

Research Article

Automatic Diagnosis of Mild Cognitive Impairment Based on Spectral, Functional Connectivity, and Nonlinear EEG-Based Features

Reza Akbari Movahed  and Mohammadreza Rezaeian 

Department of Biomedical Engineering, Hamedan University of Technology, Hamedan, Iran

Correspondence should be addressed to Mohammadreza Rezaeian; rezaeian@hut.ac.ir

Received 12 October 2021; Revised 21 May 2022; Accepted 30 July 2022; Published 11 August 2022

Academic Editor: Peng Li

Copyright © 2022 Reza Akbari Movahed and Mohammadreza Rezaeian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate and early diagnosis of mild cognitive impairment (MCI) is necessary to prevent the progress of Alzheimer's and other kinds of dementia. Unfortunately, the symptoms of MCI are complicated and may often be misinterpreted as those associated with the normal ageing process. To address this issue, many studies have proposed application of machine learning techniques for early MCI diagnosis based on electroencephalography (EEG). In this study, a machine learning framework for MCI diagnosis is proposed in this study, which extracts spectral, functional connectivity, and nonlinear features from EEG signals. The sequential backward feature selection (SBFS) algorithm is used to select the best subset of features. Several classification models and different combinations of feature sets are measured to identify the best ones for the proposed framework. A dataset of 16 and 18 EEG data of normal and MCI subjects is used to validate the proposed system. Metrics including accuracy (AC), sensitivity (SE), specificity (SP), F1-score (F1), and false discovery rate (FDR) are evaluated using 10-fold crossvalidation. An average AC of 99.4%, SE of 98.8%, SP of 100%, F1 of 99.4%, and FDR of 0% have been provided by the best performance of the proposed framework using the linear support vector machine (LSVM) classifier and the combination of all feature sets. The acquired results confirm that the proposed framework provides an accurate and robust performance for recognizing MCI cases and outperforms previous approaches. Based on the obtained results, it is possible to be developed in order to use as a computer-aided diagnosis (CAD) tool for clinical purposes.

1. Introduction

Dementia is the most prominent neurological disorder among the elderly, resulting in deterioration of cognitive abilities such as memory, thinking, behavior, and limitation in performance of daily activities [1]. While dementia mostly affects people over the age of sixty, some individuals are afflicted with this condition at a younger age. It was estimated that 44.4 million people around the world suffer from some forms of dementia [2, 3]. Alzheimer's disease (AD) is the most common type of dementia, comprising about 60% to 80% of dementia cases [4]. AD causes severe memory loss, cognitive impairment, and behavioral changes. In the United States, it is the third most costly illness and the sixth most common cause of mortality [4]. Unfortunately, there is

no effective treatment for AD and anti-AD medications merely serve to alleviate the symptoms of the disease. Hence, diagnosis of AD at an early stage is essential for controlling the progress of the disease. Mild cognitive impairment (MCI) is the intermediate stage between the normal cognitive deficit due to aging and AD or other forms of dementia, and it is frequently considered the early stage of AD. It is estimated that about 15% to 20% of dementia cases lead to AD [5]. According to some experts, diagnosis of MCI is more critical than AD, because it is more optimum to the management of the disease at the MCI stage. A more accurate method for diagnosis of MCI is therefore crucial to the control of disease progress. Nonetheless, the symptoms of MCI are often indistinguishable from those associated with growing older. Diagnosis of MCI was done integrating

multiple experiments including psychological tests such as mini-mental state examinations (MMSE), blood tests, spinal fluid analysis, neurological examination, and magnetic resonance imaging (MRI). However, these traditional methods are laborious, time-consuming, and error prone. Therefore, many MCI cases may not be diagnosed accurately and in a timely fashion. Hence, novel and automated methods for early and accurate MCI diagnosis are in high demand. Electroencephalography (EEG) is an efficient modality which records the bioelectrical activity of brain neurons corresponding to various states from the scalp surface area. EEG signals can depict the functioning of the brain with great temporal resolution and are noninvasive, inexpensive, and portable for recording. As a result, it might be employed as a biological biomarker for MCI and other types of dementia that is objective and trustworthy. The manual interpretation of these signals is tough and challenging due to their nonstationary and nonlinearity nature. Many researchers are urged to employ machine learning techniques for automated EEG signal analysis in light of the advancements in computer science and artificial intelligence. For instance, numerous studies have been conducted to automatically detect different neurological disorders such depression [6–8], epilepsy [9–11], seizure [12–15], Parkinson’s disease [16, 17], and schizophrenia [18].

EEG-based machine learning frameworks for diagnosing AD, MCI, and other forms of dementia have been reported in a number of research [19–28]. An automatic MCI diagnosis method, for instance, was presented by Kashefpoor et al. [19] and is based on the extraction of spectral characteristics of EEG signals. This method involved extracting 19 spectral features from 19-channel EEG data based on the delta, theta, alpha1, alpha2, gamma, beta1, and beta2 frequency subbands. The best discriminative features were then chosen using a correlation-based feature selection algorithm. The classification test was carried out using the neuro-fuzzy (NF) and k-nearest neighbor (KNN) classifiers, and an accuracy of 88.89% was reported in their study as the best classification result. Using spectrum and complexity analysis, McBride et al. proposed an EEG-based classification framework that distinguishes between AD, MCI, and healthy individuals [20]. The complexity analysis comprises computational activity, mobility, complexity, sample entropy, and Lempel-Ziv complexity parameters. The spectral analysis involves extracting features from the delta, theta, alpha1, alpha2, gamma, beta1, and beta2 frequency subbands. In this work, an average accuracy of 79.2% was obtained for MCI classification using the support vector machine (SVM) classifier. In a different study, Kashefpoor et al. used 19-channel EEG signals along with supervised dictionary learning techniques like label consistent K-SVD (LC-KSVD) and correlation-based label consistent K-SVD (CLC-KSVD) to diagnose MCI [21]. On two separate dictionary learning classifiers, they applied a time series signal as well as a vector of retrieved spectral features. By casting a vote between the predictions of the time series signal and spectral characteristic vector, the final forecast for each sample was determined. The CLC-KSVD approach yielded an accuracy of 88.9%, which is the highest achieved accuracy in this investigation.

A single-channel EEG-based technique for MCI diagnosis using speech-evoked brain responses was presented by Khatun et al. [22]. Using time and spectral domain analysis to extract 590 characteristics from the recorded sounds, the top 25 were chosen using the random forest (RF) method. They used logistic regression (LR) and support vector machine (SVM) classification models for the classification task, and the best classification result that they found was an accuracy of 87.9%. Based on the spectral-temporal analysis, Yin et al. proposed an integrated MCI diagnosis approach in [23]. This method uses spectral-temporal analysis to extract a collection of features and then uses a developed wrapper algorithm called the three-dimensional (3-D) evaluation algorithm to derive an ideal feature subset. The SVM classifier in this investigation had the best accuracy rate, which was 96.94%. Power spectral density (PSD), skewness, kurtosis, spectral skewness, spectral kurtosis, spectral crest factor, spectral entropy (SE), and fractal dimension (FD) properties were used by Sharma et al. to construct an automatic EEG-based MCI detection technique [24]. The study’s dataset was gathered under four different conditions: open eyes, closed eyes, the finger tapping test (FTT), and a continuous performance test (CPT). The SVM classifier achieved the best classification accuracy of this method, which was 96.94%. Fast Fourier transform (FFT) and wavelet transform (WT) feature extraction techniques were used by Durongbhan et al. in [25] to offer an automatic AD detection methodology based on EEG recordings. The presented findings show that this method used the KNN classifier to achieve an accuracy of 99%. Using the piecewise aggregate approximation (PAA) compression approach and feature extraction techniques like permutation entropy (PE) and autoregressive (AR), Siuly et al. created a system for automatically identifying MCI patients [26]. The classification models utilized in this investigation were the extreme learning machine (ELM), SVM, and KNN. According to the presented data, the ELM classifier had the best classification performance thanks to its accuracy of 98.78%. Using the interhemispheric coherence features and the properties of EEG subbands, Oltu et al. proposed another EEG-based paradigm for the classification of MCI, AD, and healthy people [27]. For feature extraction, this method used the discrete wavelet transform (DWT), PSD, and interhemispheric coherence; for classification, it used bagged trees. The best outcome of this study was 96.5%. Another strategy for differentiating between AD and healthy participants using EEG data was put out by Safi and Safi [28]. This method involved applying the DWT, PSD, and empirical mode decomposition (EMD) algorithms to the signals and then extracting certain features from the outputs of the aforementioned algorithms, including variance, kurtosis, skewness, Shannon entropy, sure entropy, and Hjorth parameters. This study compared SVM, KNN, and regularized linear discriminant analysis (RLDA) and found that KNN had the greatest classification performance with an accuracy of 97.64%.

A viable and accurate method for clinical usage has not yet been provided, despite the fact that numerous studies on automated EEG-based MCI diagnosis have been carried out. In other words, achieving an optimal/robust performance with

high accuracy, which might be used for clinical applications, is the main problem of autonomous EEG-based MCI diagnosis. So, based on EEG signals, we present a precise machine learning-based methodology for MCI detection. To do this, EEG data are processed to extract three key feature sets, including spectral, functional connectivity, and nonlinear properties. The sequential backward feature selection (SBFS) technique is used to choose the optimal feature combination. To select the optimal categorization model for the suggested methodology, various models are evaluated. Additionally, each feature set and its combinations with other feature sets are investigated in the suggested technique to determine which combination is optimal. Additionally, the most important functional connectivity aspects as well as EEG signal strength disparities in common EEG frequency subbands between MCI and HC patients were also looked into.

2. Materials and Methodology

2.1. Dataset. A public dataset consists of EEG recorded signals from MCI patients and healthy control participants, provided by Kashefpoor et al. [21], was utilized in this investigation. It contains 18 EEG data from MCI patients and 16 EEG data from healthy participants. This dataset's participants were all older than 55. At Noor Hospital in Isfahan, Iran, these participants were identified and chosen. This hospital also performed neuropsychological examinations, EEG recordings, and other experimental procedures. The individuals' diagnoses and the experimental design for data gathering were authorized by the ethics committee of the vice chancellor's office at the Isfahan University of Medical Sciences for Research and Technology in Isfahan, Iran. Based on Peterson's criteria, participants' MCI diagnoses were made. 19 EEG electrodes were put in accordance with the 10-20 International System to record EEG data from each subject while they were all resting and their eyes were closed (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2). For 30 minutes, these signals were collected. Additionally, the electrode-skin impedance was less than 5 k Ω , and the sampling rate was set to 256 Hz.

2.2. Proposed Methodology. The proposed methodology for MCI diagnosis based on EEG data is summarized in Figure 1. Preprocessing, feature extraction, feature selection, classification, and validation make up the proposed methodology. The EEG signals underwent preprocessing to remove noise and artifacts. In this step, each signal was sliced into one-minute segments after suppressing noise and artifacts. Then, using EEG segments, the spectral, functional connectivity, and nonlinear feature sets were retrieved. Following that, 10-fold crossvalidation was used to randomly divide the samples into the training and testing sets. To choose the most effective discriminative features, the training set was used during the feature selection process. The feature selection algorithm in this work was the SBFS approach. The training set was then used to train the classifier after the nonselected features had been eliminated from the testing and training sets. The trained classification model was

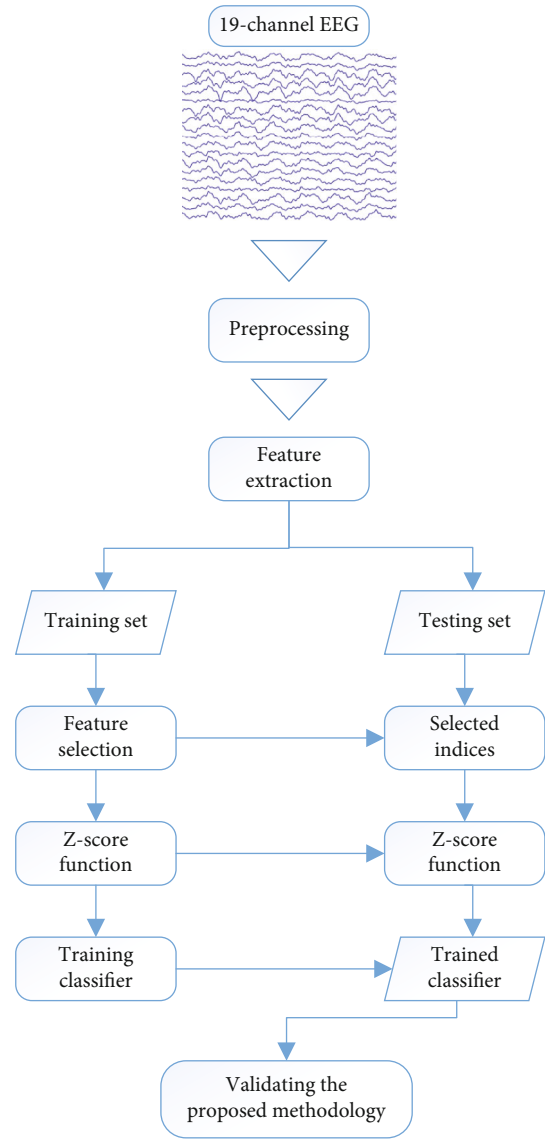


FIGURE 1: Overview of the proposed EEG-based methodology for MCI diagnosis.

then used to classify each case of the testing set for validating the suggested methodology.

2.2.1. Preprocessing. Eye movements, eye blinks, electromyogram (EMG), electrocardiograms, electrode channel drift, and power line interference are just a few of the noise and aberrations that can contaminate EEG signals and prohibit a pure portrayal of brain function in the data. In order to prevent further inaccurate analysis, it is crucial to remove noise and artifacts from the EEG signal. Using the EEGLAB toolbox in MATLAB, a preprocessing approach was conducted to all recorded EEG signals in this investigation [29]. A band-pass filter with a low cutoff frequency of 0.5 Hz and a high cutoff frequency of 32 Hz was used to filter the signals in the first step. Due to the components of the aforementioned artifacts being focused outside the frequency band between 0.5 Hz and 32 Hz, it reduces the effects

of EMG and power line interference. The independent components of the signals were then extracted using the independent component analysis (ICA) methodology, and using a voting classification method, each component was classified into artifact and nonartifact classes. The ICLabel [30] and MARA [31] automatic plugins, along with manual inspection, projected a label for each component in this voting classification approach, and the final predicted label was the one that obtained more than half of the votes. The signals were then rebuilt after the expected artifact components had been removed. The remaining noisy intervals were then taken out of the reconstructed signals using a visual analysis. Finally, to increase the amount of samples, the preprocessed EEG signals were sliced into one-minute segments using a nonoverlapping sliding window. It should be noted that the labels on the EEG data segments matched those on the EEG signal's original label.

2.2.2. Feature Extraction. By converting each raw sample's values into a select few useful features, feature extraction in machine learning is aimed at representing each sample with the qualities that are task relevant. The spectral, functional connectivity, and nonlinear feature sets are extracted from the EEG segments using the suggested methods. The explanation of each extracted feature set is as follows:

- (1) *Spectral features:* the spectral properties of the EEG signals are intended to depict the characteristics of the frequency subbands of the EEG at various scalp locations. Due to their relation to brain's function, these characteristics may serve as diagnostic criteria for neurological illnesses. In this study, the spectral feature set was created by extracting the band power of each EEG segment's interhemispheric asymmetry and theta (4 to 8 Hz), alpha (8 to 13 Hz), and beta (13 to 32 Hz) frequency subbands. To determine the band power of frequency subbands, the PSD of EEG segments was calculated using the Welch periodogram [32]. The Hamming window with a 50% overlap between the windows was used in the Welch periodogram. It is important to note that each channel of the EEG segments was used to extract the band powers of the aforementioned frequency subbands. The definition of interhemispheric asymmetry, which measures disparities in the band power of the frequency subbands in the left and right hemispheres, is as follows:

$$IA = \log(P_{RH}) - \log(P_{LH}), \quad (1)$$

where IA , P_{RH} , and P_{LH} stand for the interhemispheric asymmetry, the band power in the right hemisphere, and the band power in the left hemisphere, respectively. The interhemispheric asymmetry for the channel pairings Fp2-Fp1, F4-F3, F8-F7, C4-C3, T4-T3, P4-P3, T6-T5, and O2-O1 was computed in this work for the delta, theta, alpha, and beta frequency subbands

- (2) *Functional connectivity features:* numerous research has looked at brain connection in recent years to understand how information is processed, sent to, received by, or shared between various brain regions during various cognitive tasks and mental states. A branch of neuroscience known as functional connectivity seeks to quantify the statistical relationships between the dynamics of concurrently recorded signals [33] in order to gauge brain connectivity. Coherence, mutual information, and synchronization likelihood are only a few examples of functional connectivity metrics used to assess brain connectivity. Here, a collection of characteristics based on the statistical correlations between EEG channels in various scalp areas was extracted using the synchronization likelihood method [34]. In general, synchronization likelihood analyzes the nonlinear and linear dependencies between two signals, which may be complex and significantly different in the two signals, to estimate the synchronization between them. The possibility of synchronization between two signals is represented numerically by a value between zero and one. More synchronization between two signals is indicated by a greater value for this metric. First, a time-delay embedding method that is specified as follows [34] was used to create a state-space representation of an M -channel EEG signal:

$$X_{k,i} = [x_{k,i}, x_{k,i+l}, \dots, x_{k,i+(m-1)l}], \quad (2)$$

where $k \in \{1, 2, \dots, M\}$, $i \in \{1, 2, \dots, N\}$, m, l are the channel number, the index of each discrete sample, embedding dimension, and lag parameters, respectively. Next, $(P_{k,i}^\varepsilon)$ is defined as follows to state that $X_{k,i}$ and $X_{k,j}$ vectors are closer than a distance of ε [34].

$$P_{k,i}^\varepsilon = \frac{1}{2(\omega_1 - \omega_2)} \sum_{j=1}^N \phi(\varepsilon - |X_{k,i} - X_{k,j}|), \quad \omega_1 < |i - j| < \omega_2, \quad (3)$$

where ϕ , $|\cdot|$, ω_1 , and ω_2 stand for Heaviside step function, Euclidean distance, dimension of the Theiler correction window, and dimension of the sharpening window, respectively. Each of the Theiler and sharpening windows establish a window around the discrete sample i to modify autocorrelation consequences and sharpen the time resolution of the synchronization measure. Now, $\varepsilon_{k,i}$ is determined for each k and each i for which $P_{k,i}^{\varepsilon_{k,i}} = p_{\text{ref}}$, where $p_{\text{ref}} \ll 1$. In the next step, the number of channels where the $X_{k,i}$ and $X_{k,j}$ will be closer together than $\varepsilon_{k,i}(H_{i,j})$ is determined for each sample pair (i, j) and within the considered window ($\omega_1 < |i - j| < \omega_2$) as follows [34]:

$$H_{i,j} = \sum_{k=1}^M \phi(\varepsilon_{k,i} - |X_{k,i} - X_{k,j}|). \quad (4)$$

$H_{i,j}$ indicates how many of the embedded time series signals resemble each other, and it varies between 0 and M . Now, the synchronization likelihood for each channel (k) and each discrete sample pair (i, j), ($S_{k,i,j}$) is defined in (5) [34].

$$S_{k,i,j} = \begin{cases} \frac{H_{i,j} - 1}{M - 1}, & \text{if } |X_{k,i} - X_{k,j}| < \varepsilon_{k,i}, \\ 0, & \text{if } |X_{k,i} - X_{k,j}| > \varepsilon_{k,i}. \end{cases} \quad (5)$$

To obtain the rate of synchronization between channel k at sample i and all other $M - 1$ channels ($S_{k,i}$), an averaging over all j of $S_{k,i,j}$ is performed as follows [34]:

$$S_{k,i} = \frac{1}{2(\omega_1 - \omega_2)} \sum_{j=1}^N S_{k,i,j}, \quad \omega_1 < |i - j| < \omega_2. \quad (6)$$

$S_{k,i}$ ranges between p_{ref} and 1. $S_{k,i} = 1$ indicates the maximum synchronization of all M channels, and $S_{k,i} = p_{\text{ref}}$ corresponds with the case where all M channels have minimum synchronization. Finally, the average of $S_{k,i}$ over all i is computed and expressed as synchronization likelihood between k channels. In this work, p_{ref} , l , m , ω_1 , and ω_2 were set to 0.01, 10, 10, 100, and 410, respectively, and the synchronization likelihood between the same channels was not calculated. It is worth mentioning that these parameters were chosen using trial and error experiments to obtain the most reasonable connectivity image

(3) *Nonlinear features*: EEG signals by their very nature have complicated behavior and nonlinear dynamic properties. In light of this, nonlinear analysis techniques may be superior than conventional linear analysis techniques for describing EEG signals. In this study, certain nonlinear features, including detrended fluctuation analysis, Higuchi fractal properties, correlation dimension, Lyapunov exponent, C0-complexity, Kolmogorov entropy, Shannon entropy, and approximate entropy, were computed from each channel of the EEG segments. The features are each further detailed in the paragraphs as follows

(a) *Detrended fluctuation analysis*: it is an algorithm for estimating the statistical self-affinity of a time series [35]. Consider a finite signal, $x(t)$ of length N . First, a summation version of $x(t)$ is obtained as follows:

$$X(k) = \sum_{i=1}^k (x(i) - \bar{x}), \quad (7)$$

where \bar{x} is the mean value of $x(t)$. Next, $X(k)$ is segmented into the n windows with equal lengths and a least-square line is calculated by minimizing the squared errors within each window. $Y_n(k)$ denotes the resulting least-square line fitting.

Then, the root-mean-square deviation from the trend, the fluctuation, is computed as follows:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N (X(k) - Y_n(k))^2}, \quad (8)$$

where $F(n)$ is the fluctuation. Finally, the computing of (8) is replicated for windows with different sizes to form a logarithmic scale of $F(n)$ against n . It can be denoted by $F(n) = n^\alpha$, in which α represents the self-affinity of the signal. In other words, α is the extracted feature by detrended fluctuation analysis

(b) *Higuchi fractal properties*: Higuchi proposed an algorithm to calculate the fractal dimension of a signal in 1988 [36]. Given a signal with N samples ($x(t)$), T new signals are generated using the following equation:

$$X_\tau^T = \left\{ x(\tau), x(\tau + T), \dots, x\left(\tau + \left\lceil \frac{N - \tau}{T} \right\rceil\right) \right\}, \quad (9)$$

where $\tau = 1, 2, \dots, T$ and $\lceil r \rceil$ denotes the integer part of r . Let $L_\tau(T)$ represent the length of each signal which is defined as follows:

$$L_\tau(T) = \frac{1}{T} \frac{\sum_{i=1}^{\lceil (N-\tau)/T \rceil} |x(\tau - iT) - x(\tau(i-1)T)| \times (N-1)}{\lceil (N-\tau)/T \rceil}. \quad (10)$$

Also, $L(T)$ is defined to obtain the mean length for each signal as follows:

$$L(T) = \sum_{\tau=1}^T L_\tau(T). \quad (11)$$

Finally, (11) is computed for all T values ranging from T_{min} to T_{max} and the slope of the linear fitting of $\ln L(T)$ versus $\ln 1/T$ is considered as the Higuchi fractal dimension of $x(t)$. In this study, the T_{min} and T_{max} were set to 1 and 30 values, respectively. These parameters were selected such that the Higuchi fractal characteristic could be extracted from the study's data

(c) *Correlation dimension*: it is another approach for calculating the fractal dimension by measuring the occupied space by a set of random points. In 1983, Grassberger and Procaccia presented a method for computing correlation dimension, which is the most common method for estimating correlation dimension [37]. Firstly, it constructs an m -dimensional vector using time delay (τ) and embedding dimension (m), which can be denoted as follows:

$$X(i) = [x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)], \quad (12)$$

where $i = 1, 2, \dots, N - (m - 1)\tau$, x is the signal with N samples, and X is the m -dimensional vector. Afterwards, the correlation integral of X is defined as the probability that two points of the set are in the same partition of size r . It is obtained using the following equation:

$$C(r) = \frac{2}{N(N-1)} \sum_{i \neq j} \phi(r - |X(i) - X(j)|), \quad (13)$$

where $C(r)$ and ϕ represent the correlation integral and the Heaviside step function, respectively. In the next phase, (14) is utilized to obtain the raw correlation dimension.

$$D = \lim_{r \rightarrow 0} \frac{\ln C(r)}{\ln(r)}. \quad (14)$$

Finally, the different values of D are computed using (14) for the incremental value of m . This process causes a gradual of D , and eventually, D reaches saturation. The saturated value of D is considered as the estimated correlation dimension of the signal

- (d) *Lyapunov exponent*: the chaos of a dynamic system is quantified by the Lyapunov exponent, which estimates the development or decay rate of minor perturbations along each major axis of the phase space system [38]. Consider a dynamic system with the d dimension. It is possible to determine the d number of Lyapunov exponents for this system. However, in the majority of real-world applications, the greatest Lyapunov exponent (LLE) is regarded as the extracted feature by the Lyapunov exponent. A dynamic system's maximal Lyapunov exponent (λ_1) is defined as follows:

$$d_j(i) = d_j(0) \exp(\lambda_1 i \Delta t), \quad (15)$$

where $d_j(i)$ denotes the mean Euclidian distance between two neighbor trajectories at i time and $d_j(0)$ represents the Euclidian distance between the j th pair of initially most adjacent neighbors after i time. In order to compute the LLE, the following equation is used:

$$y(i) = \frac{1}{\Delta t} < \ln(d_j(i)) >, \quad (16)$$

where $y(i)$ and $< \ln(d_j(i)) >$ represents the approximated LLE and the average value of the natural logarithm of $d_j(i)$ over all values of j , respectively

- (e) *C0-complexity*: it is a measure that is aimed at quantifying irregularities of a signal by defining the ratio of the irregular components to the original signal [39]. Consider a signal $x(n)$ with the N number of samples. Firstly, the fast Fourier transform of the

TABLE 1: The more details of each feature set.

Feature set	Number of features
Functional connectivity	108
Spectral	171
Nonlinear	152

$x(n)$ ($X(k)$) is computed and the average value of the magnitude of $X(k)$ (M) is obtained as follows:

$$M = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2. \quad (17)$$

Now, a spectrum called $Y(k)$ is constructed using $X(k)$ and M as follows:

$$Y(k) = \begin{cases} X(k), & |X(k)|^2 > M, \\ 0, & |X(k)|^2 < M. \end{cases} \quad (18)$$

By applying inverse Fourier transform to $Y(k)$, $y(n)$ is obtained and the C0-complexity of the $x(n)$ is provided as follows:

$$C0 = \frac{A_1}{A_0} = \frac{\sum_{n=0}^{N-1} |x(n) - y(n)|^2}{\sum_{n=0}^{N-1} |x(n)|^2}, \quad (19)$$

where $C0$, A_1 , and A_0 represent C0-complexity, the power of irregular, and regular parts of $x(n)$, respectively

- (f) *Kolmogorov entropy*: it reflects the loss of information's rate of a signal to quantify its chaotic degree [40]. In order to compute it, an equation based on the average rate of the loss of information of a signal with n samples is defined as follows:

$$KE = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n\tau} \sum_{i_0 \dots i_{n-1}} P_{i_0 \dots i_{n-1}} \ln(P_{i_0 \dots i_{n-1}}), \quad (20)$$

where $P_{i_0 \dots i_{n-1}}$ and KE denote the loss of information per each sample and estimated Kolmogorov entropy, respectively. It is worth mentioning that the positive and finite value of KE represents that the dynamic phenomena in the signal are chaotic. Moreover, the zero value indicates that the signal contains regular phenomena and infinite KE corresponds with the existence of nondeterministic phenomena in the signal

- (g) *Shannon entropy*: it is another metric for quantifying the chaotic rate of a signal, proposed by Shannon [41]. Given a signal with N samples, the Shannon entropy is defined as follows:

Input: The entire feature set, $Y = \{y_1, y_2, \dots, y_d\}$
Output: The selected feature subset, $S_k = \{S_j | j = 1, 2, \dots, k; S_j \in Y\}$ where $k = 1, 2, \dots, d$

1. Start with the entire set, $S_0 = Y$.
2. Eliminate the worst feature, $s^* = \text{argmax}(J(S_k - s))$, where $s \in Y - S_k$.
3. Update $S_{k+1} = S_k - s^*$; $k = k + 1$.
4. If $J(S_k) < J(S_{k-1})$, go to step 6.
5. Else go to the step 2.
6. Stop.

ALGORITHM 1: The SBFS algorithm.

TABLE 2: The obtained results of the proposed framework using different classifiers in terms of the percentage (%) of the mean and standard deviation of AC, SE, SP, F1, and FDR metrics.

Classifier	AC (mean \pm Std)	SE (mean \pm Std)	SP (mean \pm Std)	F1 (mean \pm Std)	FDR (mean \pm Std)
LSVM	99.4 \pm 1.8	98.8 \pm 3.5	100 \pm 0	99.4 \pm 1.8	0 \pm 0
RBFSVM	95.9 \pm 3.9	100 \pm 0	91.1 \pm 8.4	96.3 \pm 3.6	6.8 \pm 6.7
LR	87.8 \pm 4.2	85.2 \pm 11.6	90.0 \pm 10.8	87.5 \pm 5.5	8.2 \pm 8.7
KNN	98.2 \pm 2.8	97.5 \pm 5.6	98.7 \pm 3.9	98.1 \pm 3.2	1.0 \pm 3.1
DT	80.2 \pm 8.8	79.3 \pm 15.6	81.6 \pm 15.2	80.1 \pm 10.1	16.8 \pm 11.7
GB	81.4 \pm 7.1	81.0 \pm 13.1	81.8 \pm 10.5	81.2 \pm 9.4	17.3 \pm 9.8
NB	77.4 \pm 7.3	63.2 \pm 15.2	92.5 \pm 7.0	73.4 \pm 11.9	9.9 \pm 9.6
RB	86.6 \pm 10.3	85.8 \pm 9.4	87.0 \pm 16.5	87.3 \pm 87.3	10.5 \pm 11.1

$$H = - \sum_{i=1}^N p_i \ln(p_i), \quad (21)$$

where H and p_i represent the Shannon entropy of the signal and the probability of having the i sample in the signal, respectively

(h) *Approximate entropy*: it is an algorithm to estimate the rate of regularity and the unpredictability of fluctuations of a signal [42]. If this estimation is higher, the signal contains more irregularity. In the first step of this algorithm, a sequence of vectors ($X(i)$) is constructed from a signal as follows:

$$X(i) = [x(i), x(i+1), \dots, x(i+m-1)], 1 \leq i \leq N-m+1, \quad (22)$$

where $x(n)$ is the signal with N samples. In the next phase, the distance between $X(i)$ and $X(j)$ ($D[X(i), X(j)]$) is computed using the following equation:

$$D[X(i), X(j)] = \max_{k=1,2,\dots,m} [|x(i+k-1) - x(j+k-1)|], \quad (23)$$

where $|\cdot|$ is the Euclidean distance. Now, $C_i^m(r)$ is calculated for each i , $i = 1, 2, \dots, N-m$ as follows:

$$C_i^m(r) = \frac{\text{number of } D[X(i), X(j)] \leq r}{N-m-1}, \quad (24)$$

where r denotes the threshold for $D[X(i), X(j)]$. Finally, the approximate entropy (ApEn) is determined as follows:

$$\text{ApEn} = \Phi^m(r) - \Phi^{m+1}(r), \quad (25)$$

where $\Phi^m(r)$ is defined as follows:

$$\Phi^m(r) = \frac{1}{N-m-1} \sum_{i=1}^{N-m-1} \ln(C_i^m(r)) \quad (26)$$

In this work, the values of m and r were set to 2 and 0.2 var (x), respectively.

The amount of features in each feature set in the suggested methods is displayed in Table 1. Table 1 shows that the spectral, functional connectivity, and nonlinear feature sets, comprising 108, 171, and 152 characteristics, respectively. It is important to note that the aforementioned features are vectorially concatenated to provide a vector for each data. There are 19 features in each frequency subband power. We recovered 8 features for theta, alpha, beta, and theta frequency subbands for the IA feature. 32 features were therefore included in the IA. Additionally, each nonlinear feature has 19 features.

2.2.3. Feature Selection. In machine learning frameworks, the goal of feature selection is to find the best subset of features to improve the classification performance by ignoring irrelevant and redundant attributes. In this study, a wrapper feature selection algorithm called SBFS was employed to obtain the best features' subset for discriminating MCI and

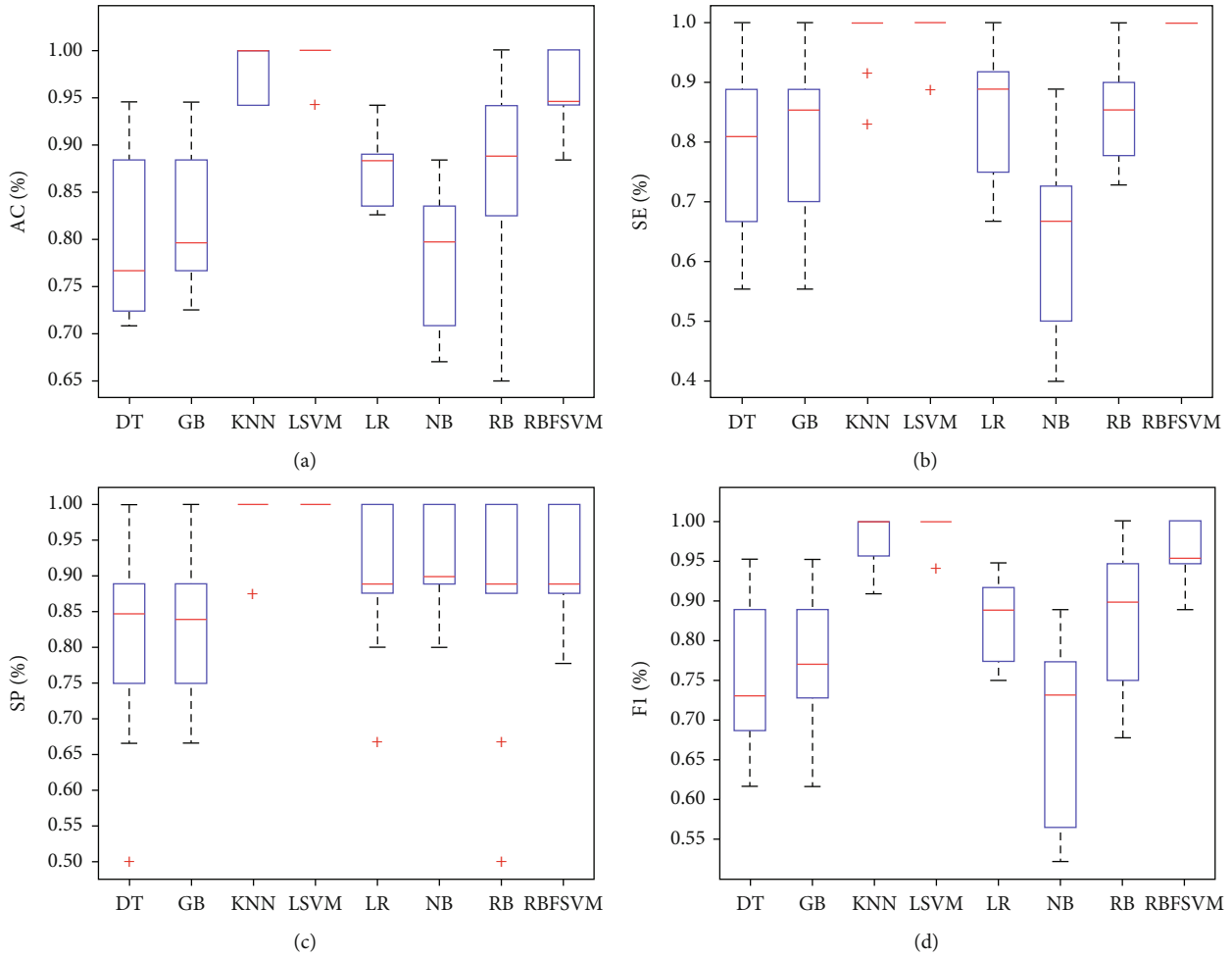


FIGURE 2: The box plots of the obtained values of AC (a), SE (b), SP (c), and F1 (d) metrics per classifier using the 10-fold crossvalidation method.

HC samples. The main steps of the SBFS algorithm are illustrated in Algorithm 1. Briefly, SBFS attempts to obtain the best features' subset by sequentially eliminating features from the entire feature set. As shown in Algorithm 1, removing features continues as long as the objective criterion is ascending [43]. In this work, the criterion (J) is set to the average accuracy in 10-fold crossvalidation. It should be mentioned that the classifier of the SBFS method was set to linear discriminant analysis (LDA) in our study.

2.2.4. Classification. The main component of supervised machine learning frameworks is classification, which is aimed at predicting a class label or a specific example of incoming data. Classification models carry out this task. The classification model for the suggested machine learning framework was chosen by comparing a number of them in this study, including SVM with linear (LSVM) and radial basis function (RBFSVM) kernels, LR, KNN, decision tree (DT), naive Bayes (NB), RUSBoost (RB), and GentleBoost (GB). Additionally, before training and testing classifiers, the training and testing sets were applied to the z -score transformation. Finding the optimum decision boundary

that can categorize an n -dimensional feature space into classes is the primary objective of SVM models. A hyperplane and a hyper radial basis curve are the ideal decision boundaries for LSVM and RBFSVM, respectively. The extreme points or vectors selected by these models aid in determining the optimal decision boundary. Support vectors are used to describe these severe situations. A nonparametric supervised learning technique that makes advantage of neighbor similarity is the KNN algorithm. A class membership is the result of this method. A datum is assigned to the class that has the highest percentage of support from its k closest neighbors after receiving a majority vote from those neighbors. A DT method is a decision-support tool that categorizes each item class using a tree-like model of decisions and their potential outcomes. A subset of probabilistic classifiers called NB is based on the Bayes theorem and assumes independence between sample features. On ensemble classifiers, the RB and GB models are built. Using different base models as a starting point, ensemble learning creates a new classifier that outperforms all of its component classifiers. The Bayesian optimizer was also utilized to optimize the hyperparameters of LSVM, RBFSVM, LR, DT, NB, RB, and GB models.

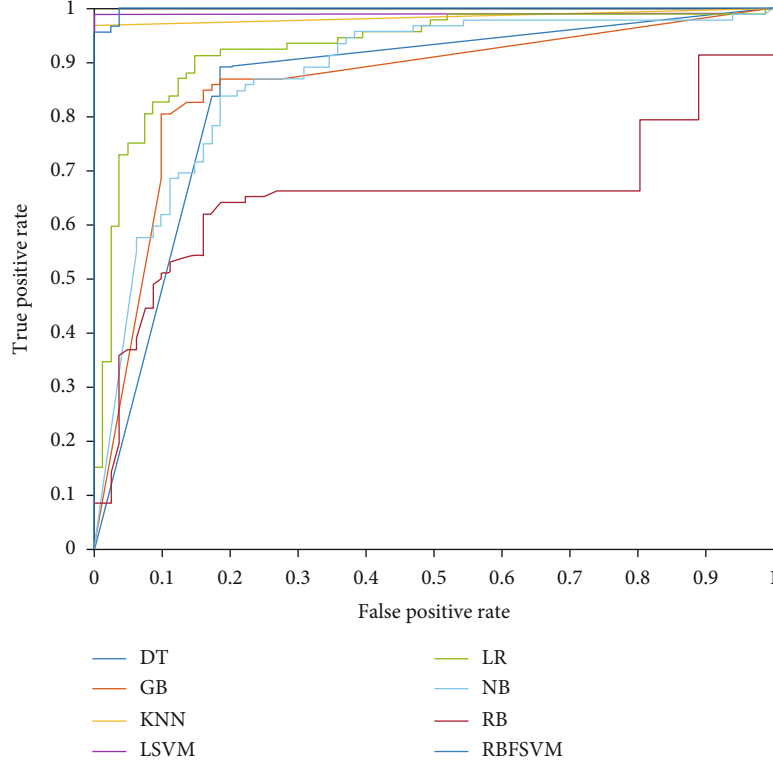


FIGURE 3: The ROC plots of the proposed framework per classifier using the 10-fold crossvalidation method.

2.2.5. Validation. The 10-fold crossvalidation strategy was utilized in this work to assess the classification performance of the suggested MCI diagnosis method. This method divides the dataset into 10 folds at random. The model is then trained using a subset of 9 folds as the training set and validated using a subset of 1 fold as the testing set. To make each fold the testing subset once, this procedure is repeated ten times. The classification performance of the suggested method may be assessed using the accuracy (AC), sensitivity (SE), specificity (SP), F1-score (F1), and false discovery rate (FDR) performance metrics by applying the testing set to the trained model and comparing the predicted and actual labels. The following is how the aforementioned metrics are calculated:

$$\begin{aligned}
 AC &= \frac{TP + TN}{TP + FN + TN + FP}, \\
 SE &= \frac{TP}{TP + FN}, \\
 SP &= \frac{TN}{FP + TN}, \\
 F1 &= \frac{2TP}{2TP + FP + FN}, \\
 FDR &= \frac{FP}{FP + TN},
 \end{aligned} \tag{27}$$

where TP is the number of MCI cases that are correctly predicted, FN is the number of MCI cases that are incorrectly

predicted as HC samples, FP is the number of HC cases that are incorrectly predicted as MCI cases, and TN is the number of HC samples that are correctly predicted.

3. Results

This section evaluates the performance of the suggested framework from a number of angles. The obtained results of the classification models using the suggested strategy are provided in the first portion in order to choose the best of them. The next step was to select the ideal combination for the suggested framework by using each set of attributes and their combinations as the input framework. Then, in order to examine the difference between the two groups based on the aforementioned features, the EEG signal strength disparities in the alpha, beta, theta, and delta frequency subbands as well as the most important functional connectivity features between MCI and HC cases are investigated. The EEG signal band powers and functional connectivity feature sets, among the recovered feature sets in this study, may offer some biological notions of MCI. To ascertain which EEG signal band powers and functional connectivity coefficients most significantly differ between MCI and HC subjects, we presented these sections. The intersection of the returned subsets by SBFS in the execution of the suggested framework's 10-fold crossvalidation is then reported and examined. Finally, a report and analysis of the leave-one-participant-out crossvalidation approach's acquired results for the suggested framework was made.

TABLE 3: The classification results of the proposed method based on RBFSVM, LSVM, and KNN classification models using different EEG feature sets in terms of the percentage (%) of the mean and standard deviation of AC, SE, SP, F1, and FDR metrics.

Feature set	Classifier	AC (mean \pm Std)	SE (mean \pm Std)	SP (mean \pm Std)	F1 (mean \pm Std)	FDR (mean \pm Std)
Functional connectivity + spectral + nonlinear	LSVM	99.4 \pm 1.8	98.8 \pm 3.5	100.0 \pm 0.0	99.4 \pm 1.8	0.0 \pm 0.0
	RBFSVM	95.9 \pm 3.9	100.0 \pm 0.0	91.1 \pm 8.4	96.3 \pm 3.6	6.8 \pm 6.7
	KNN	98.2 \pm 2.8	97.5 \pm 5.6	98.7 \pm 3.9	98.1 \pm 3.2	1.0 \pm 3.1
Functional connectivity + nonlinear	LSVM	98.8 \pm 2.4	98.7 \pm 3.9	99.0 \pm 3.1	98.7 \pm 2.6	1.1 \pm 3.5
	RBFSVM	94.8 \pm 5.6	100.0 \pm 0.0	89.3 \pm 10.8	95.2 \pm 5.3	8.6 \pm 9.3
	KNN	98.8 \pm 2.4	100.0 \pm 0.0	97.0 \pm 6.7	99.0 \pm 2.1	1.8 \pm 4.0
Functional connectivity + spectral	LSVM	98.8 \pm 2.4	98.5 \pm 4.5	98.8 \pm 3.5	98.6 \pm 2.8	1.1 \pm 3.5
	RBFSVM	97.1 \pm 4.1	100.0 \pm 0.0	93.2 \pm 9.8	97.2 \pm 3.9	6.5 \pm 7.9
	KNN	98.8 \pm 2.4	98.7 \pm 3.9	97.5 \pm 7.9	98.9 \pm 2.2	0.6 \pm 2.1
Spectral + nonlinear	LSVM	94.1 \pm 5.5	95.9 \pm 6.8	91.8 \pm 9.9	94.6 \pm 4.7	6.0 \pm 7.1
	RBFSVM	94.2 \pm 7.2	99.0 \pm 2.8	88.7 \pm 17.6	94.6 \pm 6.1	8.6 \pm 11.0
	KNN	98.2 \pm 3.9	98.0 \pm 4.2	98.5 \pm 4.5	98.4 \pm 3.4	1.0 \pm 3.1
Spectral	LSVM	93.0 \pm 7.9	92.3 \pm 9.5	93.8 \pm 8.1	93.2 \pm 8.5	5.6 \pm 8.4
	RBFSVM	95.3 \pm 3.6	99.0 \pm 3.1	90.6 \pm 9.1	95.8 \pm 3.2	6.7 \pm 6.4
	KNN	98.8 \pm 2.4	98.0 \pm 4.2	100.0 \pm 0.0	98.9 \pm 2.2	0.0 \pm 0.0
Functional connectivity	LSVM	97.6 \pm 4.9	100.0 \pm 0.0	95.1 \pm 10.4	97.8 \pm 4.6	3.8 \pm 8.3
	RBFSVM	98.1 \pm 2.5	97.7 \pm 3.4	98.6 \pm 2.3	98.2 \pm 2.3	1.2 \pm 2.1
	KNN	97.0 \pm 2.6	96.8 \pm 3.7	96.1 \pm 3.8	96.7 \pm 2.4	3.2 \pm 2.9
Nonlinear	LSVM	86.7 \pm 8.6	86.9 \pm 9.4	86.5 \pm 13.6	86.5 \pm 9.5	12.9 \pm 13.6
	RBFSVM	83.8 \pm 11.3	89.1 \pm 12.9	78.7 \pm 16.4	84.2 \pm 12.6	19.1 \pm 15.1
	KNN	87.4 \pm 8.1	85.2 \pm 11.8	91.7 \pm 11.4	87.1 \pm 9.2	8.9 \pm 12.5

3.1. Results per Classifiers. The outcomes of the proposed framework employing the aforementioned classification models are listed in Table 2. According to Table 2's findings, LSVM had the best classification performance. The given framework utilized the LSVM model to achieve an average AC of 99.4%, SE of 98.8%, SP of 100%, F1 of 99.4%, and FDR of 0%. Comparing this classifier to the others, it produced results with higher means for AC, SP, and F1 and a lower mean for FDR. Additionally, it outperformed the other classification models in terms of performance measures with the lowest standard deviation, demonstrating the greater stability of its performance within the framework that was presented. According to the results in Table 2, KNN and RBFSVM were the second and third best classification models in the framework that was provided. In comparison to other classifiers, RBFSVM earned the SP metric's greatest mean and lowest standard deviation. With the lowest mean of AC, SE, and F1 metrics among the employed classifiers, NB offered the worst classification performance. The classification performance of the other classifiers, including RB, GB, DT, and LT, was comparable. The box plots of the AC, SE, SP, and F1 metric-achieved values by the proposed framework utilizing classification models are shown in Figure 2. As depicted in Figure 2, the acquired AC, SE, SP, and F1 metric values of LSVM, KNN, and RBFSVM were reasonably high, confirming that the suggested method's

use of these classifiers produced acceptable performance. LSVM outperformed the other classifiers in terms of AC, SE, SP, and F1 performance metrics. Additionally, LSVM's boxplots for the aforementioned metrics were less negative than those for other classifiers. These findings show that, in terms of classification accuracy and performance resilience, the LSVM model is better to alternative classifiers for the suggested framework. Overall, the findings showed that LSVM, KNN, and RBFSVM were the three best classifiers for the suggested framework. The receiver operator characteristic (ROC) curves for each classification model in the proposed framework are shown in Figure 3. DT, GB, KNN, LSVM, LR, NB, RB, and RBFSVM had areas under the curve (AUC) of 0.85, 0.86, 0.98, 0.99, 0.92, 0.87, 0.65, and 0.99, respectively. These findings showed that employing the LSVM and RBFSVM classification models, the proposed framework had the best classification performance.

3.2. Results per Feature Set. The optimal input for the suggested framework was chosen in this part using each feature set and its combinations as the input classification framework. It is important to note that the three top classifiers from the previous section carried out these evaluations. The classification outcomes of the proposed framework employing each input and the LSVM, RBFSVM, and KNN classification models are shown in Table 3. Based on the

TABLE 4: The t -test results on the alpha, theta, beta, and delta EEG band powers in each EEG channel. Italic items indicate p value < 0.01 .

Brain region	EEG channel	Alpha p value	Beta p value	Delta p value	Theta p value
Frontal	Fp1	0.9474	0.2499	0.3600	0.9732
	Fp2	0.1083	0.0267	0.5997	0.3103
	F7	0.2214	0.3346	0.2985	0.2593
	F3	0.0871	0.1177	0.2291	0.2874
	Fz	<i>0.0001</i>	<i>0.0001</i>	0.0863	<i>0.0032</i>
	F4	<i>0.0011</i>	<i>0.0091</i>	0.3181	0.0335
Central	F8	<i>0.0001</i>	<i>0.0001</i>	0.0241	<i>0.0005</i>
	C3	0.2783	0.6267	0.4596	0.2667
	Cz	<i>0.0069</i>	<i>0.0001</i>	0.0527	0.0174
Occipital	C4	0.2179	0.0588	0.8330	0.5390
	O1	0.0301	0.0300	0.1801	0.1105
Parietal	O2	<i>0.0001</i>	<i>0.0001</i>	<i>0.0031</i>	<i>0.0001</i>
	P3	0.1187	0.7611	0.5157	0.3680
	P4	<i>0.0036</i>	<i>0.0001</i>	<i>0.0049</i>	<i>0.0014</i>
	Pz	<i>0.0051</i>	<i>0.0001</i>	<i>0.1577</i>	<i>0.0078</i>
Temporal	T3	0.6706	0.4975	0.4047	0.4661
	T5	0.0231	<i>0.0060</i>	0.1148	0.0863
	T4	<i>0.0033</i>	<i>0.0061</i>	0.0409	0.0164
	T6	<i>0.0006</i>	<i>0.0002</i>	0.0772	<i>0.0068</i>

reported results in Table 3, the integration of functional connectivity, spectral, and nonlinear feature sets provided the best classification performances by obtaining the highest average of AC, SE, SP, and F1 parameters and the lowest means of the FDR metric. Additionally, as compared to the other inputs, the integration of the aforementioned feature sets yielded the lowest standard deviations of evaluation metrics, proving that the proposed framework is more reliable when all EEG feature sets are used as input. LSVM, which delivered an average AC of 99.4%, SE of 98.8%, SP of 100%, F1 of 99.4%, and FDR of 0%, provided the best classification performance of the suggested framework employing the combination of functional connectivity, spectral, and nonlinear feature sets. Performance was nearly identical when functional connectivity was combined with the spectral and nonlinear feature sets. In terms of classification accuracy performance, these combinations come in second. KNN achieved an average AC of 98.8%, SE of 98.7%, SP of 97.5%, F1 of 98.9%, and FDR of 0.6%, which was the best classification performance of the suggested framework employing the combination of functional connectivity and spectral feature sets. KNN also offered the best classification performance for the combination of functional connectivity and nonlinear feature sets, with an average AC of 98.8%, SE of 100.0%, SP of 97.0%, F1 of 99.0%, and FDR of 1.8%. In terms of performance for classification accuracy, these combinations come in second. The suggested framework's functional connectivity and spectral feature set combination produced the best classification results when utilizing KNN, which had an average AC of 98.8%, SE of 98.7%, SP of 97.5%, F1 of 98.9%, and FDR of 0.6%. With an average AC of 98.8%, SE of 100.0%, SP of 97.0%, F1 of 99.0%, and FDR of 1.8%, KNN also offered the greatest classification

performance for the combination of functional connectivity and nonlinear feature sets. These pairings are in second place for categorization accuracy performance. KNN achieved the best classification results of the suggested framework employing the combination of functional connectivity and spectral feature sets, with an average AC of 98.8%, SE of 98.7%, SP of 97.5%, F1 of 98.9%, and FDR of 0.6%. KNN also offered the best classification performance for the combination of functional connectivity and nonlinear feature sets, with an average AC of 98.8%, SE of 100.0%, SP of 97.0%, F1 of 99.0%, and FDR of 1.8%.

3.3. EEG Signal Power Analysis. This section looked into the changes in alpha, theta, beta, and delta EEG band power between MCI and HC samples at various scalp locations. In order to quantify the difference between the two groups by these band powers, the band powers of the aforementioned frequency subbands for MCI and HC cases were studied using the t -test approach. Table 4 provides the t -test results on the alpha, theta, beta, and delta EEG signal powers of MCI and HC cases in each EEG channel. According to these results, alpha and beta band powers provided the most significant difference between MCI and HC cases. Using alpha and beta band powers, the frontal, parietal, and temporal regions of the scalp offered greater discrimination than other regions. Theta power was the next-best EEG band power for MCI and HC discriminating. However, the delta band power was unable to distinguish significantly between the MCI and HC patients. Figure 4 displays the alpha, delta, beta, and theta band powers on the EEG topographic maps of HC and MCI subjects. According to Figure 4, the frontal lobe used alpha band power to significantly distinguish between HC and MCI individuals. The

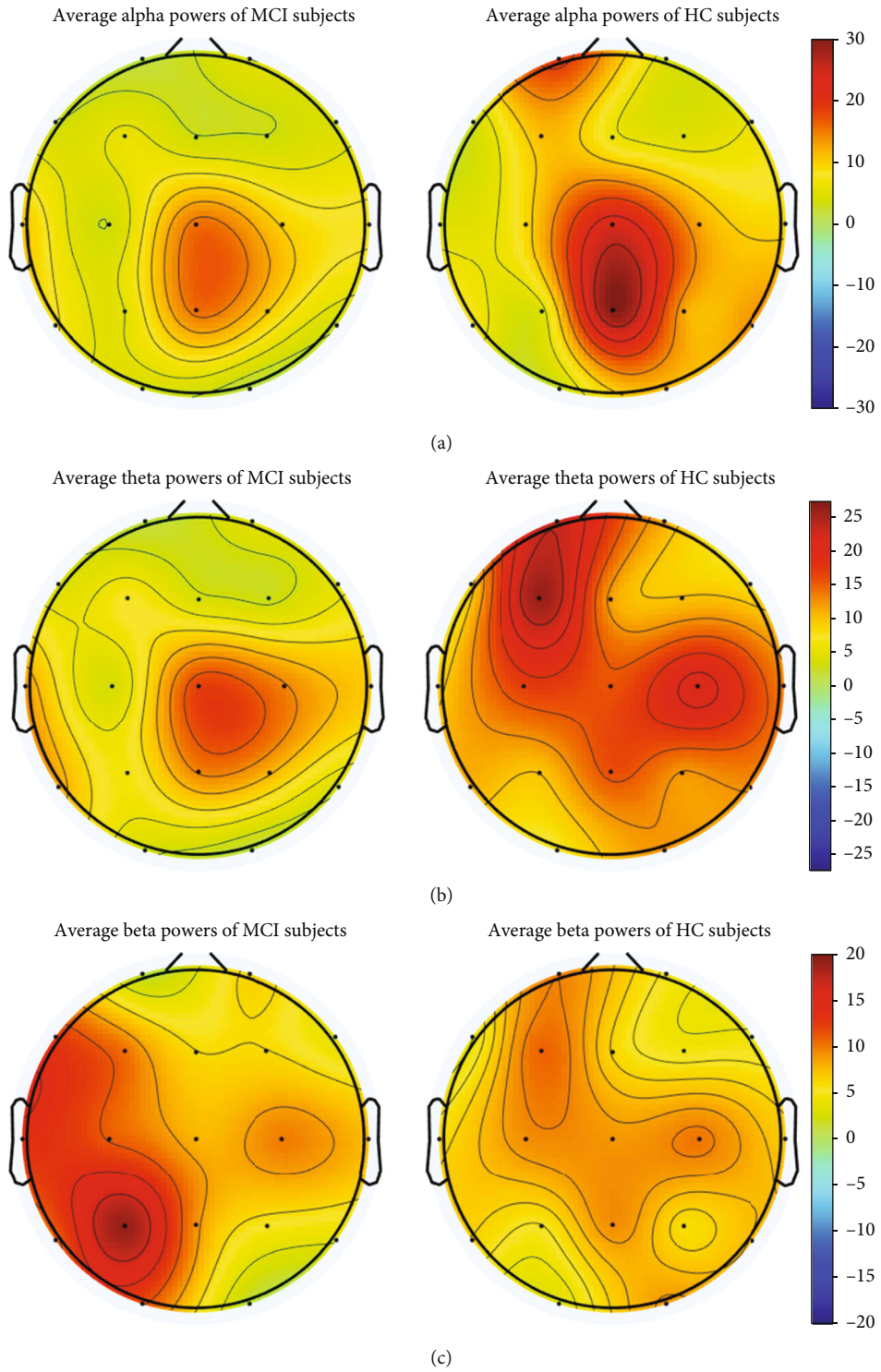


FIGURE 4: Continued.

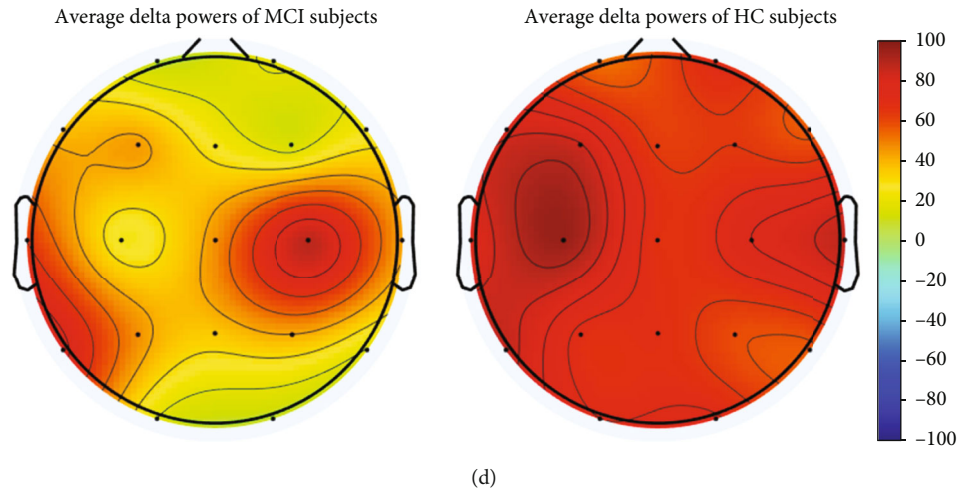


FIGURE 4: The EEG topographic maps of HC and MCI cases in terms of the alpha, delta, beta, and theta band powers.

frontal, temporal, and occipital lobes contributed significantly to the variations between HC and MCI subjects in theta band power. The major differences between HC and MCI subjects employing beta band power were found in the left temporal and occipital areas. The frontal and occipital areas supplied the substantial variations between HC and MCI subjects for delta band power.

3.4. Functional Connectivity Analysis. The major objective of this section is to use the t -test approach to identify the most important functional connectivity coefficients that distinguish between HC and MCI cases. In order to achieve this, the p value metric of the t -test method was used to rank the functional connectivity coefficients of MCI and HC cases and the top 10 values were identified. The top ten functional connectivity coefficients for distinguishing between HC and MCI samples are listed in Table 5 along with associated p value metrics. These findings show that there is a significant difference between the classes utilizing each of the top ten functional connectivity measures, with the statistical significance difference of the ten features between MCI and HC cases being less than $1e-3$. The F3-C3, Fp1-F3, F3-T5, P3-F7, P4-Cz, C3-O1, C3-C4, Fp1-Fp2, C3-Fp2, and Fz-C4 were the top 10 features which provided the most significant difference between HC and MCI classes among functional connectivity coefficients. The boxplot of the top ten functional connectivity features for MCI and HC cases is shown in Figure 5. These boxplots show that there can be large disparities between MCI and HC samples depending on these top ten features. The functional connection between the frontal, temporal, central, and parietal scalp regions is also inferred from these findings to have contributed to the most notable disparities between MCI and HC classes.

3.5. Selected Features. In the evaluation of the proposed framework using 10-fold crossvalidation, the SBFS method returns a specific subset of features as selected features in each iteration, which resulted in 10 subsets of features for all iterations. The intersection of these ten subsets is reported and examined in this subsection. The details of these traits are

TABLE 5: Ten top functional connectivity features in terms of discrimination between HC and MCI classes with their p values of the t -test method.

Functional connectivity feature	t -test p value
F3-C3	$6.46e-08$
Fp1-F3	$1.12e-06$
F3-T5	$3.97e-06$
P3-F7	$2.63e-05$
P4-Cz	$5.09e-05$
C3-O1	$1.18e-04$
C3-C4	$1.22e-04$
Fp1-Fp2	$1.75e-04$
C3-Fp2	$2.34e-04$
Fz-C4	$3.47e-04$

listed in Table 6. These features totaled 361 and were divided into three sets: the spectral set (108), the functional connectivity set (171), and the nonlinear set (82). These results show that all spectral and functional connectivity features were chosen in all iterations of 10-fold crossvalidation and had the biggest proportion of the features chosen, whereas 82 nonlinear characteristics were chosen in all iterations.

3.6. Participant-Independent Evaluation. The suggested framework was validated using the leave-one-participant-out crossvalidation approach in order to be assessed from the standpoint of the participant independence. In this method, the spectral, nonlinear, and functional connectivity properties were retrieved from each participant's EEG signals without segmenting them first. The remaining data were then utilized as the training set, with one data set serving as the testing set. Until each case was used as the testing set, this was reproduced 34 times. The classification model for the proposed framework in this evaluation was an LSVM classifier. The resulting confusion matrix by the suggested framework in this assessment is shown in Table 7. Table 7

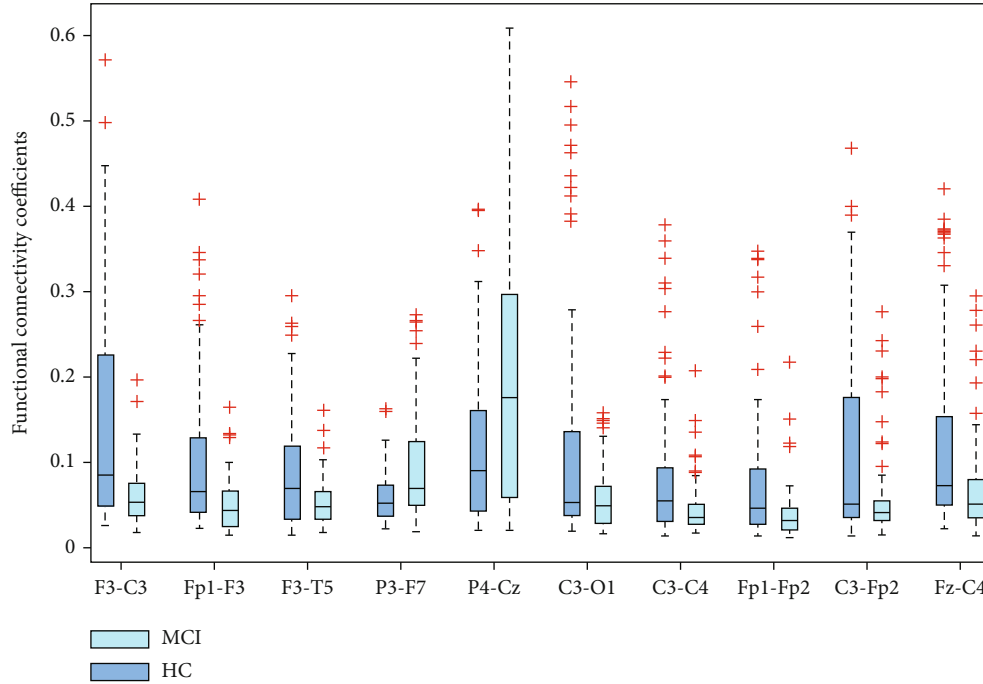


FIGURE 5: The boxplot of ten top functional connectivity features of MCI and HC cases.

TABLE 6: The intersection of the selected features in all iterations of 10-fold crossvalidation.

Feature set	Selected features
Functional connectivity	All features
Spectral	All features
	DFA (Fp1, Fp2, F4, Fz, T4, T6, Pz, O1, O2)
	Higuchi (Fp2, F8, T4, P4, Pz, O2, C4, Cz, O2)
	Correlation dimension (Fp1, Fp2, F3, Fz, T3, T4, T6, P3, C3, C4, O1, O2)
	Lyapunov exponent (F3, F4, Fz, T5, P4, O2, C3, C4, Cz, O1, O2)
Nonlinear	C0-complexity (Fp1, F4, T3, P3, P4, Pz, C4, O1)
	Kolmogorov entropy (Fp1, Fp2, F3, F8, T3, T5, T6, Pz, C4, O1, O2)
	Shannon entropy (Fp1, Fp2, F3, Fz, T3, T6, T4, T5, O1, Cz)
	Approximate entropy (Fp1, Fp2, F4, F3, Fz, F7, T6, P3, P4, O2, C3, Cz)

shows that the suggested framework used an LSVM model to achieve an AC of 91.1%, SE of 88.8%, SP of 93.7%, F1 of 91.4%, and FDR of 5.8%. These findings show that even in more demanding evaluations like the leave-one-participant-out cross-validation strategy, the suggested framework provided an accurate and acceptable classification performance.

4. Discussion

This paper presents a machine learning framework for MCI diagnosis based on EEG signals using spectral, functional

TABLE 7: The obtained confusion matrix by the proposed framework using the leave-one-participant-out cross-validation approach.

Predicted classes	True classes	
	MCI	HC
MCI	16	1
HC	2	15

connectivity, and nonlinear features. The findings demonstrate how well the proposed framework performs accurate and reliable classification. The spectral and functional connectivity feature sets outperformed the nonlinear feature set among the feature sets. This may be due to the fact that nonlinear feature sets are less likely than spectral and functional connectivity feature sets to distinguish between MCI and HC patients. These feature sets may also offer biological explanations for cognitive states. This study used the *t*-test method to examine the band powers of the alpha, delta, beta, and theta frequency subbands in each EEG channel of MCI and HC individuals. This test revealed that the most important distinction between MCI and HC cases was between alpha and beta band powers. Beta band power is more connected to the traits of MCI and other types of dementia than alpha band power. The relationship between beta band power and several MCI-related cognitive processes, including as expectation, consciousness, memory, and problem-solving, may account for the physiological differentiation made by beta band power. Furthermore, the differentiation made by alpha band power may be linked to further MCI-related cognitive states, such as difficulty focusing and extreme relaxation. Additionally, the top 10 functional connectivity characteristics distinguished between the HC and

TABLE 8: Comparison between the proposed framework and previous works for identifying MCI patients based on EEG signals.

Study	Year	EEG features	Classifiers	Reported AC
[19]	2016	Spectral features	NF and KNN	88.8%
[21]	2019	Time series signal spectral and features	LC-KSVD and CLC-KSVD	88.9%
[22]	2019	Time and spectral domain features	LR and SVM	87.9%
[23]	2019	Spectral-temporal features	SVM	96.94%
[24]	2019	Spectral, statistical, and nonlinear features	SVM	96.94%
[26]	2020	AR and PE features	ELM, SVM, and KNN	97.64%
Proposed framework	2021	Spectral, functional connectivity and, Nonlinear features	LINSVM, RBFSVM, and LR, DT, RB, NB, GB, and KNN	99.4%

MCI groups in statistical analysis of the MCI and HC samples were presented. The findings show that the most effective way to distinguish between MCI and HC cases is through the functional connectivity coefficients across the frontal, temporal, central, and parietal scalp areas. These findings may be the result of a correlation between certain MCI-related mental states and cognitive tasks, such as elicitation, attention, thinking, problem-solving, and memory-related processes and the linkages between frontal, temporal, central, and parietal regions.

A comparison of the proposed framework and earlier studies for MCI diagnosis based on EEG data is shown in Table 8. It should be noted that only [19, 21] research utilized the same dataset and evaluation process for their approaches but the content and methodologies of other publications are not the same as our work. The given framework, which has the highest mean of AC among state-of-the-art methods, outperforms the other state-of-the-art strategies for automatic MCI diagnosis based on EEG signals, according to the results described in Table 8. The fundamental improvement of the proposed framework over earlier research is the integration of the spectral, functional connectivity, and nonlinear properties, which have not been combined in such a way in other works. According to the results, the combination of spectral, functional connectivity, and nonlinear features produced the greatest classification results, exceeding earlier studies that predicated MCI diagnosis on EEG data. Another significant distinction between the proposed framework and earlier research is that the provided framework uses the SBFS algorithm as a feature selection strategy, whereas earlier works employed other feature selection approaches including RF and rank-based feature selection methods.

The main restriction is the tiny sample size of the used dataset. As a result, the proposed method's precise and reliable performance could not be highly generalizable. By dividing each EEG signal into five-minute chunks, we did our best to make up for this restriction. However, further MCI EEG datasets are necessary for the applicability of this paradigm and related strategies. On the other hand, there are very few publicly available MCI EEG datasets with more participants. It is advantageous to generalize the validation of the suggested methodologies because public datasets generally have the potential to open up new avenues for collaboration. The framework that is being given also has a heavy

computational load. The method's computing load was raised even though the integration of all feature sets produced the maximum classification accuracy and produced a high-dimensional feature matrix. It further complicates how the framework should be interpreted in terms of physiological and biomarker characteristics. Additionally, it is unclear how clinically applicable this study and all automatic EEG-based MCI detection methods are. More clinical experimental evidence is required to confirm these techniques' clinical efficacy.

However, this paper proposes an automated MCI diagnosis framework based on EEG signals with an accurate and robust classification performance according to the obtained results. It could be developed to use it as a computer-aided diagnosis (CAD) tool for clinical purposes. Future studies can also focus on providing more MCI EEG datasets and implementing deep learning approaches for automatic EEG-based MCI diagnosis.

5. Conclusion

This study provided a spectral, functional connectivity, and nonlinear feature-based machine learning framework for automatic MCI diagnosis. In this framework, SBFS was applied as a feature selection approach to choose the best subset of features and enhance classification performance. Additionally, many categorization models were assessed in order to choose the best one for the suggested framework. The optimal input for the proposed MCI diagnosis framework was also determined by applying each of the feature sets and their combinations to the proposed framework. Based on the obtained results, the LSVM classifier combined with functional connectivity, spectral, and nonlinear feature sets achieved the best classification performance of the proposed framework, which provided an average AC of 99.4%, SE of 98.8%, SP of 100%, F1 of 99.4%, and FDR of 0%. Additionally, the leave-one-participant-out crossvalidation method was used to evaluate the offered framework. The results showed that the LSVM model had an AC of 91.1%, SE of 88.8%, SP of 93.7%, F1 of 91.4%, and FDR of 5.8%. These findings demonstrate how well the newly presented framework performs accurate and reliable classification. The current methodology offered a superior classification performance in terms of robustness and accuracy when compared to earlier research for EEG-based automatic

MCI diagnosis. Based on the results, the medical equipment businesses would be encouraged to create a CAD system for MCI diagnosis utilizing the same proposed framework due to the high potential of the given framework to identify MCI patients. Future research can concentrate on supplying additional MCI EEG datasets and utilizing cutting-edge deep learning techniques for automatic EEG-based MCI diagnosis.

Data Availability

The used dataset in this study is a public dataset. The data statement was checked and it is valid.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] National Collaborating Centre for Mental Health (UK), *Dementia: a NICE-SCIE guideline on supporting people with dementia and their carers in health and social care*, Leicester (UK): British Psychological Society, UK, 2007.
- [2] M. Prince, E. Albanese, M. Guerchet, and M. Prina, "Dementia and risk reduction: an analysis of protective and modifiable factors," *World Alzheimer Report*, pp. 66–83, 2014.
- [3] A Association, "2015 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 11, no. 3, pp. 332–384, 2015.
- [4] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, "Alzheimer disease in the United States (2010–2050) estimated using the 2010 census," *Neurology*, vol. 80, no. 19, pp. 1778–1783, 2013.
- [5] A. Lobo, L. J. Launer, L. Fratiglioni et al., "Prevalence of dementia and major subtypes in europe: a collaborative study of population-based cohorts," *Neurology*, vol. 54, no. 5, p. S4, 2000.
- [6] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, "A major depressive disorder classification framework based on eeg signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis," *Journal of Neuroscience Methods*, vol. 358, p. 109209, 2021.
- [7] E. Aydemir, T. Tuncer, S. Dogan, R. Gururajan, and U. R. Acharya, "Automated major depressive disorder detection using melamine pattern with eeg signals," *Applied Intelligence*, vol. 51, no. 9, pp. 6449–6466, 2021.
- [8] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbf, "Major depressive disorder diagnosis based on effective connectivity in eeg signals: a convolutional neural network and long short-term memory approach," *Cognitive Neurodynamics*, vol. 15, no. 2, pp. 239–252, 2021.
- [9] M. Behnam and H. Pourghassem, "Seizure-specific wavelet (seizlet) design for epileptic seizure detection using correlation ellipse features based on seizure modulus maximas patterns," *Journal of Neuroscience Methods*, vol. 276, pp. 84–107, 2017.
- [10] S. Rukhsar, Y. Khan, O. Farooq, M. Sarfraz, and A. Khan, "Patient-specific epileptic seizure prediction in long-term scalp eeg signal using multivariate statistical process control," *Irbm*, vol. 40, no. 6, pp. 320–331, 2019.
- [11] S. Janjarasjitt, "Spectral exponent characteristics of intracranial eegs for epileptic seizure classification," *IRBM*, vol. 36, no. 1, pp. 33–39, 2015.
- [12] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.
- [13] G. Chandel, P. Upadhyaya, O. Farooq, and Y. Khan, "Detection of seizure event and its onset/offset using orthonormal _triadic_ wavelet based features," *IRBM*, vol. 40, no. 2, pp. 103–112, 2019.
- [14] K. Polat and M. Nour, "Epileptic seizure detection based on new hybrid models with electroencephalogram signals," *IRBM*, vol. 41, no. 6, pp. 331–353, 2020.
- [15] M. Deriche, S. Arafat, S. Al-Insaf, and M. Siddiqui, "Eigen-space time frequency based features for accurate seizure detection from eeg data," *IRBM*, vol. 40, no. 2, pp. 122–132, 2019.
- [16] T. J. Hirschauer, H. Adeli, and J. A. Buford, "Computer-aided diagnosis of Parkinsons disease using enhanced probabilistic neural network," *Journal of Medical Systems*, vol. 39, no. 11, p. 179, 2015.
- [17] R. Yuvaraj, M. Murugappan, U. R. Acharya, H. Adeli, N. M. Ibrahim, and E. Mesquita, "Brain functional connectivity patterns for emotional state classification in Parkinson's disease patients without dementia," *Behavioural Brain Research*, vol. 298, pp. 248–260, 2016.
- [18] M. Shim, H.-J. Hwang, D.-W. Kim, S.-H. Lee, and C.-H. Im, "Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level eeg features," *Schizophrenia Research*, vol. 176, no. 2-3, pp. 314–319, 2016.
- [19] M. Kashefpoor, H. Rabbani, and M. Barekatin, "Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features," *Journal of Medical Signals and Sensors*, vol. 6, no. 1, pp. 25–32, 2016.
- [20] J. C. McBride, X. Zhao, N. B. Munro et al., "Spectral and complexity analysis of scalp eeg characteristics for mild cognitive impairment and early Alzheimer's disease," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 2, pp. 153–163, 2014.
- [21] M. Kashefpoor, H. Rabbani, and M. Barekatin, "Supervised dictionary learning of eeg signals for mild cognitive impairment diagnosis," *Biomedical Signal Processing and Control*, vol. 53, p. 101559, 2019.
- [22] S. Khatun, B. I. Morshed, and G. M. Bidelman, "A single-channel eeg-based approach to detect mild cognitive impairment via speech-evoked brain responses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1063–1070, 2019.
- [23] J. Yin, J. Cao, S. Siuly, and H. Wang, "An integrated mci detection framework based on spectral-temporal analysis," *International Journal of Automation and Computing*, vol. 16, no. 6, pp. 786–799, 2019.
- [24] N. Sharma, M. Kolekar, K. Jha, and Y. Kumar, "Eeg and cognitive biomarkers based mild cognitive impairment diagnosis," *IRBM*, vol. 40, no. 2, pp. 113–121, 2019.
- [25] P. Durongbhan, Y. Zhao, L. Chen et al., "A dementia classification framework using frequency and time-frequency features based on eeg signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 826–835, 2019.
- [26] S. Siuly, Ö. F. Alçin, E. Kabir et al., "A new framework for automatic detection of patients with mild cognitive impairment

- using resting-state eeg signals,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 9, pp. 1966–1976, 2020.
- [27] B. Oltu, M. F. Akşahin, and S. Kibaroglu, “A novel electroencephalography based approach for Alzheimer’s disease and mild cognitive impairment detection,” *Biomedical Signal Processing and Control*, vol. 63, p. 102223, 2021.
- [28] M. S. Safi and S. M. M. Safi, “Early detection of Alzheimer’s disease from EEG signals using Hjorth parameters,” *Biomedical Signal Processing and Control*, vol. 65, p. 102338, 2021.
- [29] A. Delorme and S. Makeig, “Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis,” *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.
- [30] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, “Iclabel: an automated electroencephalographic independent component classifier, dataset, and website,” *NeuroImage*, vol. 198, pp. 181–197, 2019.
- [31] I. Winkler, S. Haufe, and M. Tangermann, “Automatic classification of artifactual ica-components for artifact removal in eeg signals,” *Behavioral and Brain Functions*, vol. 7, no. 1, p. 30, 2011.
- [32] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing: Prentice-Hall Englewood Cliffs*, 1989.
- [33] L. R. Squire, N. Dronkers, and J. Baldo, *Encyclopedia of Neuroscience*, Elsevier, 2009.
- [34] C. J. Stam and B. Van Dijk, “Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets,” *Physica D: Nonlinear Phenomena*, vol. 163, no. 3–4, pp. 236–251, 2002.
- [35] M. Jospin, P. Caminal, E. W. Jensen et al., “Detrended fluctuation analysis of eeg as a measure of depth of anesthesia,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 840–846, 2007.
- [36] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory,” *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.
- [37] P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” in *The Theory of Chaotic Attractors*, pp. 170–189, Springer, 2004.
- [38] J. Röscke, J. Fell, and P. Beckmann, “Nonlinear analysis of sleep eeg data in schizophrenia: calculation of the principal lyapunov exponent,” *Psychiatry Research*, vol. 56, no. 3, pp. 257–269, 1995.
- [39] S. En-hua, C. Zhi-jie, and G. Fan-ji, “Mathematical foundation of a new complexity measure,” *Applied Mathematics and Mechanics*, vol. 26, no. 9, pp. 1188–1196, 2005.
- [40] L. I. Aftanas, N. V. Lotova, V. I. Koshkarov, V. L. Pokrovskaja, S. A. Popov, and V. P. Makhnev, “Non-linear analysis of emotion eeg: calculation of kolmogorov entropy and the principal lyapunov exponent,” *Neuroscience Letters*, vol. 226, no. 1, pp. 13–16, 1997.
- [41] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [42] S. M. Pincus, “Approximate entropy as a measure of system complexity,” *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [43] P. Pudil, J. Novovičová, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.