where (x_i) is the i^{th} sample of IMF.

$$Th(x_i) = \begin{cases} 1, & |x_i| > \lambda \\ 0, & \text{elsewhere} \end{cases} \tag{7}$$

so,

$$Th = \#\{i \text{ such that } |x_i| > \lambda\}. \tag{8}$$

where x_i and λ are the i^{th} sample of IMF and a positive threshold, respectively. In this work, λ is set empirically to 0.2.

$$Su = \sum_{i=1}^{n} \min(x_i^2, \lambda^2) + \lambda^2 \sum_{i=1}^{n} I(|x_i| > \lambda) - \sigma^2 \tag{9}$$

$$I(|x_i| > \lambda) = \begin{cases} 1 & \text{if } |x_i| > \lambda \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where x_i is the i^{th} sample of IMF, λ is the threshold parameter that is set to 3, σ^2 is the noise variance, and I is the indicator function.

$$Nr = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{11}$$

where x_i is the i^{th} sample of IMF, and p is the norm parameter which is set to 1.1.

3.3 Classification

Cascade-forward neural network (CFNN) architecture is an artificial intelligence architecture that is designed based on the principles of feed-forward neural network (FFNN) architecture [2]. In CFNN architecture, like FFNN architecture, the flow of information is from the input layer to the output layer, and there is also no connection between the neurons in each layer. The CFNN has: input layer, hidden layers, and output layer the number of neurons in the input layer and the output layer is equal to the number of features and the number of classes, respectively. The only difference between the FFNN architecture and CFNN architecture is that in CFNN architecture, there is a connection between each layer and all the previous layers including the input layer, while in the FFNN architecture, there is a connection only between the successive layers. Thus, the CFNN can generalize the patterns and structures of data better than traditional FFNN architectures. The advantage of the CFNN architecture is that it maps any input to output, like the FFNN. Additionally, CFNN architecture finds the

nonlinear relationships between the input and output layers while preserving the linear relationships between the layers.

In the paper, a CFNN architecture is used to classify R or NR groups. The performance of CFNN architecture is compared to an FFNN architecture with the same number of layers and neurons to show the effectiveness of the CFNN architecture rather than FFNN. Additionally, the performance of the CFNN is compared to other traditional machine learning algorithms: k-nearest neighbors (KNN), support vector machine (SVM), and random forest (RF). The 10-fold cross-validation (CV) strategy is employed to train the model to avoid any bias in the results. Table 1 shows the parameter setting of the classification.

3.4 Results and Discussion

The EEG signals are divided into packages, and each package has a length of 2560 samples. The size of each package is (19×2560) where 19 is the number of channels and 2560 is the length of the package. The Mumtaz database has a total of 862 packages: 522 packages NR and 340 packages R to SSRI therapy. Atieh Hospital database has 444 packages: 168 packages NR and 276 packages R to rTMS therapy. Two binary classification tasks are implemented to predict the outcome of the two therapies using pre-treatment EEG signals. One task is to classify the pretreatment EEGs to R and NR to SSRI therapy using the Mumtaz database. The other task is to classify the pretreatment EEGs to R and NR to rTMS therapy using the Atieh Hospital database. Figure 2 illustrates the 19-channel EEG signal after the segmentation process into packages for both datasets. Figure 2 depicts the complexity of visually predicting treatment results for depression using the multi-channel EEG signal.

Table 1. Parameters setting of classification methods.

Classification method	Parameter	Value
CFNN and FFNN	Architecture	[10, 10, 10, 10, 10, 10, 10]
	Maximum number of epochs	50
RF	Number of trees	20
SVM	Kernel Function	linear
KNN	Distance metric	Euclidean
	Number of neighbors	5
CV	Number of folds	10

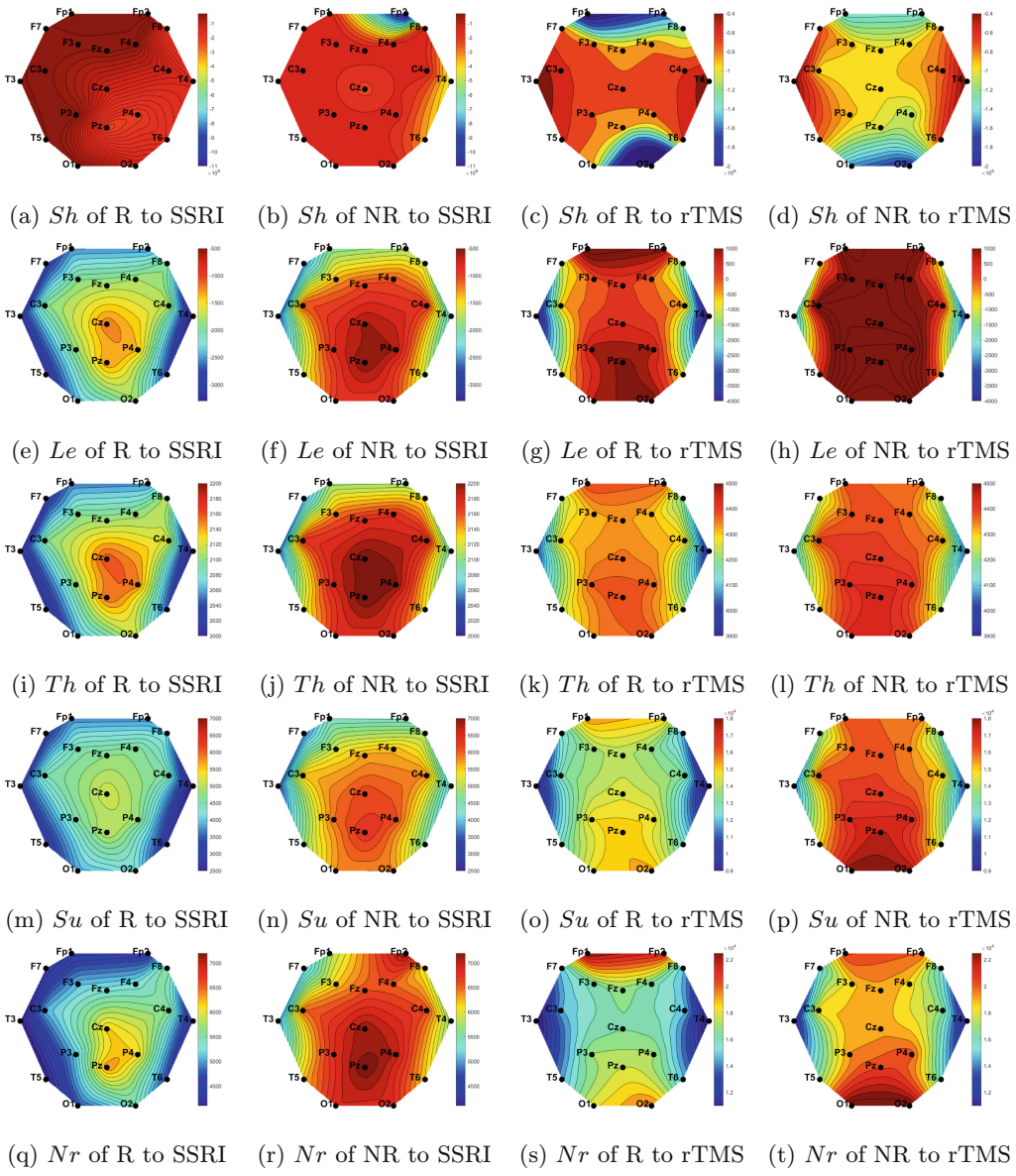
The steps of the proposed TOP-EEG software are shown in Fig. 3. First step, the EEG signals are decomposed into the IMFs using DGMD. Then, five entropy features are computed from each IMF, and the mean values of the features are considered final features. Assume that the EEG signals are decomposed into M IMFs, then five features are computed from each IMF of each channel, which

means $5 \times M$ features are extracted from each channel. The mean value of each feature is computed for all IMFs (i.e., mean of $(1 \times M)$). For example, the Sh entropy is calculated for IMFs of one channel, then the mean value of all IMFs is considered as a feature. This process is repeated for all channels and entropy features. Thus, five features are extracted for each channel, and a total of $19 \times 5 = 95$ features are extracted for each 19-channel EEG signal package. Giving the fact that the number of modes in DGMD are automatically defined and changed from one package to another, we compute the mean values of entropy features for all IMFs to extract the same number of features for all packages. The variation of the entropy-based features on the scalp is shown in Fig. 4. Figure 4 shows that the extracted entropy-based features in DGMD domain can show a significant difference between R and NR EEG signals in both SSRI and rTMS therapies. It shows that the values of entropy-based features in NR group are higher than R group. Entropy measures the value of randomness and also show the status of human brain activity. The lower value of entropy feature in R group than NR group, it might be due to the brain is in more stable state, which can be considerd as a biomarker. This fact show that EEG signals in R group are more predictable with less irregular behavior than in the NR group. Thus, the R EEG signals show the possibility to respond positively to therapeutic interventions than the NR.

The KW test shows statistically significant difference between different groups (features). Features that have a p-value less than 0.01 are considered statistically significantly difference. The results show that 60 and 49 features out of 95 are statistically significantly difference. Classification accuracy, sensitivity, and specificity are used to evaluate the performance of the proposed TOP-EEG software. The accuracy (ACC) is computed to classify R and NR patients. Sensitivity (SEN) is computed to show how confident the patient will be NR to a specific therapy. Specificity (SPE) is computed to show how confident the patient will be R to specific therapy.

Figure 5a and 5b show the accuracy, sensitivity, and specificity using different classifiers. The results show that CFNN architecture is the best classifier for both SSRI and rTMS therapies. Table 2 show the numerical values for all classifiers. The CFNN architecture achieves a classification accuracy of 93.16%, sensitivity of 96.36%, and specificity of 88.24% for SSRI therapy and accuracy of 94.59%, sensitivity of 90.48%, and specificity of 97.10% for rTMS therapy. Although the FFNN architecture achieves an acceptable accuracy of 90.77% for rTMS therapy, it is lower than CFNN. Among the traditional machine learning algorithms, the performance of RF is better than KNN and SVM in both SSRI and rTMS therapies. However, the performance of CFNN and FFNN architectures as deep learning models is still better than all traditional machine learning algorithms.

In this paper, we introduce a software called Therapy Outcomes Predictor by EEG (TOP-EEG) that can be installed and utilized on any personal computers. Figure 6 shows the developed out TOP-EEG software. The developed TOP-EEG software is built on the DGMD, entropy-based features. It is accurate and reliable by achieving promising classification performance in a 10-fold CV strategy for



(a) *Sh* of R to SSRI (b) *Sh* of NR to SSRI (c) *Sh* of R to rTMS (d) *Sh* of NR to rTMS

(e) *Le* of R to SSRI (f) *Le* of NR to SSRI (g) *Le* of R to rTMS (h) *Le* of NR to rTMS

(i) *Th* of R to SSRI (j) *Th* of NR to SSRI (k) *Th* of R to rTMS (l) *Th* of NR to rTMS

(m) *Su* of R to SSRI (n) *Su* of NR to SSRI (o) *Su* of R to rTMS (p) *Su* of NR to rTMS

(q) *Nr* of R to SSRI (r) *Nr* of NR to SSRI (s) *Nr* of R to rTMS (t) *Nr* of NR to rTMS

Fig. 4. The variation of the extracted entropy-based features on the skin.

prediction the outcomes of the SSRI therapy and rTMS therapy. In addition, it is fast, because it is simple. Thus, the developed TOP-EEG software does not require many computational resources and can be run on any standard computer available in clinics and hospitals. Figure 6 shows the developed TOP-EEG in four possible outcomes.

In Table 3, the performance of TOP-EEG is compared to recently proposed models in the literature for predicting the outcomes of SSRI and rTMT therapies. It achieves better performance in terms of ACC, SEN, and SPE compared to [8] method. The database used in [8] for SSRI therapy was also Mumtaz database.

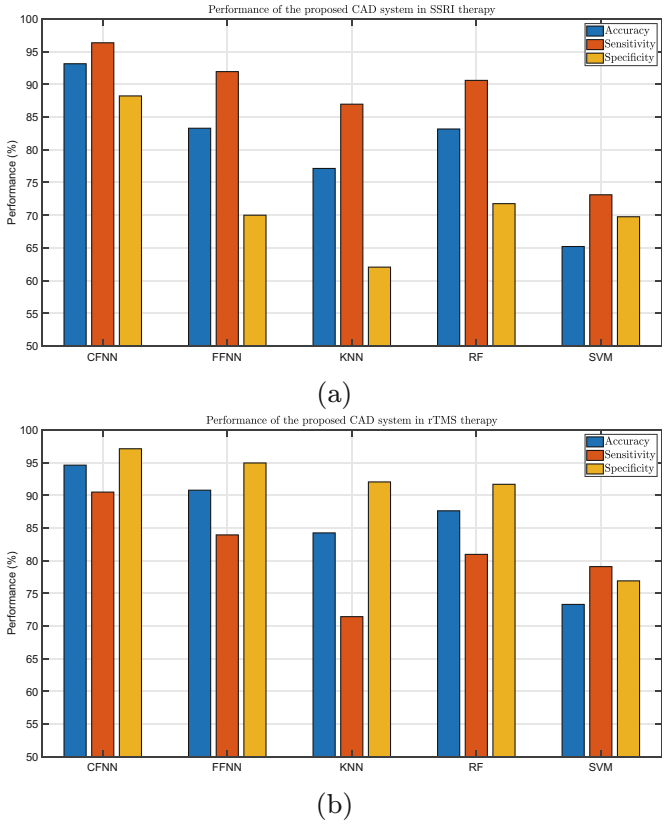


Fig. 5. Comparison of the performance of classification methods on SSRI (a) and rTMS (b) therapies.

The performance of our TOP-EEG is lower than the proposed method in [14]. In [14], the continuous wavelet transform was used to convert EEG signals into time-frequency images to feed two pre-trained convolutional neural networks: EfcientNetB0 and VGG16 to extract the feature matrix. Then, a fully connected layer was aligned with two successive long-term short memory (LSTM) networks to classify the feature matrix.

Although the performance of the proposed method in [14] achieved higher accuracy than ours, the developed TOP-EEG extracts entropy-based features from the EEG signals in DGMD domain, which is faster than generating the time-frequency images by continuous wavelet transform and using two pre-trained architectures to extract the feature matrix, as proposed in [14]. In addition, we propose a simple CFNN architecture for classification, while in [14], a fully connected layer was aligned with two successive LSTM layers, which is more complex and computationally expensive. Furthermore, the proposed method in [14] was trained on more samples 46 patients to predict the outcome of rTMS therapy, while ours is trained using 17 depressed patients.

Table 2. Performance of classification methods.

	SSRI therapy				
	CFNN	FFNN	KNN	RF	SVM
ACC(%)	93.16	83.29	77.15	83.18	65.21
SEN(%)	96.36	91.95	86.97	90.61	73.12
SPE(%)	88.24	70.00	62.06	71.76	69.76
	rTMS therapy				
	CFNN	**FFNN**	**KNN**	**RF**	**SVM**
ACC(%)	94.59	90.77	84.23	87.61	73.30
SEN(%)	90.48	83.93	71.43	80.95	79.08
SPE(%)	97.10	94.93	92.03	91.67	76.90

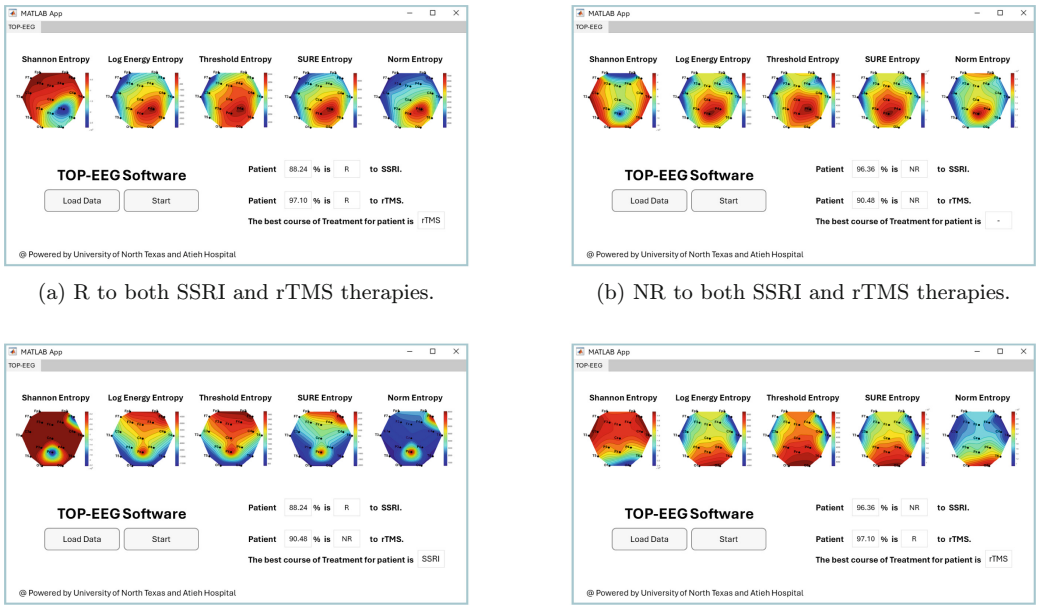
By reviewing the literature, it should be noted that this is the first time a software is developed to predict the outcomes of two different therapies to treat depression.

Table 3. Comparison of the performance of the proposed TOP-EEG software with previous works.

Ref.	Database	Therapy	Accuracy (%)	Sensitivity (%)	Specificity (%)
[8]	12 R vs. 18 NR	SSRI	91.60	90.00	90.00
TOP-EEG	12 R vs. 18 NR	SSRI	93.16	96.36	88.24
[14]	23 R vs. 23 NR	rTMS	97.10	97.30	97.00
TOP-EEG	9 R vs. 6 NR	rTMS	94.59	90.48	97.10

The contributions and important points of the current study are listed as follows:

- This is the first time in the literature that an application is developed for the DGMD technique as a robust and adaptive signal processing technique.
- The is first time in the literature, the effectiveness of entropy-based features is evaluated to predict the outcomes of depression therapies rather than detecting depression disorders.
- It is the first time in the literature that a method is developed to predict the outcome of two depression therapies (i.e., SSRI and rTMS); all other previous studies were developed to predict the outcome of one depression therapy (i.e., SSRI or rTMS).
- The value of entropy features in NR patients is higher than in R patients. We conclude that the recorded behavior of human brain in the R situation is more stable than NR situation. On the other hand, a higher entropy value in the NR patients indicates that the human brain might be in unstable



(a) R to both SSRI and rTMS therapies.

(b) NR to both SSRI and rTMS therapies.

(c) R to SSRI therapy and NR to rTMS therapy.

(d) NR to SSRI therapy and R to rTMS therapy.

Fig. 6. The developed TOP-EEG software.

status. We conclude that the unstable behavior might reflect greater neural activities compared to the R state, which might affect NR patients to respond to depression therapies.

- This is the first time to use entropy feature as a biomarker to detect the outcome of different depression therapies as shown in the software GUI Fig. 6. The topograph shows the distribution of entropy values on human scalp, which is used as a biomarker for therapies outcomes.
- A simple CFNN architecture is proposed, which is less computationally expensive than other studies in the literature.
- The proposed TOP-EEG software is evaluated on two databases with different sampling frequencies, which shows the robustness of our proposed software and its insensitivity to sampling frequencies.
- The TOP-EEG is reliable to be used in real-world applications since it is trained and tested using 10-fold CV strategy.
- We offer a cost-effective, highly accurate, and efficient TOP-EEG software that utilizes pretreatment EEG signals to predict the outcomes of SSRI and rTMS depression therapies. The design and implementation of our proposed software is performed using MATLAB software. The simple design and implementation shows that it can be utilized in hospitals and clinics to predict therapy outcomes for depressed patients.

4 Conclusion

Suicide and self-harm are the major consequences of depression if it is left untreated. The SSRI and rTMS therapies are two common treatments for depression. Although the effectiveness of SSRI and rTMS therapies has been approved by the FDA as a depression treatment, the response rates of these two therapies are about 50%. It means if they are ineffective, the risk of suicide and self-harm for the depressed patients will increase. In this study, the TOP-EEG software is developed using the DGMD technique and CFNN architecture to predict the outcomes of SSRI and rTMS therapies based on the pretreatment EEG signals of depressed patients. The TOP-EEG is trained and tested using the EEG signals of 45 depressed patients, including 30 patients for SSRI therapy and 15 patients for rTMS therapy. The software achieved classification accuracy levels of 93.16% and 94.59% to predict the outcomes of the SSRI and rTMS therapies, respectively. The results indicate that the entropy-based features of EEG signals in the DGMD domain are high for non-responding (NR) patients, and we interpret the increase in entropy level due to more brain activities. Each depression therapy is suitable to suppress a certain level of brain activity that is caused by depression. Thus, NR patients might not respond to a certain therapy, because the therapy cannot suppress the level of brain activity for those patients. In the future, the performance of the TOP-EEG software will be extended to predict the outcome of more depression treatments such as psychological therapies like tricyclic antidepressants (TCAs) and Monoamine oxidase inhibitors (MAOIs), and somatic treatments such as electroconvulsive therapy (ECT), vagus nerve stimulation (VNS), deep brain stimulation (DBS), transcranial direct current stimulation (tDCS), light therapy, ketamine, and esketamine.

Acknowledgements. This work was supported by the National Science Foundation (NSF) under grant numbers 2225229 and 2231519.

References

1. Akbari, H., Esmaili, S.: A novel geometrical method for discrimination of normal, interictal and ictal EEG signals. Traitement du Signal **37**(1), 59–68 (2020)
2. Akbari, H., Ghofrani, S., Zakalvand, P., Sadiq, M.T.: Schizophrenia recognition based on the phase space dynamic of EEG signals and graphical features. Biomed. Signal Process. Control **69**, 102917 (2021)
3. Akbari, H., Korani, W.: Early detection of depression and alcoholism disorders by EEG signal. In: Luo, B., Cheng, L., Wu, Z.-G., Li, H., Li, C. (eds.) Neural Information Processing: 30th International Conference, ICONIP 2023, Changsha, China, November 20–23, 2023, Proceedings, Part X, pp. 439–452. Springer Nature Singapore, Singapore (2024). https://doi.org/10.1007/978-981-99-8141-0_33
4. Akbari, H., et al.: Recognizing seizure using poincaré plot of EEG signals and graphical features in dwt domain. Bratislava Med. J. (2023)
5. Akbari, H., Sadiq, M.T., Payan, M., Esmaili, S.S., Baghri, H., Bagheri, H.: Depression detection based on geometrical features extracted from SODP shape of EEG signals and binary PSO. Traitement du Signal **38**(1) (2021)

6. Akbari, H., Sadiq, M.T., Rehman, A.U.: Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain. Health Inf. Sci. Syst. **9**, 1–15 (2021)
7. Dragomiretskiy, K., Zosso, D.: Variational mode decomposition. IEEE Trans. Signal Process. **62**(3), 531–544 (2013)
8. Mumtaz, W., Xia, L., Mohd Yasin, M.A., Azhar Ali, S.S., Malik, A.S.: A wavelet-based technique to predict treatment outcome for major depressive disorder. PLoS ONE **12**(2), e0171409 (2017)
9. Nazari, M.: Multiscale dynamic graph signal analysis toolbox (MDGABox) (2024). https://www.mathworks.com/matlabcentral/fileexchange/158116-multiscale-dynamic-graph-signal-analysis-toolbox-mdgabox
10. Nazari, M., Korshøj, A.R., ur Rehman, N.: Multiscale dynamic graph signal analysis. Signal Process. **222**, 109519 (2024)
11. ur Rehman, N., Aftab, H.: Multivariate variational mode decomposition. IEEE Trans. Signal Process. **67**(23), 6039–6052 (2019)
12. Riera-Serra, P., et al.: Clinical predictors of suicidal ideation, suicide attempts and suicide death in depressive disorder: a systematic review and meta-analysis. Eur. Arch. Psych. Clin. Neuroscience, 1–21 (2023)
13. Shahabi, M.S., Shalbaf, A., Maghsoudi, A.: Prediction of drug response in major depressive disorder using ensemble of transfer learning with convolutional neural network based on EEG. Biocybern. Biomedical Eng. **41**(3), 946–959 (2021)
14. Shahabi, M.S., Shalbaf, A., Rostami, R., Kazemi, R.: A convolutional recurrent neural network with attention for response prediction to repetitive transcranial magnetic stimulation in major depressive disorder. Sci. Rep. **13**(1), 10147 (2023)
15. Sinyor, M., Ming Chan, P.P., Niederkrotenthaler, T., Scott, V., Platt, S.: The development, progress, and impact of national suicide prevention strategies worldwide. Crisis: J. Crisis Interv. Suicide Prev. (2024)
16. U.S. Food and Drug Administration: Selective Serotonin Reuptake Inhibitors (SSRIs) Information (2014). https://www.fda.gov/drugs/information-drug-class/selective-serotonin-reuptake-inhibitors-ssris-information
17. U.S. Food and Drug Administration: Repetitive Transcranial Magnetic Stimulation (rTMS) Systems: Class II Special Controls Guidance (2018). https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/repetitive-transcranial-magnetic-stimulation-rtms-systems-class-ii-special-controls-guidance