

Mild cognitive impairment detection with optimally selected EEG channels based on variational mode decomposition and supervised machine learning

Majid Aljalal^{a,*}, Marta Molinas^b, Saeed A. Aldosari^a, Khalil AlSharabi^a, Akram M. Abdurraqeb^a, Fahd A. Alturki^a

^a Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh, Saudi Arabia

^b Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway



ARTICLE INFO

Keywords:

Channel selection
EEG
Entropy
VMD
Machine learning
MCI
Multi-objective optimization
NSGA

ABSTRACT

Detecting mild cognitive impairment (MCI), which is typically the earliest stage of dementia, is essential for managing dementia. Recently, researchers have explored the use of electroencephalography (EEG) to detect MCI. EEG involves recording brain activity by placing a number of electrodes (channels) directly on the scalp. Here we present effective variational mode decomposition (VMD)-based methods for detection of MCI with optimally selected EEG channels. To achieve this, we first decompose the EEG signals from each channel into sub-signals called variational mode functions (VMFs) with different frequency bands. From each VMF, an entropy measure is used to extract a feature. Several supervised machine learning techniques are investigated to classify the resulting features and obtain full-channel classification accuracy. We assess the classification performance of the proposed methods on 19-channel EEGs recorded from 32 healthy subjects and 29 MCI patients in the resting state. In a second step, we explore the feasibility of reducing the number of EEG channels while maintaining classification accuracy by using several EEG channel selection methods. Specifically, we study the effectiveness of non-dominated sorting genetic algorithm (NSGA)-based methods (multi-objective optimization methods), in addition to greedy algorithms such as back-elimination and forward-addition. Results show that selecting a few suitable channels can lead to similar or even superior accuracy compared to a classifier that utilizes all EEG channels. The results also show that the multi-objective NSGA optimization methods have a greater ability than the greedy methods to select few compatible channels. For example, 99.51% is the maximum accuracy achieved using all 19 EEG channels when processed with a combination of VMD, log energy entropy, and a K-nearest neighbor (KNN) classifier. The same accuracy is achieved with only 9 channels selected by NSGA, compared to 13 channels for greedy algorithms. With an ensemble KNN classifier, the highest accuracy, 99.81%, is achieved using 11 channels selected by an NSGA-based method. The results are encouraging, demonstrating that this brain disorder can be accurately diagnosed **using a minimal number of electrodes and paving the way for its adoption** in clinical practice.

1. Introduction

Dementia is a term used to describe a progressive loss of mental abilities such as memory, speech, and cognition that interferes with daily functioning [1]. People over the age of sixty are more susceptible to this condition. The most typical early symptom is difficulty recalling recent events [2]. Mild cognitive impairment (MCI) is the initial phase of Alzheimer's disease (AD) and other dementia types. It leads to cognitive alterations that are observable to the affected individuals or their family members, but do not significantly impair their daily activities [3,4].

Since MCI does not meet the diagnostic criteria for AD or dementia, it does not usually interfere with daily life. Nonetheless, people with MCI are at an increased risk of developing AD or other forms of dementia in the future, with approximately 15–20% of MCI patients progressing to AD each year [5]. While a new drug for AD, Lecanemab [6], has been recently approved by the Food and Drug Administration (FDA), diagnosing AD during the MCI stage can significantly slow down disease progression. This could help AD patients take advantage of future therapies before reaching a more severe stage of this disease. To diagnose MCI/AD, various procedures are often required, including

* Corresponding author.

E-mail address: maljalal@ksu.edu.sa (M. Aljalal).

magnetic resonance imaging (MRI), computed tomography (CT), mini-mental state examinations (MMSE), blood tests, neurological examinations, positron emission tomography (PET), and spinal fluid analyses.

Electroencephalography (EEG) is a non-invasive method for capturing alterations in brain activity (electric potentials) generated by many neurons. These electric potentials are measured by electrodes placed on the scalp. The spatial resolution of EEG is influenced by the number of electrodes used and how they are positioned on the scalp. In contrast to MRI, CT, and PET, EEG recording methods provide good temporal resolution and EEG devices are more affordable and can be portable. There have been many studies on EEG analysis using machine learning techniques to detect various neurological conditions, such as Parkinson's disease (PD), major depressive disorder, autism spectrum disorder (ASD), epilepsy, AD, schizophrenia, and [7–16] and emotion recognition as well [17]. Several studies have also employed EEG signals with machine learning techniques for the automatic detection of MCI using task-state or resting-state. In the case of task-state EEG, the participants are required to do predefined tasks. For example, in [18], experiments with five contrastive speech sounds along vowel continuums were conducted. The vowel continuums were played successively, and the subjects were asked to identify the sounds while their EEG data was gathered from only one channel. In the resting state, individuals' EEGs are recorded when their eyes are closed or open, with no task performed. Recording resting-state EEG data is easier, more practical in real-world circumstances, and more comfortable, especially for the elderly. According to a recent systematic review [19] on the use of resting-state EEG for AD diagnosis and progression evaluation, a sizable body of literature (48 studies) focused on differentiating AD from healthy controls, but the differences between the MCI and health control (HC) have received less attention.

In order to early detect AD in resting-state EEGs, J. Dauwels et al. [20] employed granger causality and stochastic event synchrony measurements as features and linear/quadratic discriminant analysis (L/QDA) as classifiers. They used data from 22 MCI and 38 HC participants with 21 channels to evaluate their approach, and the results showed an 83% classification accuracy. M. Kashefpoor et al. [21] proposed a technique to differentiate between individuals with MCI and healthy controls using basic spectral EEG data. The authors utilized a neuro-fuzzy algorithm and a K-nearest neighbor (KNN) classifier in combination to categorize selected features from 11 MCI patients and 16 healthy controls. The EEGs from 19 channels were segmented for 1 sec, with the segments then being 50% overlapped to achieve an accuracy of 88.89%. Later, the same group [22] increased the dataset by using 29 MCI and 32 HC and proposed an analysis of EEG signals based on supervised dictionary learning, called correlation-based Label consistent K-SVD (CLC-KSVD). After implementing single-channel classification and brain-region-based classification, an accuracy of 80% was obtained based on the left-temporal region that includes F7, T3, and T5 channels. By applying their proposed method to the data in [21], an accuracy of 89% was obtained for the same brain region. The same dataset studied in [21] was considered by S. Hadiyoso et al. [23], who used a KNN classifier with power spectral features to reach a classification accuracy of 81.5%. Y.T. Hsiao et al. [24] proposed relative power-based features (KERP) for categorizing MCI and healthy controls. In this study, 30-channel recordings of EEGs from 24 MCI and 27 HC subjects were divided into 3-sec segments, and the segments were then 50% overlapped. Before the classification, Fisher's method was used to select the extracted features. The selected KERP features were classified using a support vector machine (SVM) with leave-one-subject-out (LOSO) cross-validation (CV) technique, which yielded an accuracy of 90.20%. J. Yin et al. [25] used the same dataset as in [21], but they removed five HC instances to balance the proportion of MCI patients and HC cases. The authors applied stationary wavelet transformation (SWT) to improve the signal to noise ratio, and then extracted nine statistical features including median, standard deviation, mean, and mode. The SVM classifier achieved an accuracy of 96.94% when the feature vectors from all MCI and

HC are divided to 60% for training, 20% for validation, and 20% for testing. S. Siuly et al. [26] extracted features from the same dataset in [21] using auto-regressive and permutation entropy models, which were then categorized by an extreme learning machine (ELM) resulting in a 10-fold CV accuracy of 98.78%. A recent study by A.M. Alvi et al. [27] employed a deep learning-based framework, specifically the Long Short-Term Memory (LSTM) model, to differentiate between MCI and healthy controls. The authors generated 20 alternative LSTM models and evaluated their performance on the same dataset as used in [21], ultimately identifying the best model with 96.41% accuracy using 5-fold CV. K. Lee et al. [28] extracted several features for MCI detection, including absolute and relative power spectrum density (PSD), differential and rational asymmetry, phase-amplitude coupling, Shannon entropy, Hjorth parameters, Lyapunov exponent, Hurst exponent, and Kolmogorov complexity. The total number of features extracted from the 32 channels was 1500. The authors used their own dataset of 21 MCI and 21 HC participants. For channel selection, the authors evaluated the classification accuracies for symmetric two-, four-, six-, and eight-electrode combinations using an SVM classifier. The highest classification accuracies of 74.04%, 82.43%, 86.28%, and 86.85% were achieved for the two-, four-, six-, and eight-electrode configurations, respectively, with leave-pair-subject-out (LPSO) CV. Also more recently, R.A. Movahed et al. [29] extracted 425 features (spectral, functional connectivity, and nonlinear features) from EEGs of 18 MCI and 16 HC participants recorded using 19 channels. Using the combination of all feature sets, linear SVM achieved an accuracy of 99.4% with 10-fold CV.

Numerous studies have looked into how to discriminate between AD, MCI, and HC. For instance, N. Sharma et al. [30] used an SVM classifier to classify 44 people (15 dementia, 16 MCI, and 13 HC), under four conditions: eye-open, eye-close, finger tapping test, and continuous performance test, to explore the categorization of MCI, dementia, and HC. The study analyzed a total of eight EEG features obtained from 21 channels. These features included PSD, kurtosis, spectral kurtosis, spectral entropy, spectral crest factor, skewness, spectral skewness, and fractal dimension. With 10-fold CV and an open-eye resting state, 84.1% accuracy was achieved when diagnosing MCI from HC in this study. With a decision tree (DT) classifier, G. Fiscon et al. [31] evaluated the use of Fourier and wavelet transforms on 109 individuals (19-channels EEG signals from 49 AD, 37 MCI, and 23 HC). For the MCI vs. HC classification task, a combination of discrete wavelet transform (DWT) and DT yielded an accuracy of 83.3% with holdout validation and 93.3% with 10-fold CV. Recently, B. Oltu et al. [32] employed DWT, PSD, and interhemispheric coherence to extract features from 16 MCI, 8 AD, and 11 HC. Using a bagged tree classifier and 5-fold CV, an accuracy of 95.50% was achieved. In a more recent study, D. Pirrone et al. [33] estimated power intensity for each high- and low-frequency band of EEGs recorded from 105 subjects (19-channels EEG signals from 48 AD, 37 MCI, and 20 HC). Using SVM, DT, and KNN classifiers, they studied several classification problems. KNN achieved a 95% accuracy in the MCI vs. HC classification problem with 10-fold CV.

It is important to mention that the classification accuracy can be considerably improved through careful design of the feature extraction stage. In addition, the accuracy can be improved by selecting the most relevant channels. With the exception of [28], current state-of-the-art efforts focus on enhancing MCI classification accuracy through the feature extraction stage but ignore the investigation of the optimal channel combination for MCI detection using resting-state EEG. Although [28] is a good step forward in EEG channel selection, the study only considers the two-, four-, six-, and eight-channels under the restrictive condition of symmetric channel pairs. For example, in the case of evaluating the accuracies of two channels, the Fp1 channel is selected only with the Fp2 channel, the F7 channel is selected with only the F8, the O1 channel is selected with only the O2, and so on. This strategy ignores a lot of two-channel subsets that may lead to higher accuracy. This also applies to four-, six-, and eight-channel combinations. The authors of [28] used this strategy, symmetric channel pairs,

because, of course, it would be very difficult to evaluate all channel subsets of two, four, six, or eight channels. In our study, we consider the entire search space and propose to use heuristic optimization methods to select the optimal channel subsets. Accordingly, the contributions of the present study are presented in two parts. In the first part, we provide new and effective variational mode decomposition (VMD)-based approaches for detecting MCI from resting-state EEG signals. To the best of our knowledge, no study has used the VMD to extract the features of EEGs for diagnosing MCI. In the second part, we explore the use of various channel reduction methods with the aim of maintaining or even exceeding the full-channel classification accuracy. To extract features, the study combined VMD with sure entropy, log energy entropy, Shannon entropy, threshold entropy, or norm entropy. This combination produced effective biomarkers for MCI. The resulting features were then classified using various machine learning techniques, including LDA, QDA, SVM, KNN, and Random Forest (RF). Regarding EEG channel selection, we apply greedy algorithms such as back-elimination (BE) and forward-addition (FA). In addition, multi-objective optimization methods (non-dominated sorting genetic algorithm II (NSGA-II) and referenced-point NSGA-II (R-NSGA-II)) are also used for selecting EEG channels as well as classifier parameters. In the case of using NSGA-II and R-NSGA-II, the obtained feature vectors are repeatedly tested for solving two unconstrained objectives to maximize classification accuracy while minimizing the number of EEG channels required for MCI classification. As in [22], we evaluate the performance of the proposed methods on the public dataset used in [21,23,25–27] (11 MCI and 16 HC), combined with the public dataset used in [29] (18 MCI and 16 HC), making the number of subjects balanced and greater.

The remainder of this article is organized as follows: [Section 2](#) provides a detailed description of the EEG data used, and the methods employed for pre-processing, feature extraction, classification, and EEG channel selection. [Sections 3](#) presents the results and corresponding discussions. Finally, [Section 4](#) concludes the article with recommendations for future research.

Abbreviation	Definition
AD	Alzheimer's disease
ASD	Autism spectrum disorder
BE	Back-elimination
BPF	Band-pass filter
CA	Classification accuracy
CLC	Correlation-based Label Consistent
CT	Computed tomography
CV	Cross-validation
DA	Discriminant analysis
DT	Decision tree
DWT	Discrete wavelet transform
EEG	Electroencephalography
ELM	Extreme learning machine
EMD	Empirical mode decomposition
ENKNN	Ensemble k-nearest neighbor
FA	Forward-addition
FDA	Food and Drug Administration
FE	Feature extraction
Full-channel acc.	Classification accuracy when all channels are used
GA	Genetic algorithm
HC	Healthy controls
HHT	Hilbert-Huang transform
KERP	kernel eigen-relative-power
KNN	K-nearest neighbor
K-SVD	K-singular value decomposition
LDA	Linear discriminant analysis
LogEn	Log energy entropy
LOSO	Leave-one-subject-out
LPSO	Leave-pair-subject-out
LSTM	Long short-term memory
MCI	Mild cognitive impairment
MMSE	Mini-mental state examinations
MOOP	Multi-objective optimization problem
MRI	Magnetic resonance imaging
No_ch	Number of channels
NoEn	Norm entropy

(continued on next column)

(continued)

Abbreviation	Definition
NSGA	Non-dominated sorting genetic algorithm
NSGA-II	Second version of NSGA
NUCOG	Neuropsychiatry Unit Cognitive Assessment Tool
Param	Classifier parameter
PD	Parkinson's disease
PET	Positron emission tomography
PSD	Power spectral density
QDA	Quadratic discriminant analysis
R-NSGA-II	Referenced-point NSGA-II
RF	Random forest
ShEn	Shannon entropy
SuEn	Sure entropy
SVM	Support vector machine
SWT	Stationary wavelet transformation
ThEn	Threshold entropy
VMD	Variational mode decomposition
VMF	Variational mode function

2. Methods

[Fig. 1](#) provides a broad outline of the various stages involved in processing EEG data and selecting channels. The initial stage involves pre-processing of the raw EEG dataset to eliminate any artifacts. In this stage, a band-pass filter is employed to isolate the frequency range of interest. The EEG signals are then divided into equal-length segments without overlapping. During the feature extraction stage, each segment signal from each channel is decomposed into variational mode functions (VMFs) with distinct frequency bands using VMD (more details in [Fig. 4](#)). One feature is extracted from each VMF by applying one of the following entropy measures: sure entropy, log energy entropy, Shannon entropy, threshold entropy, or norm entropy. The collection of feature values from all channels within a segment constitutes a feature vector. The remaining feature vectors are built from the other segments using the same process. The classification can then be done either on all features corresponding to all EEG channels or a subset of them when channel selection is implemented. For EEG channel selection, back-elimination, forward-addition, NSGA-II, or R-NSGA-II are employed. In the classification stage, the feature vectors from the selected channels are classified using RF, L/QDA, SVM, or KNN, to distinguish MCI feature vectors from HC ones. Each stage is covered in more detail in the subsequent subsections.

2.1. Data description and pre-processing

The suggested methods are examined with [two public datasets that are both available in \[34\]](#). The first includes 11 MCI and 16 HC cases, while the second includes 18 MCI and 16 HC. Similar to [22], to create a larger and more balanced dataset, we merged the two datasets. The combined dataset has 61 people over the age of 55, including 29 amnestic MCI patients and 32 normal cases. All participants had finished at least primary school. Subject selection, cognitive tests, EEG recordings, and other experimental procedures were handled by Noor Hospital in Isfahan, Iran. Individuals with dementia, severe physical issues, substance abuse, brain traumas, or serious mental disorders were barred from participating in the study. Prior to participation, each subject was given a thorough description of the research protocol and supplied informed consent. A detailed neuropsychiatric interview was used to assess each individual using Peterson's MCI criteria [35]. MCI was defined as Mini-Mental State Examination (MMSE) scores between 21 and 26, whereas normal was defined as scores over 27. The Neuropsychiatry Unit Cognitive Assessment Tool (NUCOG) was also used to diagnose MCI, with scores ranging from 75 to 86.5 [36]. The mean and standard deviation for the subjects' age, education, and results from the psychiatric test are displayed in [Table 1](#). During a resting state, all participants underwent EEG recordings while lying down in a quiet

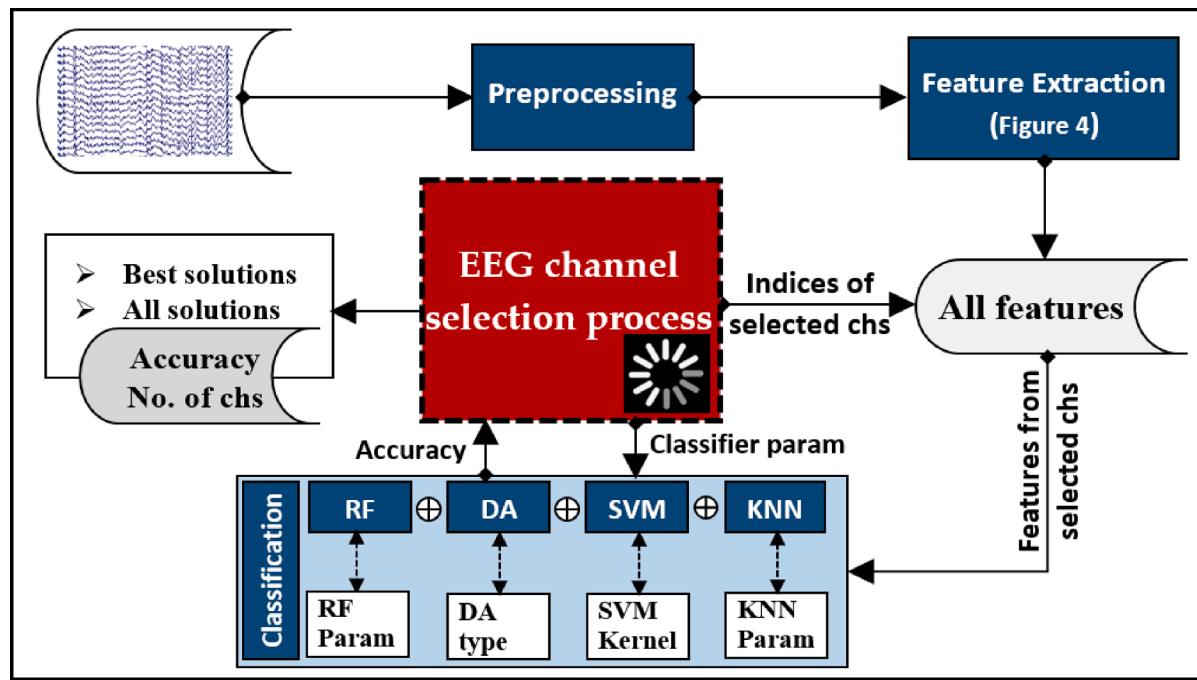


Fig. 1. High-level overview of EEG signal processing and channel selection process.

Table 1
Subject's demographics.

Characteristic	HC (mean \pm st)	MCI (mean \pm st)
No. of cases	32	29
Ages (years)	63.8 ± 4.3	65.7 ± 4.9
Education (years)	8.7 ± 2.3	8.3 ± 1.8
MMSE scores	28.8 ± 0.9	26.9 ± 0.7
NUCOG scores	92.5 ± 3.1	81.5 ± 2.4

room with their eyes closed. Following the 10–20 International System, 19 EEG electrodes were placed. The locations of the 19 channels in this study are depicted in Fig. 2. For EEG recording, a 32-channel digital EEG

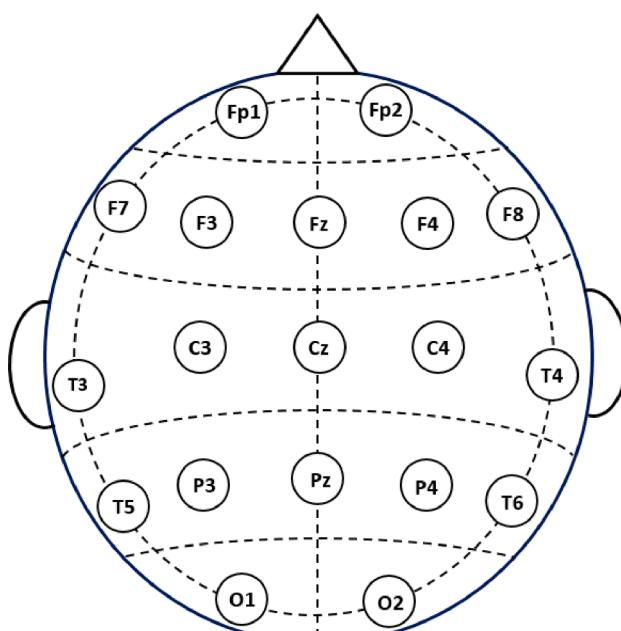


Fig. 2. The 19 channels locations used in this study.

device (Galileo NT, EBneuro, Italy) was used, with electrode-to-skin resistance of less than 5 kΩ. Each subject's EEG data was recorded for a duration of 30 min [22]. To avoid participant fatigue resulting from prolonged recording, only the first 10 min of the EEG recordings were utilized in our study. Fig. 3 displays the electrode maps and PSD of the HC and MCI EEGs. Additionally, the electrode maps are offered for four additional arbitrary frequencies: 2, 6, 10, and 22 Hz. In general, in both cases, the low-frequency spectrum has a higher power density than the high-frequency spectrum.

Each EEG recording is filtered using a fifth-order band-pass Butterworth filter. The frequency range between 0.5 and 32 Hz, where the highest MCI and HC signals are present, is chosen in order to remove baseline artifacts and AC power-line noise seen in Fig. 3. For additional preprocessing, the EEG data are split into non-overlapping segments with a size of $ch \times N$, where ch denotes the number of channels and N is the number of EEG samples per channel at a given time period T . In this study, a 10-second time window is used. The possibility of reducing the number of channels, ch , while improving, or at the very least maintaining, the classification accuracy of MCI vs. HC will be investigated.

2.2. Feature extraction (FE)

2.2.1. Variational mode decomposition

Variational mode decomposition or VMD is a modern signal decomposition technique, first proposed by Dragomiretskiy and Zosso [37]. VMD decomposes a signal $y(t)$ into N mode signals, called variational mode functions (VMFs), and one residual signal. Compared to other decomposition techniques such as empirical mode decomposition (EMD), VMD is more robust to noise and sampling, allowing for the accurate reconstruction of the original signal from its decomposed modes. Among the improvements that VMD brings is that the resulting sub-signals (or VMFs) have non-overlapped frequency bands [37]. The fundamental characteristic of VMFs in VMD is that each of them have cosine wave shapes, slowly varying and positive envelopes, and non-decreasing instantaneous frequency patterns. To decompose a signal using VMD, several processes are involved, including Wiener filtering, Hilbert transform, frequency mixing, and heterodyne demodulation. The primary objective is to identify a discrete set of VMFs and

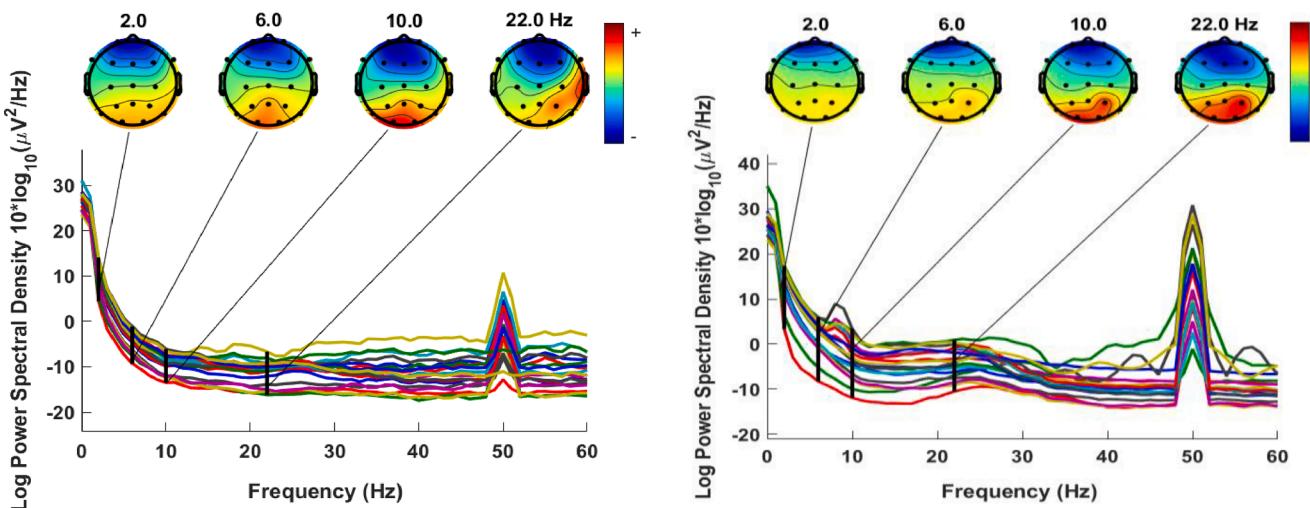


Fig. 3. PSD and electrodes map for HC EEG (left) and MCI EEG (right).

their respective central frequencies that minimize the constrained variational problem, as defined by the following equation;

$$\min(VMF_i, w_i) \left\{ \sum_i^N \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * VMF_i(t)] e^{jw_i t} \right\|_2^2 \right\} \quad (1)$$

In (1), $\delta(t)$ represents the Dirac distribution, $\{VMF_i\} = \{VMF_1, VMF_2, \dots, VMF_N\}$ denotes the complete set of modes, $\{w_i\} = \{w_1, w_2, \dots, w_N\}$ are the central frequencies of all modes, $(\delta(t) + \frac{j}{\pi t}) * VMF_i(t)$ represents the Hilbert transform for $VMF_i(t)$, and the symbol $*$ denotes convolution. For more details on the VMD algorithm, we refer readers to [37].

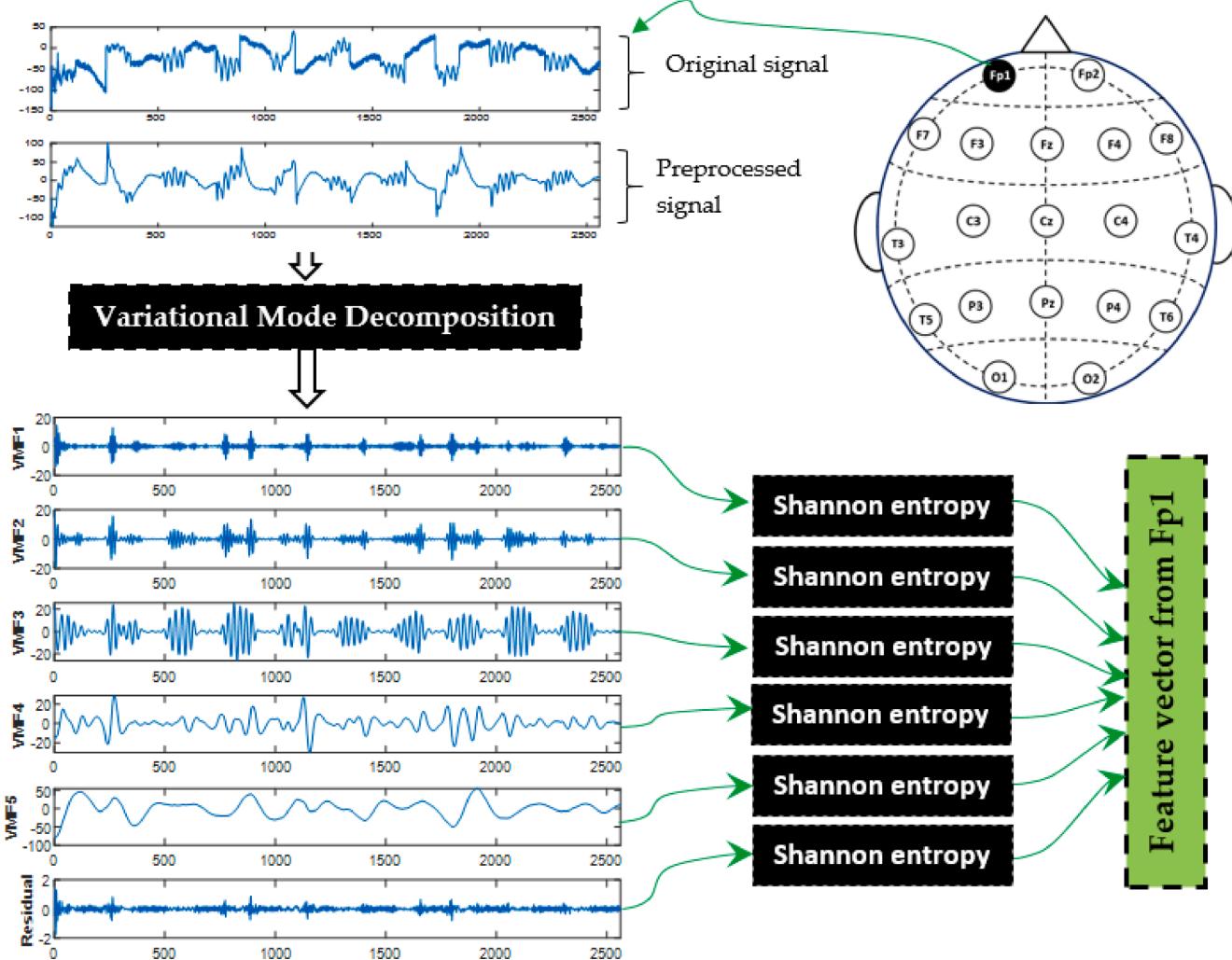


Fig. 4. Example of feature extraction from a 10 sec segment using VMD and Shannon entropy.

Delta (<4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz) are the five sub-bands of EEG signals [38]. As a result, in this study, the number of VMFs is adjusted to 5 in order to acquire VMF central frequencies relating to these sub-bands. Fig. 4 depicts a feature extraction example from a 10-second section using VMD. VMD first decomposes each pre-processed-segmented signal from each channel into five VMFs and one residual. According to Hilbert-Huang transform, the central frequencies of the produced modes VMF1, VMF2, VMF3, VMF4, and VMF5 are 27.20 Hz, 17.19 Hz, 9.28 Hz, 3.41 Hz, and 1.1 Hz, respectively, which belong to different EEG sub-bands. The modes included in beta, alpha and theta (VMF1, VMF2 and VMF3) are used for feature extraction while modes included in delta band (VMF4 and VMF5) are excluded to avoid on the baseline noise. Because residual signal still contains some details of the original signal (see Fig. 4), it is also included for feature extraction.

2.2.2. Features

As shown in Fig. 4, the next step to extract features is computing entropy from the resulting VMFs and residual signals. Entropy is a standard measure of complexity, regularity, and statistical quantification of time series. Several research have demonstrated the efficacy of entropy in analyzing and discovering biomarkers for various mental disorders, such as PD [7,8], epilepsy [39], attention deficit hyperactivity disorder [40], and ASD [41]. This has inspired us to employ entropy as a feature for MCI detection. Rather than directly extracting EEG features, we suggest computing the entropy of the generated VMFs, which might aid in developing valuable biomarkers for identifying MCI.

Several entropy measures are examined: Shannon entropy, log energy entropy, norm entropy, sure entropy, and threshold entropy. These measures are defined as follows. If x_i is the probability frequency of the i^{th} unique value and k is the number of unique values in the discrete signal $S(n)$, then the entropy is defined by [42]:

Shannon entropy (ShEn);

$$\text{ShEn} = \sum_{i=1}^k |x_i|^2 \log|x_i|^2 \quad (2)$$

Log energy entropy (LogEn);

$$\text{LogEn} = \sum_{i=1}^k \log|x_i|^2 \quad (3)$$

Norm entropy (NoEn);

$$\text{NoEn} = \sum_{i=1}^k |x_i|^p, \quad (4)$$

where p is the power of entropy. According to [42], the value of p must be greater or equal to 1.

Sure entropy (SuEn);

$$\text{SuEn} = k - \#\{i \text{ such that } |x_i| \leq \ell\} + \sum_i \min(x_i^2, \ell^2), \quad (5)$$

where ℓ is the threshold value, and usually $\ell > 2$, according to [42].

Threshold entropy (ThEn);

$$\text{ThEn} = \#\{i \text{ such that } |x_i| > \alpha\}, \quad (6)$$

where ThEn is the number of time instants when the signal is greater than a threshold α . According to [42], the threshold should be set to a value less than 1. After experimental fine-tuning, we have discovered that setting $p = 1.1$, $\ell = 3$, $\alpha = 0.2$ leads to improved results in terms of accuracy.

One of these entropy measures (Eq. 2–6) is used to compute four features: three features from the VMFs (VMF1, VMF2, and VMF3) while the fourth feature is from the residual signal. In other words, four feature values are extracted from each channel segment (see Fig. 4). If the number of channels is ch , then for each segment, a $4 \times ch$ feature vector is extracted. The procedure depicted in Fig. 4 is replicated across all segments of the data to obtain a complete set of feature vectors (in the

form of a feature matrix) that represent both MCI and HC cases. Classification is then implemented on the resulting feature matrix.

2.3. Classification and performance evaluation

Several machine learning techniques exist for classifying EEG data. Choosing a suitable classifier can be a complex task that relies on diverse factors such as the dataset size and complexity, computational complexity, and desired accuracy. Computational complexity of a machine learning algorithm is influenced by several factors, such as input data size, model complexity, feature dimensionality, and hyperparameters selected [43]. Classification models can be broadly categorized into linear and nonlinear models. Linear classifiers usually require fewer parameter adjustments when compared to nonlinear classifiers, making them more robust and less susceptible to overfitting. Nonetheless, in certain scenarios, such as those involving extensive or intricate datasets, nonlinear classifiers may provide better outcomes [38]. The present study examines the effectiveness of L/QDA, RF, SVM, and KNN, which are among the most commonly used machine learning techniques for EEG data classification, in differentiating between MCI and HC features. LDA is a binary classification method that distinguishes between individual classes by utilizing mean vectors and covariance matrices of feature vectors. It employs a hyperplane to reduce variance within a class while increasing variance across classes [44]. LDA is a favorable option due to its satisfactory performance, low computational cost, and lack of necessity for substantial pretraining. However, because of its linearity, it may perform poorly when dealing with large nonlinear EEG data. QDA extends LDA by enabling non-linear separation between classes. QDA generates a quadratic decision boundary, while LDA produces a linear decision boundary [45]. LDA's linear boundary may be advantageous when the number of observations is lesser than the number of dimensions, but it may not perform well when the model's assumptions are violated. On the other hand, QDA can capture complex decision boundaries and may deliver better results when the variance of data is non-uniform across classes [45]. Regarding the SVM, given a set of training data, the SVM develops a model or hyperplane that maximizes the distance between patterns from distinct classes [46]. SVMs use a kernel trick to address nonlinear classification problems, which maps the original data or feature space into a different higher dimensional space, making it easier to solve the classification problem. KNN, like other classifiers, can distinguish between two or more patterns. KNN is a simple classification technique that finds the K nearest data points to a test point (a test feature vector) using a distance metric and then classifies that test point based on the majority label of its K nearest neighbors using the distance metric. KNN is a basic algorithm with only one parameter. Any distance metric, including those established by the user, can be entered. However, for large datasets, KNN is computationally expensive and consumes a lot of memory [47]. RF is an ensemble learning method that, during training, builds several decision trees and outputs the class that is the mode of the classes of the individual trees. Although RF is effective in high-dimensional domains and huge datasets, it is slow to analyses and may overfit noisy data [48]. An additional ensemble-based method, ensemble KNN (ENKNN), is investigated in this paper. It mixes numerous KNN models to improve classification performance, usually using a weighted voting mechanism that favors more accurate models. A comprehensive discussion of those classification methods is available in [44–48]. For a more comprehensive discussion of these classification methods, including additional references, readers can refer to [49], which presents a review of supervised machine learning classification algorithms, and [43], which discusses their computational complexity.

In this study, the aim of investigating those classifiers is to compare different types of classifiers and determine which one provides better results. The selection of classifier parameters is optimized using NSGA. In other words, NSGA works by treating the hyperparameters of the classifier as the parameters to be optimized. In the case of RF, different

tree depths (1, 2, ..., 35) are investigated. For SVM, three kernels—linear, polynomial, and radial basis functions—are investigated using NSGA. KNN is tested with 1–10 neighbors. Table 2 summarizes the hyperparameters of each classifier in the present study and clarifies which ones are optimized. For evaluating the performance of each classifier, we employ 10-fold CV. This method randomly splits the dataset into ten equal subsets, with one subset used for validation (test) and the other nine for training [50]. The subset assignment (validation/testing) is changed 10 times (10-fold). Sensitivity, specificity, classification accuracy (CA), and F-score are computed using Equations (7–10) for each time (round), and the results are then averaged over the ten rounds.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (8)$$

$$CA = \frac{N_{\text{correct}}}{N_{\text{total}}} \times 100\% \quad (9)$$

$$F\text{-score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \times 100\% \quad (10)$$

where TP stands for true positives, TN for true negatives, FN for false negatives, and FP for false positives, N_{correct} represents the number of feature vectors that are correctly classified, and N_{total} represents the total number of feature vectors that need to be classified. Sensitivity measures the ability of a classifier to accurately identify individuals with the condition, while specificity measures the ability of the classifier to correctly identify individuals without the condition [51]. The number of accurate positive predictions made is measured by the *precision* metric, which is defined as

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (11)$$

2.4. EEG channel selection and reduction

It is known that more data is not necessarily more information, so some channels may contain redundant, useless, or suboptimal information [52]. The primary goals of using channel selection are to decrease the computational cost of EEG signal processing, shorten the preparation period, and improve classification accuracy by decreasing the possibility of over-fitting that occurs when all channels are used. To this end, two greedy methods and two multi-objective optimization methods are employed for selecting EEG channels. In this study, like in [28], the channel selection is accomplished based on a classifier-specific strategy in which the classifier's training and test processes are involved and evaluated using 10-fold cross-validation throughout the entire selection process.

2.4.1. Greedy algorithms

Using greedy algorithms is an easy and quick way to implement [53]. Two greedy strategies are described for selecting EEG channels in the next subsections.

Table 2
Classifiers' hyperparameters.

Classifier	Hyperparameters
SVM	Kernel type = (optimized: linear, polynomial, or radial basis function), method='least squares', C = 2e-1
KNN/ ENKNN	No. of neighbors = (optimized), distance = 'Euclidean', rule='nearest'
RF	Learner type = 'decision tree', ensemble = 'bag', tree depths = (optimized: 1 to 35)
DA	Discriminant type = (optimized: linear, or quadratic)

2.4.1.1. Back-Elimination algorithm (BE). BE, a well-known greedy algorithm, has been applied to the selection of feature subsets [54,55]. BE is used in the present study to choose the optimum channel combination at each stage (iteration) from a pool of 19 channels. BE requires 19 iterations to achieve this goal. In the first iteration, the method initially removes one of the 19 channels before using the remaining 18 channels to determine the classification accuracy $Acc_{1,1}$. Then the removed channel is placed back with channels group while another channel is removed. The classification accuracy of the resulting channel group $Acc_{1,2}$ is computed. This process is repeated by going over all 19 channels resulting in a set of accuracy values $Acc_{1,1}, Acc_{1,2}, \dots, Acc_{1,19}$. The highest accuracy $MaxAcc_1 = \max(Acc_{1,1}, Acc_{1,2}, \dots, Acc_{1,19})$, is preserved along with its corresponding subset of channels (first local optimum), $ch_selected_1$, that will be used in the next iteration. Note that the subset $ch_selected_1$ contains 18 out of the original 19 channels. In the second iteration, the whole process is repeated, obtaining $Acc_{2,1}, Acc_{2,2}, \dots, Acc_{2,18}$, $MaxAcc_2$, $ch_selected_2$ (second local optimum). The same process is repeated such that one channel is removed at each iteration. The number of selected channels is reduced to one during the 19th iteration. The BE algorithm produces two vectors as outputs. The first vector consists of the maximum accuracies, denoted as $[MaxAcc_1, MaxAcc_2, \dots, MaxAcc_{19}]$. The second vector contains the corresponding subset of channels selected by the algorithm, represented as $[ch_selected_1, ch_selected_2, \dots, ch_selected_{19}]$. The classification procedure is run $19 \times 20 / 2 = 190$ times for the 19 channels.

2.4.1.2. Forward-Addition algorithm (FA). FA works in a similar manner to BE but with a reversed channel selection process, where the size of selected channel subset is increased at every iteration. During the first iteration, the classification accuracy is computed for each channel individually. The channel with the highest accuracy, denoted as $MaxAcc_1 = \max(Acc_{1,2}, Acc_{1,3}, \dots, Acc_{1,19})$, is selected along with its corresponding channel (the first local optimum), represented as $ch_selected_1$. In the following iteration, one of the remaining 18 channels is added to $ch_selected_1$ to form a subset of two channels, and the classification accuracy of this subset is computed. The process is repeated by going over all 18 channels to obtain the accuracy using each subset, $Acc_{2,1}, Acc_{2,2}, \dots, Acc_{2,18}$. The highest accuracy $MaxAcc_2 = \max(Acc_{2,1}, Acc_{2,2}, \dots, Acc_{2,18})$, is preserved along with its corresponding channels subset (second local optimum,), $ch_selected_2$. The FA algorithm repeats this process for 19 iterations, with one channel added at each iteration. By the 19th iteration, the number of selected channels becomes 19. As with BE, the final outputs of the FA algorithm consist of two vectors: the first vector contains the maximum accuracies, $[MaxAcc_1, MaxAcc_2, \dots, MaxAcc_{19}]$, while the second vector includes the corresponding subset of channels, $[ch_selected_1, ch_selected_2, \dots, ch_selected_{19}]$. The classification procedure is run 190 times, like BE.

2.4.2. Multi-objective optimization algorithms

Simultaneous optimization (maximization or minimization) of two or more objective functions is required for multi-objective optimization problems (MOOP). Because the objective functions are incompatible with one another, MOOP can only improve one goal possibly at the expense of another. A MOOP may have no restrictions or it may contain a set of constraints that a solution must satisfy. The best outcome (minimum or maximum) for the objective functions indicates that a solution is optimal and is therefore viable if it satisfies all of the restrictions [56]. In this study, two NSGA methods are used to solve our multi-objective problem and evaluate the outcomes using different feature extraction and classification techniques. These two algorithms, along with the problem, are briefly discussed in the following subsections.

2.4.2.1. NSGA-II and R-NSGA-II. Charles Darwin's theory of natural evolution is the source of inspiration for a search heuristic known as a

genetic algorithm (GA). It is generally applied to complex optimization and search issues. The population for GAs consists of a collection of possible solutions (chromosomes), and each solution or chromosome is made up of a set of parameters (variables) called genes [57]. To produce the best results, GAs frequently use population initialization, fitness function computation, crossover, mutation, survivor selection, and termination criteria [58].

A Pareto-optimal solution, sometimes referred to as a non-dominated solution, is one that outperforms all other solutions (dominated solutions) in a multi-objective optimization problem. In the first version of NSGA, stable sub-populations of good points (Pareto-front) are maintained using a niche method, where good candidates are highlighted using a non-dominated sorting ranking selection strategy [59]. The second edition of NSGA (NSGA-II) was introduced to mitigate a number of issues in the first version, such as population diversity, non-elitist methodology, and computational complexity. The computational cost in NSGA-II has been decreased from $O(PM^3)$ to $O(PM^2)$, where P and M are the population size and the number of objectives, respectively [60]. The elitism technique was also introduced by contrasting the current population with the previously discovered best non-dominated solutions [60]. The NSGA-II elitism does not necessitate the establishment of any new parameters in addition to the conventional genetic algorithm parameters such as population size, termination parameter, crossover and mutation probability.

Reference point NSGA-II (R-NSGA-II), a modified version of NSGA-II, is effective at solving 2-to 10-objective optimization problems [61]. R-NSGA-II provides a set of Pareto-optimal solutions near a defined set of reference points. To guarantee a well-distributed solution set in the front, a predetermined set of reference points is used. The reference point approach is a common methodology in multi-criterion decision-making in which one or more reference (goal) points are specified by the decision-maker beforehand. The target in such an optimization task is then to identify the Pareto-optimal region closest to the reference points. In the present study, three reference points, (1,0), (5,100), and (10,100), are used. The implementation details of R-NSGA-II can be found in [61].

2.4.2.2. Definition of the problem to optimize and variables.

The problem

that needs to be optimized in this study is the selection of the most pertinent EEG channels for MCI classification while enhancing or at least preserving the classification accuracy. The problem with two objective functions is presented in its generic form in Eq. (12).

Minimize No_ch

$$\text{Maximize } CA(\text{channels}, Param) \quad (12)$$

Subject to $No_ch \geq 1$

Also subject to $CA \leq 100 \text{ \%}$

where No_ch is the number of EEG channels used, CA is the classification accuracy of MCI vs. HC calculated using Eq. (9), and $Param$ is a classifier parameter.

Solving this problem using NSGA-II or R-NSGA-II requires organizing the dataset and variable representation for both algorithms. As seen in Eq. (12), the classification accuracy CA depends on the channels $ch_1, ch_2, \dots, ch_{19}$, and $Param$, which can be represented by 20 genes to form a single chromosome (twenty variables are required). In other words, the first 19 variables correspond to the 19 EEG channels as shown in Fig. 5, while the last variable $Param$ represents the classifier's parameter to be optimized. Each channel variable can have one of two values: 1 or 0, where 1 means that the channel is selected for the classification process while 0 means that a channel is being excluded (see Fig. 5). This representation is inspired by [17], which investigated channel selection for emotion classification. Regarding the classifier's parameter, $Param$ can take values of 1, 2, ..., or 10 to reflect the number of neighbors selected in the KNN. In the case of SVM, $Param$ is used to select kernel type: 1 for linear, 2 for polynomial or 3 for radial basis function (RBF). In the case of discriminant analysis (DA) classifiers, 1 for linear DA and 2 for quadratic DA. In the case of RF, $Param$ denotes the depth of the tree, with values ranging from 1 to 35.

The entire procedure, which is organized into three stages (feature extraction, classification, and optimization), is shown in Fig. 6. The steps are briefly described as follows:

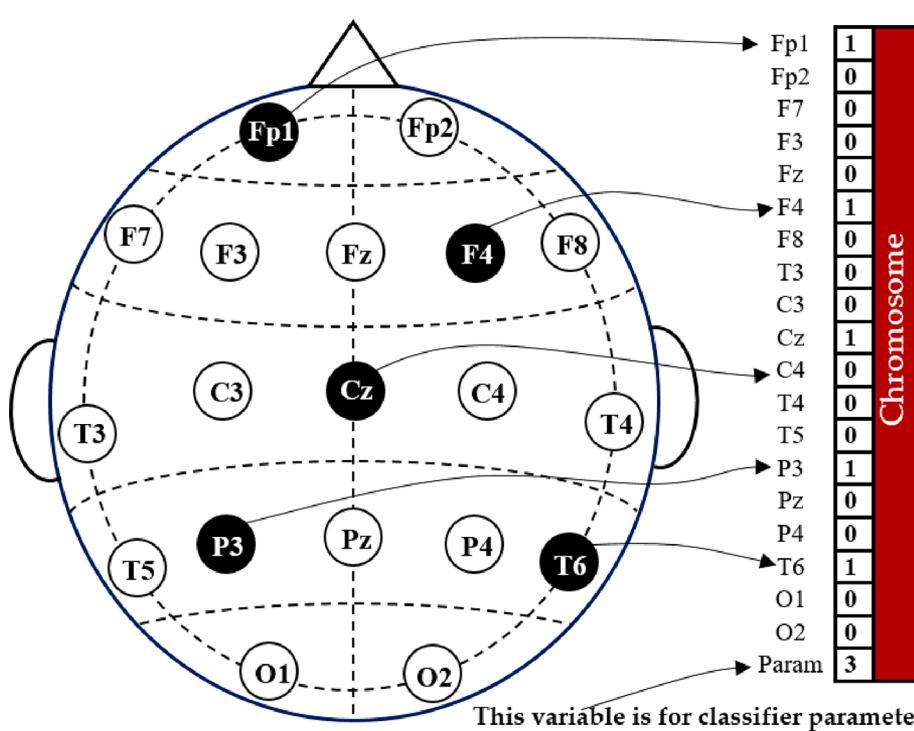


Fig. 5. Example of channel representation in a chromosome for NSGA.

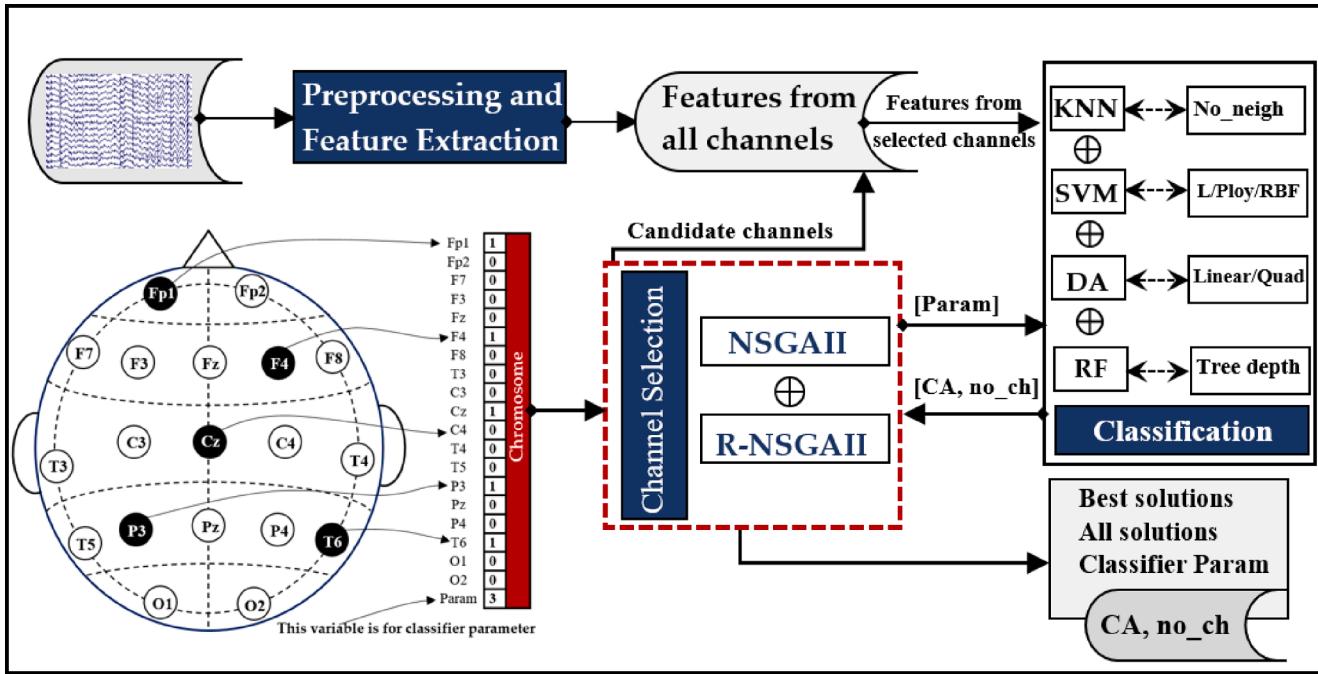


Fig. 6. Process overview for EEG channel selection using NSGA for MCI classification.

1. The raw EEG signals from all EEG channels are first read, pre-processed, and then decomposed using VMD as illustrated in Fig. 4.
2. Second, the features are computed by one of the entropy measures (Eq. 2–6), organized, and then stored for iterative use.
3. NSGA-II or R-NSGA-II is then applied to search for a reduced set of EEG channels and to optimize the classifier's parameter with respect to classification accuracy. This process starts by first constructing an initial population composed of a set of chromosomes. Each chromosome represents a potential solution for the optimal selection of EEG channels and classifier parameter.
4. The relevant subset of features corresponding to channels represented by 1 in the chromosome are considered for the classification procedure while the remaining channels' features are excluded.
5. Using one of the classifiers (RF, DA, SVM, KNN, or ENKNN), accuracy is evaluated while taking into account 10-fold cross validation and the classifier parameter, *Param*. At this stage, the values of the two objective functions, classification accuracy (*CA*) and number of channels (*No_ch*), are calculated to be used by NSGA-II or R-NSGA-II for evaluating each solution (chromosome) in the current population.
6. Steps 3 to 5 are repeated, which results in a gradually evolving population of potential solutions. The process is continued until a stopping criterion is met.

The population size in each iteration is set to 200. The termination criterion is defined by a maximum number of iterations (*MaxIter*), which is set to 50. The total number of evaluated solutions in this case is 10000. Matlab 2022 is used to implement all steps shown in Fig. 6.

3. Results and discussion

In what follows we discuss results and evaluate the proposed methods using all EEG channels in addition to results that are based on a reduced set of the EEG channels.

3.1. Results using all EEG channels (Full-channels)

In this sub-section, we report the classification results of MCI versus

HC without the use of channel reduction and classifier parameter optimization techniques. In other words, a non-optimized classifier is employed to compute the classification accuracy using all EEG channels (full-channel accuracy). For each channel, the signal is passed through a 0.5–32 Hz band-pass filter and then divided into 3660 segments (1740 segments from MCI cases and 1860 from HC cases). Using the adopted VMD and entropy FE approaches, each segment is transformed into a feature vector of length 76 collected from 19 channels. This results in a 3660×76 feature matrix, which is used as input to the classifier. Table 3 shows the results of KNN ($K = 3$) when applied with five different FE methods. All results are averaged using 10-fold CV. According to the results in Table 3, the best classification accuracies of 99.51% and 98.88% are achieved by VMD + LogEn and VMD + ShEn methods, respectively. VMD + ThEn performs well too, with accuracy results exceeding 98%. On the other hand, VMD + NoEn FE techniques perform the worst.

In addition to KNN, five more classification methods are used for further investigation. Table 4 presents the accuracy results of RF (tree depth of 30), LDA, QDA, SVM (polynomial kernel), and ENKNN (sub-space) methods. The LDA and QDA classifiers have the lowest classification accuracies, as shown in the table. VMD + ShEn + ENKNN and VMD + LogEn + KNN both achieve accuracy values of 99.54% and 99.51%, respectively. Table 4 also demonstrates that the VMD + NoEn

Table 3

The full-channel classification results (mean \pm st) of MCI vs. HC using KNN ($k = 3$) classifier.

F methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-score (%)	Kappa (%)
VMD + ShEn	98.88 \pm 0.37	99.08 \pm 0.76	98.71 \pm 0.65	98.82 \pm 0.39	97.75 \pm 0.75
VMD + NoEn	94.70 \pm 1.13	94.86 \pm 1.74	94.59 \pm 1.23	94.40 \pm 1.19	89.37 \pm 2.27
VMD + SuEn	97.19 \pm 1.09	97.52 \pm 1.51	96.91 \pm 1.07	97.03 \pm 1.14	94.36 \pm 2.18
VMD + ThEn	98.39 \pm 0.69	98.90 \pm 0.84	97.94 \pm 0.80	98.29 \pm 0.73	96.77 \pm 1.38
VMD + LogEn	99.51 \pm 0.38	99.60 \pm 0.60	99.43 \pm 0.38	99.48 \pm 0.40	99.01 \pm 0.77

Table 4

The full-channel classification accuracy (mean \pm st) of MCI vs. HC using different classification models.

FE methods	RF	LDA	QDA	SVM	KNN	ENKNN
VMD + ShEn	97.54 \pm 0.64	80.98 \pm 1.31	96.12 \pm 0.71	97.90 \pm 0.39	98.88 \pm 0.37	99.54 \pm 0.39
VMD + NoEn	97.54 \pm 0.85	71.53 \pm 2.37	67.62 \pm 1.45	95.03 \pm 0.76	94.70 \pm 1.13	96.17 \pm 1.09
VMD + SuEn	97.76 \pm 0.68	77.49 \pm 2.98	74.18 \pm 1.42	94.70 \pm 1.31	97.19 \pm 1.09	97.92 \pm 0.80
VMD + ThEn	98.33 \pm 0.69	77.70 \pm 1.57	97.08 \pm 0.78	97.98 \pm 0.58	98.39 \pm 0.69	99.45 \pm 0.39
VMD + LogEn	98.39 \pm 0.49	79.81 \pm 1.16	97.10 \pm 1.00	98.06 \pm 0.67	99.51 \pm 0.38	99.48 \pm 0.42

feature extraction method performs the worst with most classifiers. Consequently, this method will not be considered throughout the remainder of this paper. The classifiers are later compared using optimized parameters.

3.2. Results using a reduced set of EEG channels

The results of various channel selection methods are presented in this section with the goal of minimizing the number of EEG channels while maintaining classification accuracy. One possible method for channel selection is to simply evaluate the classifier's accuracy for each individual EEG channel (single-channel classification) and then select the channels with the best result (incremental evaluation). In other words, in the incremental evaluation method, the channels are ordered in descending order according to their single classification accuracy, then the incremental evaluation of a subset of channels is computed. Fig. 7 shows the KNN classification accuracy for each of the EEG channels when using the VMD + LogEn and VMD + ThEn FE methods, while Fig. 8 shows the classification accuracies based on incremental evaluation of the channels. As shown in Fig. 8, incremental evaluation increase the accuracy with increasing the subset of EEG channels, but it failed to achieve, with fewer channels, the full-channel accuracy presented in Table 3. For example, in the case of VMD + ThEn FE method, the full-

channel accuracy is 98.39%. However, incremental evaluation failed to achieve this accuracy score with fewer channels. This is also confirmed by the results of [25], in which there was an attempt to select a few EEG channels using incremental evaluation. According to [25], incremental evaluations improves the accuracy as the number of EEG channels is increased. But the highest accuracy of 96.94% was achieved only when all channels (19 channels) were used. Therefore, in the present study, other methods for selecting EEG channels are investigated to explore the feasibility of achieving full-channel accuracy with fewer channels, as discussed in the next subsections.

3.2.1. Results based on greedy algorithms

Fig. 9 shows the classification accuracy scores based on the EEG channels selected using the back-elimination (BE) method. The classification is performed using the KNN ($K = 3$) while VMD + ShEn, VMD + SuEn, VMD + ThEn, and VMD + LogEn are used for feature extraction. The red dashed line in the figure represents the full-channel classification accuracy obtained using 19 channels. The black curve represents the classification accuracy scores (local optimal) obtained at each iteration of BE. Results in Fig. 9 show that fewer channels selected by the BE method could lead to accuracy scores higher than those obtained with all channels (full-channel accuracy). For example, in Fig. 9a (VMD + ShEn), the full-channel accuracy is 98.88%, but the same accuracy is achieved with only nine channels selected by BE. Moreover, a better accuracy of 99.15% is achieved when 11 channels are selected using BE. Another example in Fig. 9d (VMD + LogEn), with 13 selected channels, the full-channel accuracy is achieved. The highest accuracy of 99.59 is achieved using this FE method with 15 channels selected by BE.

Fig. 10 shows the classification accuracy evaluated using subsets of EEG channels selected using the forward-addition (FA) method. Like BE, FA managed to select a smaller number of EEG channels that are able to achieve the full-channel accuracy over all FE methods. It can be noted that the performance of the FA method is comparable to the performance of BE method. As shown in Fig. 10, the full-channel accuracy scores are exceeded with 11, 9, 18, and 12 channels using VMD + ShEn, VMD + SuEn, VMD + ThEn, and VMD + LogEn, respectively. From Figs. 9 and 10, it can be noted that the performances of BE and FA are also related to the performances of feature extraction methods. For

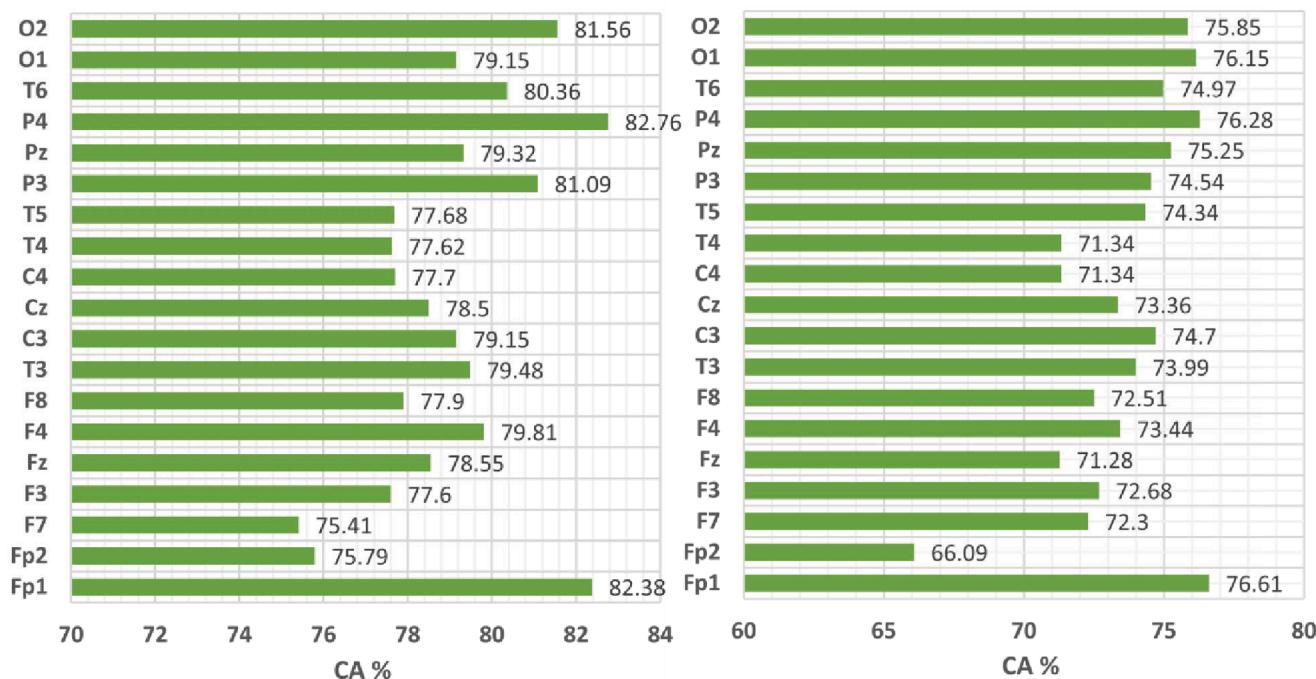


Fig. 7. KNN classification accuracy when applied to a single-channel with VMD + LogEn (left) and VMD + ThEn (right).

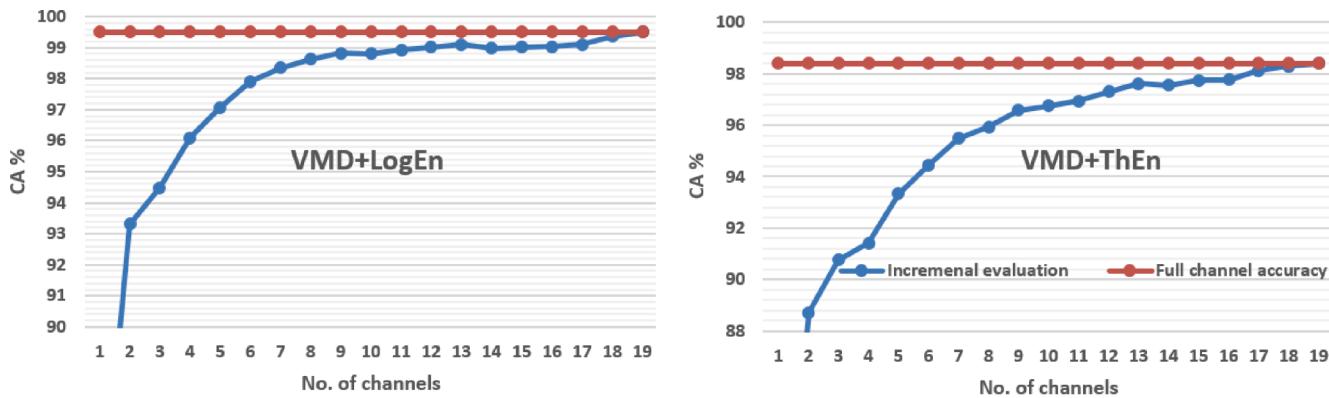


Fig. 8. KNN classification accuracy based on incremental evaluations with VMD + LogEn (left) and VMD + ThEn (right). The incremental evaluation method failed to select the most relevant channels that lead to achieve, with fewer channels, the full-channel accuracy.

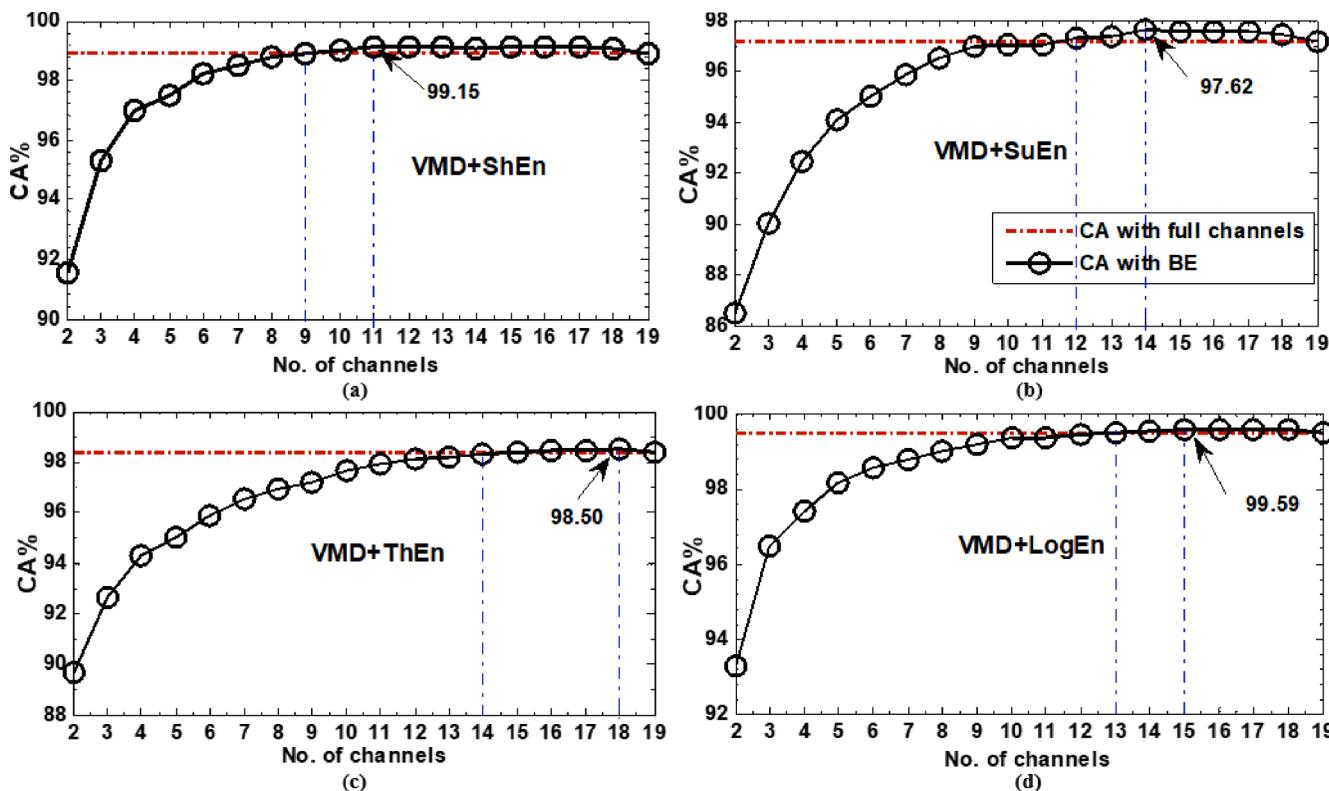


Fig. 9. KNN classification accuracy of a subset of EEG channels selected using back-elimination method with features extracted by (a) VMD + ShEn, (b) VMD + SuEn, (c) VMD + ThEn, and (d) VMD + LogEn.

example, VMD + LogEn achieved the highest full-channel accuracy (see Table 3) and also achieved the maximum accuracy scores when BE and FA (Fig. 9d and Fig. 10d) are used with a lower number of channels.

3.2.2. Results based on NSGA methods

In this subsection, NSGA-based methods are employed to select subsets of EEG channels and the value of K for the KNN classifier. Fig. 11 shows the classification results using NSGA-II with four FE methods. The red dashed line represents the full-channel accuracy while the black curve represents the NSGAII-based accuracy scores. As shown in the figure, NSGA-II is able to select a subset of channels that achieves a similar accuracy compared to full-channel accuracy. For example, with only 8, 12, 10, and 10 channels for VMD + ShEn, VMD + ThEn, VMD + SuEn and VMD + LogEn, respectively, the KNN classifier's accuracy is similar to that of using all 19 EEG channels. Moreover, NSGA-II is able to

find channel subsets for which the accuracy is actually higher than that of full-channel classifier. The maximum accuracy scores are 99.48%, 97.70%, 98.56%, and 99.64%, when using 12, 16, 14, and 11 channels with VMD + ShEn, VMD + SuEn, VMD + ThEn, and VMD + LogEn FE methods, respectively. Among the four FE methods, VMD + LogEn and VMD + ShEn are the best methods for achieving the maximum accuracy with the fewest number of channels, as shown in Fig. 11d and Fig. 11a. Regarding R-NSGA-II, Fig. 12 shows the classification results when using subsets of EEG channels selected using R-NSGA-II with the four FE methods. Results show that, for all FE methods, R-NSGA-II achieves the full-channel accuracy with 8, 10, 12, and 9 channels. Using subsets of 12 channels, VMD + SuEn, VMD + ThEn, and VMD + LogEn achieve the maximum accuracy scores of 97.65%, 98.55%, and 99.67%, respectively, which exceed the accuracy of full-channel classifier. In the case of VMD + ShEn, a maximum accuracy of 99.51% is achieved using a subset

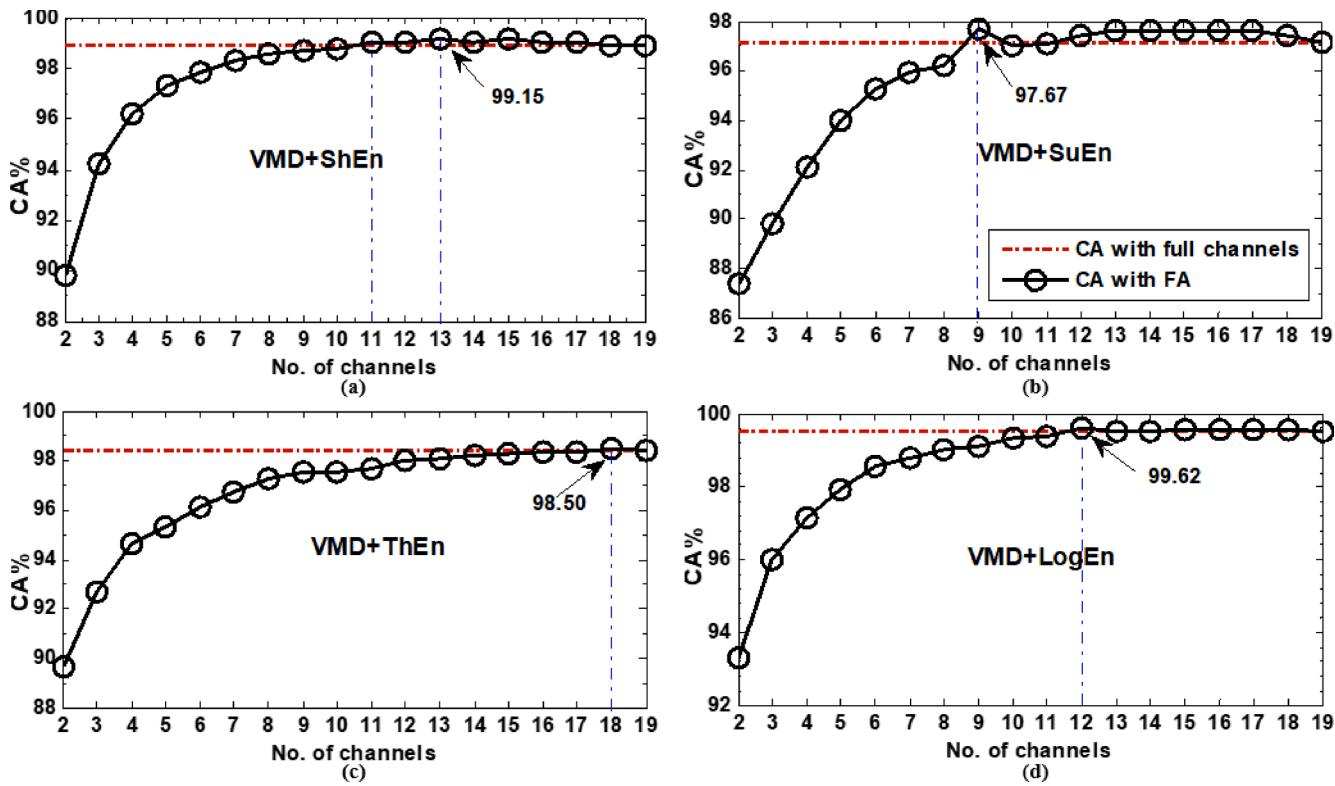


Fig. 10. KNN classification accuracy of a subset of EEG channels selected using forward-addition method with features extracted by (a) VMD + ShEn, (b) VMD + SuEn, (c) VMD + ThEn, and (d) VMD + LogEn.

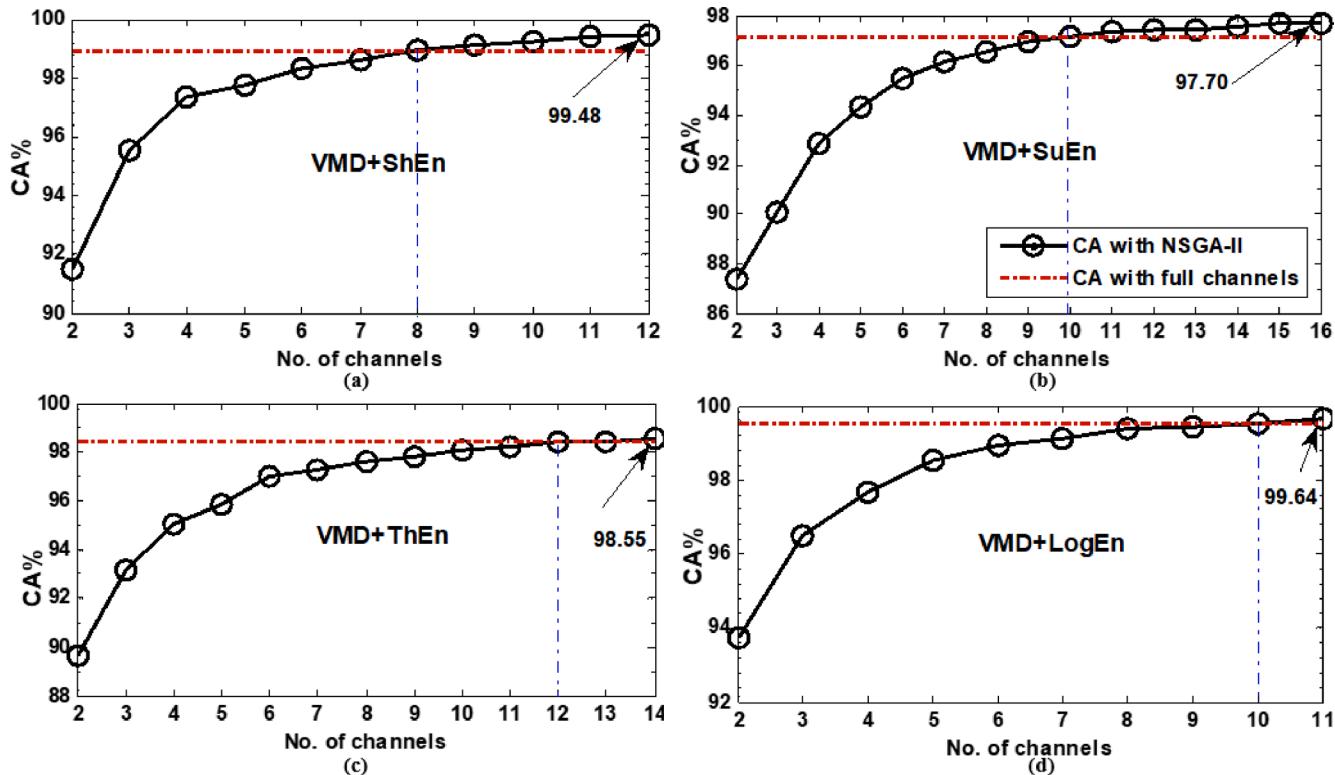


Fig. 11. KNN classification accuracy of subsets of EEG channels selected using NSGA-II with features extracted by (a) VMD + ShEn, (b) VMD + SuEn, (c) VMD + ThEn, and (d) VMD + LogEn.

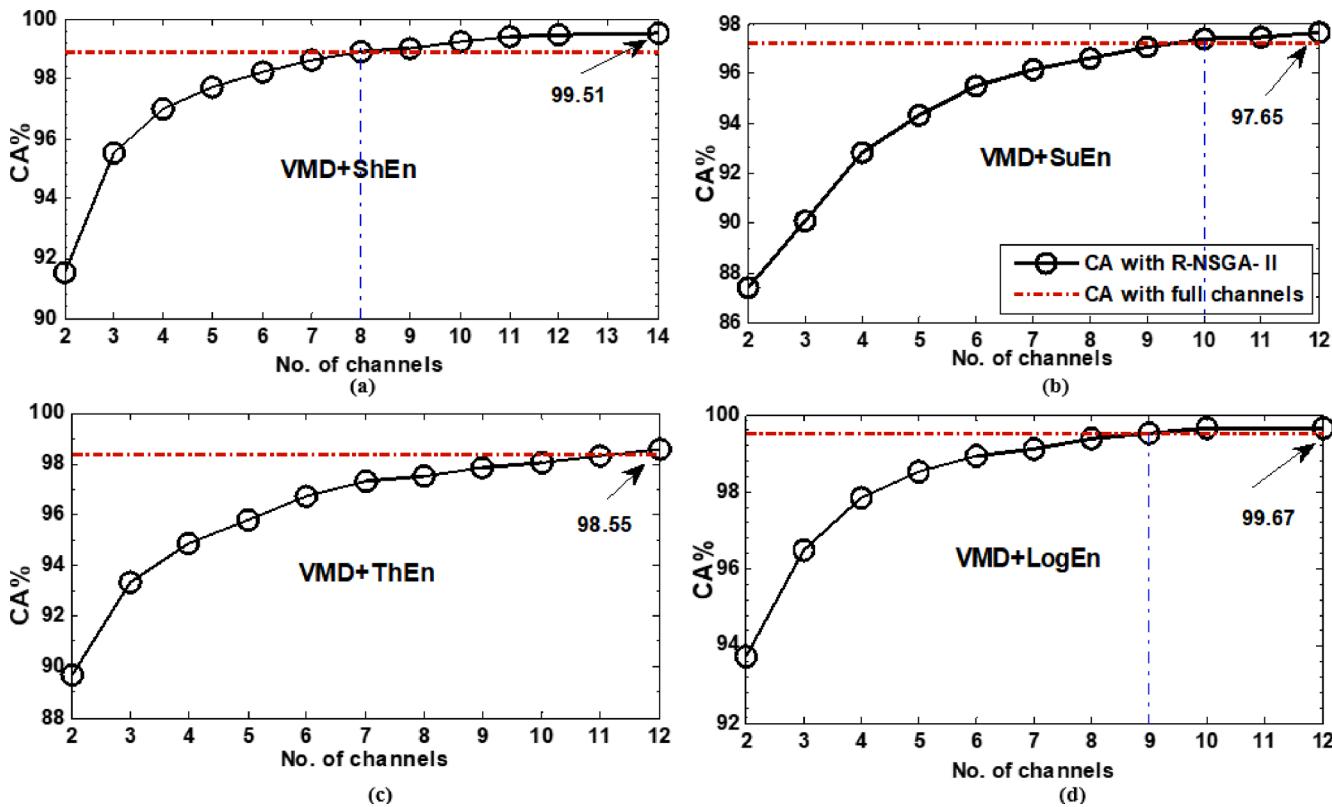


Fig. 12. The KNN Classification accuracy of EEG channels selected using R-NSGA-II with features extracted by (a) VMD + ShEn, (b) VMD + SuEn, (c) VMD + ThEn, and (d) VMD + LogEn.

of 14 channels.

Results in Figs. 9, 10, 11, and 12 demonstrate the success of the optimization and greedy methods in selecting a smaller subset of channels that lead to a higher classification accuracy than those obtained using all channels. This may indicate two things. The first is that there are some unimportant channels that may be redundant and negatively affect the accuracy of the classification. There are also some channels that contain some information, but their presence together tends to attenuate each other. Therefore, the use of effective methods to select a group of channels (the most relevant channels) that are compatible and positively affect the classification accuracy is promising for reducing the number of channels.

3.2.3. Comparison between channel selection methods

In this subsection, results of greedy algorithms (BE and FA) and multi-objective optimization algorithms (NSGA-II and R-NSGA-II) are compared and discussed. Fig. 13 shows the KNN classification accuracy results based on those algorithms with VMD + ShEn and VMD + LogEn FE methods. In the case of VMD + ShEn results (Fig. 13a), the full-channel accuracy is 98.88% (as reported before in Table 3). All channel selection methods are able to reach and exceed this accuracy with fewer channels. For example, BE is able to achieve an accuracy of 99.02% using 10 channels. With the same number of channels, NSGA-II and R-NSGA-II achieved an accuracy of 99.24%. From Fig. 13a, it can be seen that the BE has a slightly better performance than the FA when the number of channels is from two to twelve, and after that, the performance of the two algorithms becomes similar. Regarding NSGA-based methods, NSGA-II and R-NSGA-II have similar performance throughout most channel combinations (two best channels, three best, five best, and so on). Fig. 13a also shows that the NSGA-based methods achieve better performance than greedy methods over most channel subsets. With this FE method, namely VMD + ShEn, the highest accuracy of 99.51% is achieved with 14 channels selected by R-NSGA-II.

Fig. 13b shows the VMD + LogEn-based results. Like in Fig. 13a, BE has a slightly better performance than FA, while NSGA-II and R-NSGA-II have similar performance over most channel subsets. Fig. 13b also shows that the NSGA-II and R-NSGA-II methods achieve better performance than the BE and FA methods. The full-channel accuracy achieved with VMD + LogEn is 99.51% as reported in Table 3. However, all channel selection methods reach and exceed this accuracy with less numbers of channels. For example, an accuracy of 99.51% is achieved by FA, BE, NSGA-II, and R-NSGA-II using 13, 13, 10, and 9 channels, respectively. With this FE method, namely VMD + LogEn, the highest accuracy of 99.67% is achieved with a subset composed of 12 channels selected by R-NSGA-II. The most important note in Fig. 13 is that the multi-objective optimization methods (NSGA-based methods) have a greater ability than the conventional methods (greedy methods) to select few compatible channels. This is due to the way NSGA methods work to find the best solutions. NSGA produces populations of chromosomes that are assessed based on their accuracies, and those with the highest accuracy are reutilized to generate new populations. This cycle is repeated until the best solutions are obtained.

It is noteworthy that, at each iteration, both greedy algorithms (BE and FA) make a decision about the selected subset (local optimum), which is generally not optimal. Furthermore, greedy algorithms may have high accuracy at one iteration and then lower accuracy at the next iteration, despite the fact that the number of channels is greater. For example, in the case of the VMD + ShEn result shown in Fig. 10a, FA has an accuracy of 99.15% with a 13-channel subset, and in the next iteration, the selected 14-channel subset achieves a lower accuracy of 99.07%. On the other hand, NSGA-II and R-NSGA-II employ a multi-objective optimization methodology to search for a solution (Pareto front) that minimizes the number of channels and maximizes the accuracy. For example, in the case of the VMD + ShEn result shown in Fig. 11a, the highest accuracy is 99.48% with a 12-channel subset. After that, the accuracy decreases as the number of channels increases. In

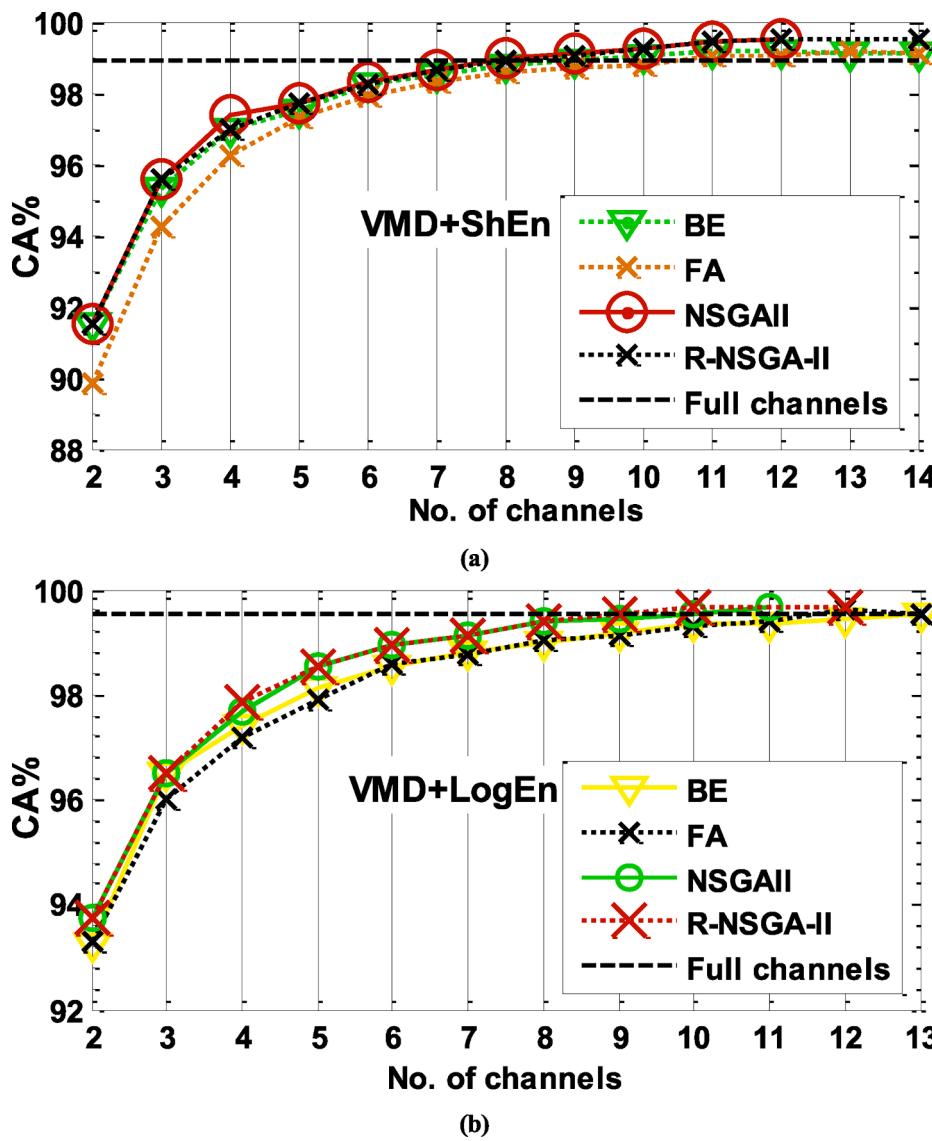


Fig. 13. Comparison of channel selection methods in terms of KNN classification accuracy with two feature extraction methods.

other words, NSGA-II could not find a solution with more than 12 channels that produce accuracy higher than 99.48%, so no solution is returned.

Fig. 14 shows the channel topographies of seven-channel solutions returned by BE, FA, NSGA-II, and R-NSGA-II with different feature extraction methods. Looking at the channels selected by the selection methods in each FE method (i.e., looking at each row in Fig. 14), it can be seen that the solutions returned by NSGA-II and R-NSGA-II are almost similar, while the solutions returned by BE and FA are not. This is possibly because NSGA-II and R-NSGA-II follow similar methodologies of generating solutions, while BE and FA work in reverse orders. While the solutions across each row of Fig. 14 do not match exactly, there are still some similarities in the form of common channels that are selected by all methods. For example, in the case of VMD + LogEn (Fig. 14d), channels Fp1, F8, T6 and O1 are selected by BE, NSGA-II, and R-NSGA-II. On the other hand, by looking at the solutions found by each channel selection algorithm over the various FE methods (i.e., looking at each column in Fig. 14), it can be seen that the solutions differ from one FE method to another. This is because each FE method extracts different biomarkers from each channel, which may lead to different solutions. However, some channels appear in most solutions at each column in Fig. 14. For example, in the case of NSGA-II and R-NSGA-II, the channels

Fp1, F8, T6, and O1 are found in the solutions of VMD + ShEn, VMD + SuEn, and VMD + LogEn FE methods. The fact that the selected channels differ depending on the used FE method is also confirmed by [24]. The authors of [24] used Fisher's method to select the features extracted by four-relative power methods (i.e., four different FE methods that are all based on relative power). Results in [24] showed that for each FE method a different optimum channel subset was selected. For example, with their second FE method, five channels were selected that were not identical to those selected with the third and fourth FE methods.

The optimal subset of channels is defined as the subset of channels that contains fewer channels while achieving the highest classification accuracy. Fig. 15 shows the channel topographies for the solutions returned by R-NSGA-II that produce the maximum classification accuracy with different feature extraction methods. In the case of VMD + LogEn, VMD + ThEn, and VMD + SuEn FE methods, each solution includes 12 channels, achieving the maximum classification accuracy of 99.67%, 98.55%, and 97.65, respectively. In the case of VMD + ShEn, shown in Fig. 15d, the best solution includes 14 channels, achieving a maximum classification accuracy of 99.51%. As can be seen in Fig. 15, Fp2, F8, T5, Pz, and T6 are common channels in all the presented solutions. By looking again at the figure, it can be seen that Fp1, F7, T3, P3, and O1 appear in the majority of the solutions.

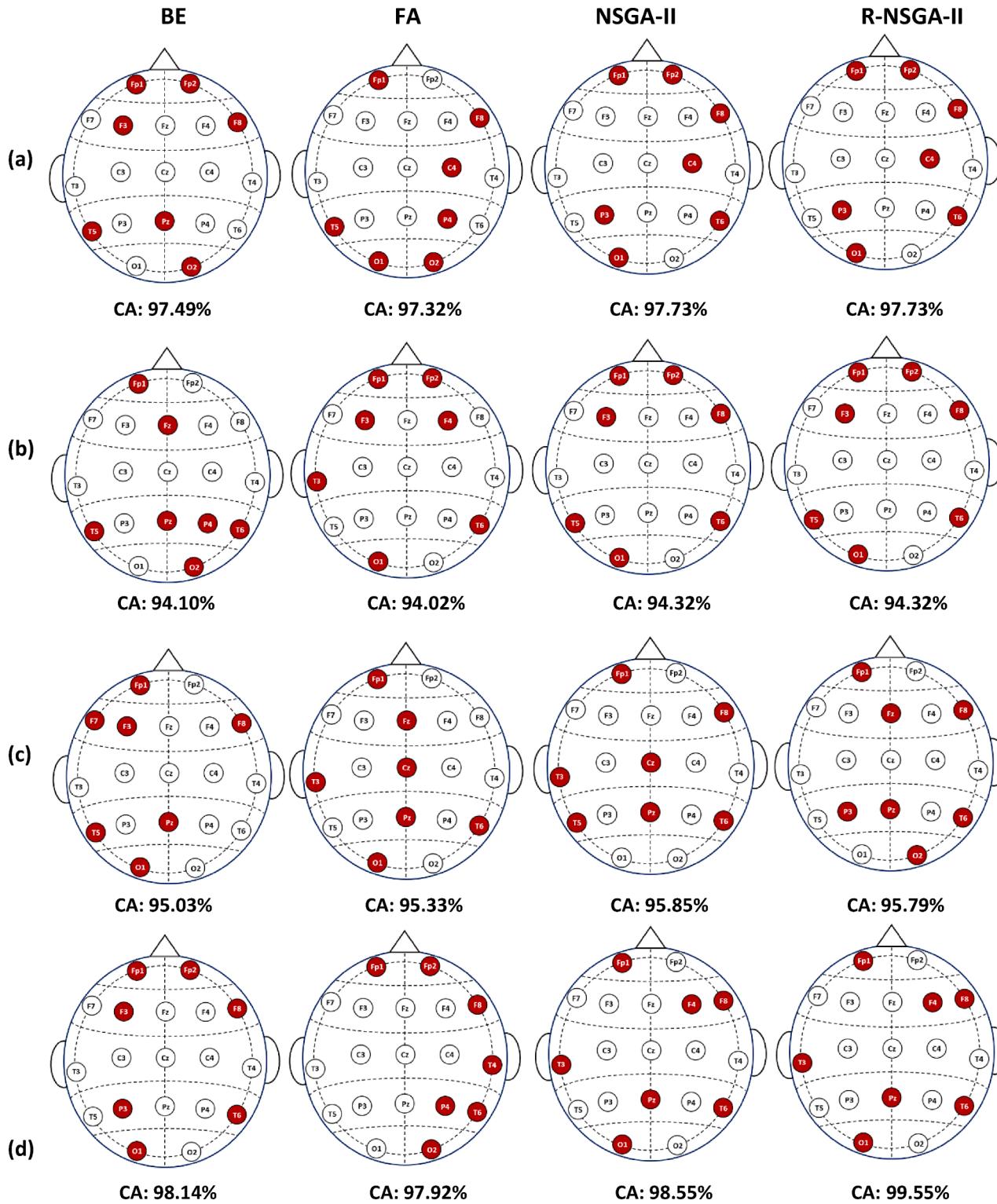


Fig. 14. Seven-channel solutions selected using different channel selection methods with features extracted by (a) VMD + ShEn, (b) VMD + SuEn, (c) VMD + ThEn, and (d) VMD + LogEn.

To provide an impression of the most and the least relevant channels, the number of times each channel is selected in each combination (subsets with 1–10 channels) using BE, FA, NSGA-II, and R-NSGA-II over all FE methods is counted and presented in Fig. 16. Each cell in the figure contains a value that represents the number of times a particular channel is selected within a subset of channels. To make it easier to interpret, the cells are colored in varying degrees of brown matching their values. For

example, whenever the number of times a channel is selected is large, the cell is represented in dark brown, and vice versa. As shown in the figure, channel Fp1 is the most selected channel within the subsets of 1–10 channels. Channels T6, F8, O1, Pz, Fp2, and T5 are also frequently chosen within the majority of subsets. The figure also shows that channels C4, T4, C3, and Cz are the least selected channels.

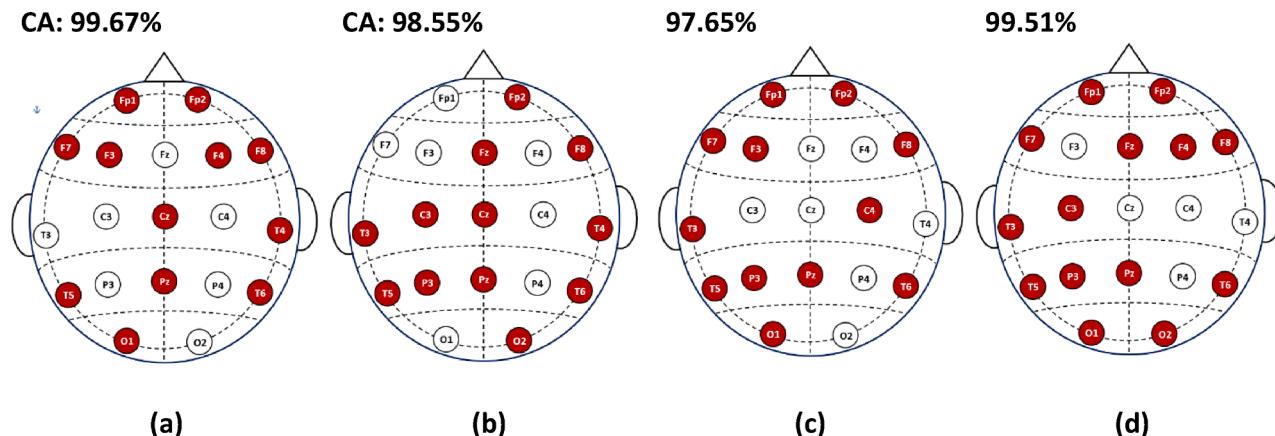


Fig. 15. The optimal channels selected by R-NSGA-II producing the maximum KNN CA with four FE methods: (a) VMD + LogEn, (b) VMD + ThEn, (c) VMD + SuEn, and (d) VMD + ShEn. The solutions (channels) differ from one FE method to another. Each FE method extracts different biomarkers from each channel, which may lead to different solutions.

No. of chs	Fp1	Fp2	F7	F3	Fz	F4	F8	T3	C3	Cz	C4	T4	T5	P3	Pz	P4	T6	O1	O2
1	12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	
2	16	0	0	0	0	0	1	2	0	0	0	0	0	0	4	2	3	4	0
3	15	1	0	1	1	0	4	0	0	0	0	0	0	5	1	4	2	5	8
4	15	7	1	1	1	0	8	1	1	2	0	0	4	0	6	2	6	7	1
5	16	5	2	3	1	0	10	1	1	1	1	0	6	0	8	2	10	8	5
6	16	7	3	5	2	0	12	4	1	1	1	1	7	3	8	3	10	7	5
7	16	8	1	6	3	3	13	5	1	2	3	1	7	4	8	3	13	10	6
8	16	10	3	7	4	3	13	6	1	2	4	2	9	2	11	4	14	10	7
9	16	12	5	6	4	6	13	6	3	3	2	4	10	3	12	5	14	12	9
10	16	14	5	6	5	8	13	9	6	4	1	5	11	5	13	7	12	12	9

Fig. 16. The number of times the EEG channels are selected by all the channel selection algorithms and FE methods.

3.2.4. Performance of classifiers and selected parameters

In this sub-section, performance results of RF, DA, SVM, KNN, and ENKNN classifiers are compared. Based on results discussed above, NSGA-II is adopted as the channel selection method. In addition, NSGA-II is also used to optimize the classifiers' parameters. As earlier mentioned, the last variable in the chromosome, *Param*, holds the parameter of a classifier to be optimized (for more details, see the sub-section: **Definition of the problem to optimize and variables**). Results indicate that the best value of *Param* may vary when the number of the selected channels changes. For example, in the case of KNN, NSGA-II selects the value of 3 ($k = 3$) when the number of channels is one, while $K = 1$ is selected when the number of channels is greater. In the case of SVM, the polynomial kernel is always selected for both VMD + ThEn and VMD + LogEn. With VMD + ShEn, the RBF kernel is selected when the number of channels is small while polynomial kernel is selected when the number of channels is large. In the case of DA, quadratic type (QDA) is always selected regardless of the number of channels. It is worth noting that the optimization algorithm never selected linear DA or linear SVM. This may be due to the non-linear nature of the EEG data that needs non-linear classifiers. The selected values of the tree depth in RF classifier are ranging from 29 to 35 in all cases. In the case of ENKNN, the best number of neighbors found by NSGA-II is between 32 and 50 in all cases. The classification accuracy of RF, SVM, KNN, and ENKNN with

three FE methods is illustrated in Fig. 17. Different markers with different colors are employed to make the results clear. In the cases of VMD + ShEn and VMD + LogEn FE methods (Fig. 17a and Fig. 17c), RF achieves the highest accuracy when the number of channels is one, whereas KNN and ENKNN perform better when the number of selected channels is high. In the case of VMD + ThEn (Fig. 17b), RF achieves the highest accuracy when the number of channels is smaller than five but, beyond that, ENKNN performs the best. With this FE method, namely VMD + ThEn, RF outperforms KNN and SVM, regardless of the number of selected channels. In general, over the three FE methods, ENKNN and KNN achieve the best performance while SVM performs the worst, in most cases. With the DWT + LogEn FE method, ENKNN, KNN, RF, and SVM classifiers achieve their highest accuracy scores (99.81%, 99.64%, 98.58%, and 98.50%) with 11, 11, 12, and 15 channels, respectively. Finally, we discuss the complexity of the classifiers. During the implementation of those classifiers, it was noted that some of them required more time than others. For example, the ensemble classifiers, RF and ENKNN, require more time compared to others because they involve combining multiple models to improve predictive performance. On the other hand, linear classifiers, such as linear SVM and LDA, required less time but generally at the expense of classification accuracy. The KNN classifier demonstrated both good performance, close to ensemble classifiers, and minimal time, similar to linear classifiers. Hence, KNN

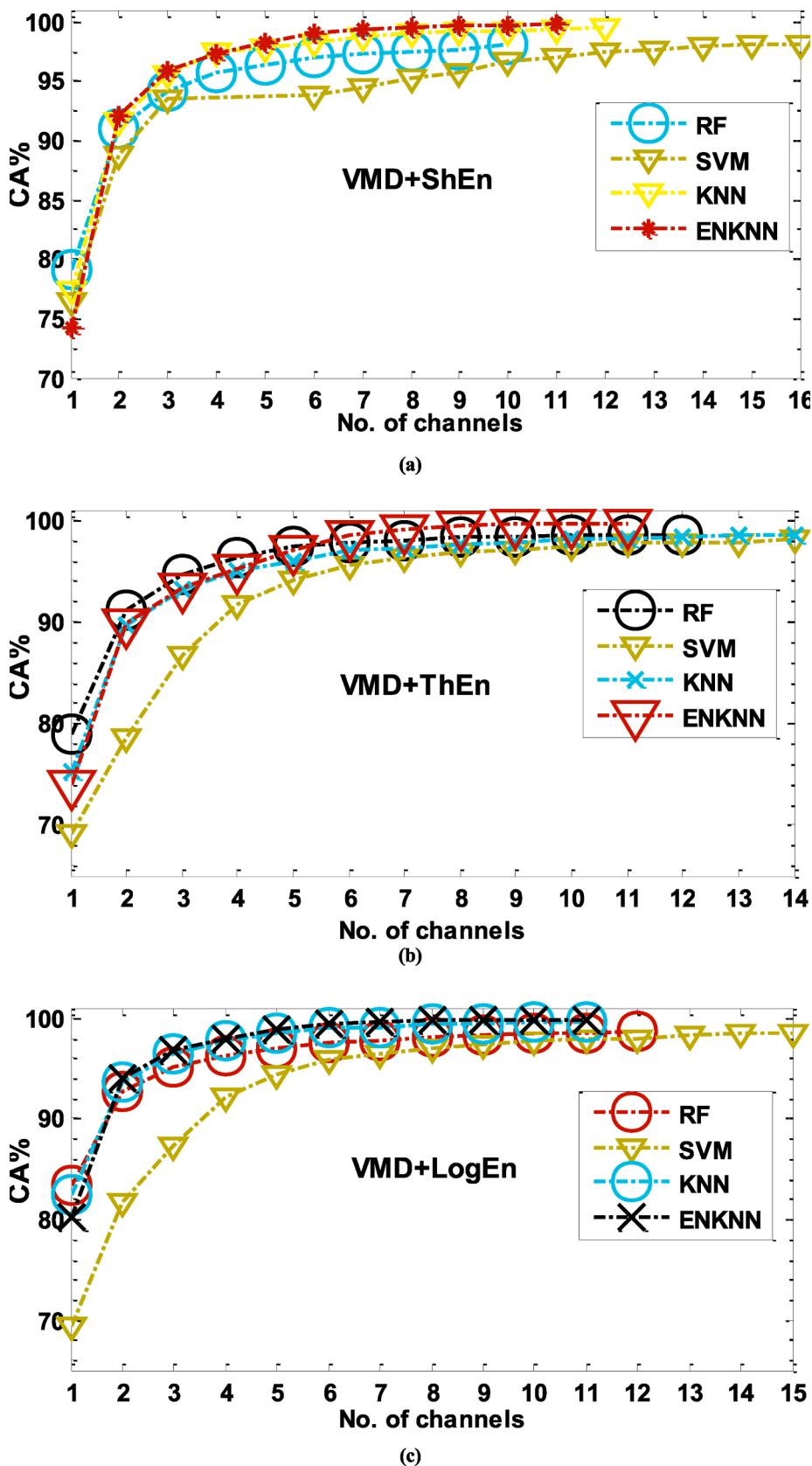


Fig. 17. The classification accuracy of MCI vs. HC using different classifiers with channels selected by NSGA-II. The genes of the NSGA are allocated to predict the best features for a given classifier.

was adopted for most of the investigations in the current study.

3.3. Comparison with previous studies

In the field of neurology, numerous studies have investigated the neurobiological underpinnings of MCI using MRI-based methods. It is known that EEG is limited to the retrieval of cortical activation, while MRI can retrieve activation not only in the cortex but also in subcortical brain structures. For example, R. C. Petersen et al. [62] concluded that MRI-based hippocampal volumetry accurately depicts the structural-functional relationships between memory loss and hippocampal damage across the spectrum from normal aging to dementia. S. E. Rose et al. [63] utilized MRI-based DTI to examine microstructural disparities in the brains of MCI patients relative to controls, revealing significantly increased mean diffusivity measurements in the left and right entorhinal cortices (BA28), posterior occipital-parietal cortex (BA18 and BA19), right parietal supramarginal gyrus (BA40), and right frontal precentral gyrus (BA4 and BA6) among MCI patients. In a study investigating anatomical and functional impairments in patients with MCI, Han et al. [64] reported notable reductions in gray matter volume in the bilateral prefrontal, left temporal regions, and posterior cingulate cortex. The study also found decreased fluctuations mainly in the prefrontal, left parietal, and right fusiform gyrus and increased fluctuations in the limbic and midbrain. The studies [63,64] concluded that there are multiple brain regions that are different in MCI patients compared to healthy people.

Regarding EEG, Table 5 summarizes the methods and findings of previous studies, as well as our own, focusing on the classification of MCI patients versus the HC group. It is not straightforward to compare our results with those of the previous studies due to differences in methodology, dataset, and evaluation strategy. Regarding the channels used, most of the studies in the table ([20,23,26,27,29–31,33]) used all channels to extract features and calculate classification accuracy. However, there are attempts in [21,22,24,25,28] to reduce the number of channels. For example, in [21], the brain region was divided into five

regions: frontal, left temporal, central, right temporal, and occipital. After implementing brain-region-based classification, the same accuracy of 88.89% was achieved from each region, except the frontal region, which achieved an accuracy of 77.78%. The same authors used a different dataset in [22] and implemented single-channel classification and brain-region-based classification. The highest classification accuracy of 80% was obtained based on the left-temporal region that includes F7, T3, and T5 channels. The study [21] shows that MCI involves multiple brain regions, while [22] indicates that the left-temporal region is the most affected.

Accordingly, our present study, similar to [24,25,28], tries to search for the most relevant channels in different brain regions (the cerebral cortex) for detecting MCI. Authors in [24] applied Fisher's class separability criterion to determine the best electrodes (channels) as well as the frequency subbands for extracting the most sensitive relative power features. Out of thirty channels, five were selected from different brain regions (Fp2, F7, T3, T5, and T6), which led to the highest classification accuracy of 90.25% using the SVM classifier. The study [25] investigated decreasing the number of channels by selecting the best subset of channels using the incremental evaluation method. According to the results in [25], incremental evaluations increase the accuracy with increasing the subset of EEG channels, but the highest accuracy of 96.94% was achieved when all channels (19 channels) are used. In our study, we also confirmed the failure of incremental evaluations (see Fig. 8) for selecting a lower number of channels that achieve accuracy higher than that obtained using all channels. Accordingly, selecting optimal channel combinations for MCI detection is not straightforward and requires effective methods. The study [28] manually selected symmetric two-, four-, six-, and eight-electrode combinations, achieving accuracies of 74.04%, 82.43%, 86.28%, and 86.85%, respectively. Although [28] is a good step forward in EEG channel selection, this study evaluated the classification accuracies for only the two-, four-, six-, and eight channels, with only symmetric channel pairs. For example, in the case of evaluating the accuracies of two-channel combinations, the Fp1 channel is selected only with the Fp2 channel, the F7 channel is

Table 5
Comparison of our results with other studies' accuracy for classifying MCI vs. HC.

Study	FE methods	Classifiers	Used dataset	No. of chs.	Classification accuracy (%)
J. Dauwels et al. [20, 2010]	Stochastic event synchrony and Granger causality	LDA/QDA	22 MCI/ 38 HC	21	83% using LOSO CV
M. Kashepoor et al. [21, 2016]	Power, ratio power, and relative power for different bands	Neurofuzzy + KNN	11 MCI/ 16 HC	3	88.89%; 9 MCI and 9 HC for training.
G. Fiscon et al. [31, 2018]	DWT	DT(C4.5)	37MCI/ 23 HC	19	93.3% using 10-fold CV 83.3 using hold out
M. Kashepoor et al. [22, 2019]	Correlation-based Label Consistent K-SVD with spectral features		29 MCI/ 32 HC	3	80%; 41 subject for training and 20 for test.
S. Hadiyoso et al. [23, 2019]	Power spectral-based features	KNN	11 MCI/ 16 HC	19	81.5%
J. Yin et al. [25, 2019]	SWT + statistical features	SVM	11 MCI/ 11 HC	19	96.94%; features vectors are divided as: 60% for training, 20% for validation and 20% for test.
N. Sharma et al. [30, 2019]	PSD, spectral entropy, spectral kurtosis, spectral skewness, and fractal dimension	SVM	16 MCI/ 13 HC	21	84.1%; eyes open while 79.5%; eyes closed using 10-fold CV.
S. Siuly et al. [26, 2020]	Auto-regressive and permutation entropy and	ELM	11 MCI/ 16 HC	19	98.78% using 10-fold CV
Y.T. Hsiao et al. [24, 2021]	kernel Eigen-relative-power	SVM	24 MCI/ 27 HC	5	90.2% using LOSO CV
A. M. Alvi et al. [27, 2022]	–	LSTM	11 MCI/ 16 HC	19	96.41% using 5-fold CV
K. Lee et al. [28, 2022]	Several features using 10 measures	SVM	21 MCI/ 21 HC	8	86.85% using LPSO CV
R. A. Movahed et al. [29, 2022]	Spectral, functional connectivity, and nonlinear features	SVM	18 MCI/ 16 HC	19	99.4% using 10-fold CV
D. Pirrone et al. [33, 2022]	Power intensity for each high-and low-frequency band	KNN	37MCI/ 20 HC	19	95% using 10-fold CV
Present	VMD + ShEn	KNN/ENKNN	(11 MCI/ 16 HC combined with 18 MCI/ 16 HC)	19	98.88/99.54 using 10-fold CV
	VMD + LogEn				99.51/99.48 using 10-fold CV
	VMD + ShEn			11	99.40/99.75 using 10-fold CV and NSGA-II
	VMD + LogEn				99.64/99.81 using 10-fold CV and NSGA-II

M. Aljalal et al.

Biomedical Signal Processing and Control 87 (2024) 105462

selected with only the F8, the O1 channel is selected with only the O2, and so on. This strategy ignores many asymmetric two-channel subsets that may lead to higher accuracy. This also applies to four-, six-, and eight-channel combinations. The authors of [28] used this strategy, symmetric channel pairs, because, of course, it would be very difficult to evaluate all channel subsets of two, four, six, or eight channels.

To tackle these limitations, heuristic optimization methods have been employed in the present study, in which the obtained feature vectors are repeatedly tested for solving two unconstrained objectives to maximize classification accuracy while minimizing the number of EEG channels required for MCI classification. Here, multi-objective optimization methods (NSGA-II and R-NSGA-II) are employed for selecting optimal EEG channel combinations and classifier parameters as well as comparing their results with those of greedy algorithms. According to the results (Figs. 11 and 12), multi-objective NSGA optimization methods succeeded in selecting a few suitable channels that led to superior accuracy compared to a classifier that utilized all EEG channels. To confirm the effectiveness of NSGA-based methods for channel selection, four FE methods were used separately. In each FE method, the NSGA succeeded in selecting suitable channels (see Figs. 11, 12, and 13). According to the results in Figs. 15 and 16, the optimal channels belong to different brain regions (frontal, parietal, temporal, and occipital), confirming the results of [21,63,64], which reported also that several brain regions are affected in MCI patients. It is worth noting that the subcortical brain regions (like hippocampus, entorhinal cortices, and limbic system) reported in [62–64] cannot be accessed using EEG technique.

In summary, we evaluated the performance of our proposed methods on the public dataset (11 MCI and 16 HC) used by [21,23,25–27], combined with the public dataset (18 MCI and 16 HC) used by [29]. We first evaluated the proposed FE methods with all 19 channels, and we were able to achieve a classification accuracy that is higher than what was reported in earlier studies. The accuracy is further improved after implementing the proposed multi-objective optimization method for EEG channel selection. Table 5 demonstrates that when employing the VMD + ShEn and VMD + LogEn FE methods, approximately 99.50% accuracy is obtained using KNN and 11 channels selected by NSGA-II.

4. Conclusion and future work

Electroencephalography (EEG) has gained popularity in identifying brain diseases, as it involves placing several electrodes (channels) directly on the scalp to record brain activity. In this paper, we propose new and effective VMD-based methods for MCI detection from resting-state EEG signals. We also explore several methods for channel number reduction to maintain or even improve the classification accuracy compared to a classifier that uses all available EEG channels. For feature extraction, VMD is combined with log energy entropy, norm entropy, sure entropy, or Shannon entropy. RF, LDA, QDA, SVM, and KNN, are investigated to classify the resulting MCI features from normal ones. Regarding EEG channel selection, we employed multi-objective optimization methods (NSGA-II and R-NSGA-II) and compare their results with those of conventional greedy algorithms (back-elimination and forward-addition) with different FE and classification methods. In addition to EEG channel selection, NSGA-II and R-NSGA-II methods are also used for selecting classifiers' parameters. In the case of using NSGA-II or R-NSGA-II, the obtained feature vectors are repeatedly tested for solving two unconstrained objectives to maximize classification accuracy while minimizing the number of EEG channels required for MCI classification. We investigated the proposed methods on EEGs recorded by 19 electrodes from 32 healthy subjects and 29 patients with MCI.

Results show that the proposed selection methods were successful in finding a few suitable channels that are able to reach and exceed the full-channel accuracy. Moreover, results demonstrate the clear superiority of optimization methods over greedy methods. For instance, the highest full-channel accuracy (i.e., using all 19 EEG channels) is 99.51% when VMD, log energy entropy, and KNN classifier are combined. The same

accuracy was achieved using only 13, 13, 10, and 9 channels selected by FA, BE, NSGA-II, and R-NSGA-II, respectively. When 11 channels were selected using NSGA-II, the ENKNN classifier provided the highest accuracy of 99.81%. Results also show that T6, F8, O1, Pz, Fp2, and T5 are the most selected channels within most solution subsets.

Finally, it can be concluded that choosing fewer suitable channels inevitably leads to an improvement in accuracy. However, the results indicate that the selected channels are affected by numerous factors, including feature extraction methods, classifiers and their parameters, and the approach used for channel selection. The present study represents a first and important step towards utilizing heuristic optimization methods to identify the optimal EEG channels. The final objective is to develop a portable and user-friendly MCI detection system that can be tested in real-time. The authors plan to verify and validate the results using additional feature extraction methods, classifiers, and channel selection approaches. Another future exploration is to identify the common channels, build a new model using those channels, and assess the classification accuracy. In addition to EEG channel selection, the authors plan to utilize heuristic optimization techniques to identify the most relevant features within each channel. The genes of NSGA can be also arranged to automatically predict the best classifier with the best parameters. This can be done in our future explorations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to the data used in the manuscript.

Acknowledgements

This research was supported by King Saud University, Riyadh, Saudi Arabia, under the Researchers Supporting Project number (RSPD2023R651).

References

- [1] World Health Organization, Dementia fact sheet, WHO Press, Geneve, Switzerland, 2020.
- [2] A. Burns, S. Iliffe, Alzheimer's disease, *BMJ* 338.feb05 1 (2009): b158-b158.
- [3] M. Prince, E. Albanese, M. Guerchet, M. Prina, *World Alzheimer Report 2014: Dementia and risk reduction: An analysis of protective and modifiable risk factors*, 2014.
- [4] Alzheimer's Association, 2015 Alzheimer's disease facts and figures, *Alzheimer's & Dementia* 11.3 (2015) 332–384.
- [5] Alzheimer's Association, Treatments and Research, https://www.alz.org/research/science/alzheimers_disease_treatments.asp.
- [6] US Food and Drug Administration, FDA grants accelerated approval for Alzheimer's disease treatment (2023).
- [7] M. Aljalal, S.A. Aldosari, K. AlSharabi, A.M. Abdurraqeb, F.A. Alturki, Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques, *Diagnostics* 12 (5) (2022) 1033.
- [8] M. Aljalal, S.A. Aldosari, M. Molinas, K. AlSharabi, F.A. Alturki, Detection of parkinson's disease from EEG signals using discrete wavelet transform, different entropy measures, and machine learning techniques, *Scientific Reports* 12 (1) (2022) 22547.
- [9] Y. Chen, C. Gong, H. Hao, Y.i. Guo, S. Xu, Y. Zhang, G. Yin, X. Cao, A. Yang, F. Meng, J. Ye, H. Liu, J. Zhang, Y. Sui, L. Li, Automatic sleep stage classification based on subthalamic local field potentials, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (2) (2019) 118–128.
- [10] F.A. Alturki, M. Aljalal, A.M. Abdurraqeb, K. AlSharabi, A.A. Al-Shamma'a, Common spatial pattern technique with EEG signals for diagnosis of autism and epilepsy disorders, *IEEE Access* 9 (2021): 24334–24349.
- [11] S. Ibrahim, R. Djemal, A. Alsuwailem, *Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis*, *Biocybernet. Biomed. Eng.* 38 (1) (2018) 16–26.
- [12] F.A. Alturki, K. AlSharabi, A.M. Abdurraqeb, M. Aljalal, EEG signal analysis for diagnosing neurological disorders using discrete wavelet transform and intelligent techniques, *Sensors* 20 (9) (2020) 2505.

- [13] L.A. Moctezuma, M. Molinas, EEG channel-selection method for epileptic-seizure classification based on multi-objective optimization, *Front. Neurosci.* 14 (2020) 593.
- [14] J. Sheng, B. Wang, Q. Zhang, Q. Liu, Y. Ma, W. Liu, M. Shao, B. Chen, A novel joint HCPMMP method for automatically classifying alzheimer's and different stage MCI patients, *Behavioural Brain Res.* 365 (2019) 210–221.
- [15] K. AlSharabi, Y.B. Salamah, A.M. Abdurraqeb, M. Aljalal, F.A. Alturki, EEG signal processing for alzheimer's disorders using discrete wavelet transform and machine learning approaches, *IEEE Access* 10 (2022) 89781–89797.
- [16] V. Jahmunah, S. Lih Oh, V. Rajinikanth, E.J. Ciaccio, K. Hao Cheong, N. Arunkumar, U.R. Acharya, Automated detection of schizophrenia using nonlinear signal processing methods, *Artif. Intell. Med.* 100 (2019) 101698.
- [17] L.A. Moctezuma, T. Abe, M. Molinas, Two-dimensional CNN-based distinction of human emotions from EEG channels selected by multi-objective evolutionary algorithm, *Scientific Reports* 12 (1) (2022) 3523.
- [18] S. Khatun, B.I. Morshed, G.M. Bidelman, A single-channel EEG-based approach to detect mild cognitive impairment via speech-evoked brain responses, *IEEE Trans. Neural Syst. Rehabilitat. Eng.* 27 (5) (2019) 1063–1070.
- [19] R. Cassani, M. Estrellas, R. San-Martin, F.J. Fraga, T.H. Falk, Systematic review on resting-state EEG for alzheimer's disease diagnosis and progression assessment, *Disease Markers* 2018 (2018) 1–26.
- [20] J. Dauwels, F. Vialatte, T. Musha, A. Cichocki, A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG, *NeuroImage* 49 (1) (2010) 668–693.
- [21] M. Kashepoor, H. Rabbani, M. Barekatain, Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features, *Journal of medical signals and sensors* 6 (1) (2016) 25.
- [22] M. Kashepoor, H. Rabbani, M. Barekatain, Supervised dictionary learning of EEG signals for mild cognitive impairment diagnosis, *Biomed. Signal Process. Control* 53 (2019) 101559.
- [23] S. Hadiyoso, C.L.F.A.R. Cynthia, M.T.L. ER, H. Zakaria, Early detection of mild cognitive impairment using quantitative analysis of EEG signals, 2019 2nd International Conference on Bioinformatics, Biotechnology and Biomedical Engineering (BioMIC)-Bioinformatics and Biomedical Engineering. Vol. 1. IEEE, 2019.
- [24] Y.T. Hsiao, C.F. Tsai, C.T. Wu, T.T. Trinh, C.Y. Lee, Y.H. Liu, MCI Detection Using Kernel Eigen-Relative-Power Features of EEG Signals, *Actuators*. Vol. 10. No. 7. MDPI, 2021.
- [25] J. Yin, J. Cao, S. Siuly, H. Wang, An integrated MCI detection framework based on spectral-temporal analysis, *Int. J. Automat. Comput.* 16 (6) (2019) 786–799.
- [26] S. Siuly, Ö.F. Alçin, E. Kabir, A. Şengür, H. Wang, Y. Zhang, F. Whittaker, A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals, *IEEE Trans. Neural Syst. Rehabilitat. Eng.* 28 (9) (2020) 1966–1976.
- [27] A.M. Alvi, S. Siuly, H. Wang, A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals, *IEEE transactions on emerging topics, Computat. Intell.* 7 (2) (2023) 375–388.
- [28] K. Lee, K.M. Choi, S. Park, S.H. Lee, C.H. Im, Selection of the optimal channel configuration for implementing wearable EEG devices for the diagnosis of mild cognitive impairment, *Alzheimer's Res. Therapy* 14 (1) (2022) 170.
- [29] R.A. Movahed, M. Rezaeian, P. Li, Automatic diagnosis of mild cognitive impairment based on spectral, functional connectivity, and nonlinear EEG-based features, *Computat. Math. Methods Med.* 2022 (2022) 1–17.
- [30] N. Sharma, M.H. Kolekar, K. Jha, Y. Kumar, EEG and cognitive biomarkers based mild cognitive impairment diagnosis, *Irbrn* 40 (2) (2019) 113–121.
- [31] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. De Salvo, A. Bramanti, P. Bramanti, M.C. De Cola, Combining EEG signal processing with supervised methods for Alzheimer's patients classification, *BMC Med. Inform. Decision Mak.* 18 (1) (2018).
- [32] B. Oltu, M.F. Akşahin, S. Kibaroğlu, A novel electroencephalography based approach for alzheimer's disease and mild cognitive impairment detection, *Biomed. Process. Control* 63 (2021) 102223.
- [33] D. Pirrone, E. Weitschek, P.D. Paolo, S.D. Salvo, M.C.D. Cola, EEG signal processing and supervised machine learning to early diagnose Alzheimer's disease, *Appl. Sci.* 12 (11) (2022) 5413.
- [34] EEG Signals from Normal and MCI (Mild Cognitive Impairment) Cases, Available: <https://misp.mui.ac.ir/en/eeg-data-0>.
- [35] M.F. Weiner, A.M. Lipton, eds, *The American Psychiatric Publishing textbook of Alzheimer disease and other dementias*, American Psychiatric Pub, 2009.
- [36] M. Barekatain, H. Askarpour, F. Zahedian, M. Walterfang, D. Velakoulis, M. Maracy, M.H. Jazi, the relationship between regional brain volumes and the extent of coronary artery disease in mild cognitive impairment, *J. Res. Med. Sci.: Off. J. Isfahan Univ. Medi. Sci.* 19 (8) (2014) 739.
- [37] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans Signal Proces.* 62 (3) (2013) 531–544.
- [38] M. Aljalal, S. Ibrahim, R. Djemal, W. Ko, Comprehensive review on brain-controlled mobile robots and robotic arms based on electroencephalography signals, *Intell. Service Robot.* 13 (4) (2020) 539–563.
- [39] N. Kannathal, M.L. Choo, U.R. Acharya, P.K. Sadashivan, Entropies for detection of epilepsy in EEG, *Comput. Methods Progr. Biomed.* 80 (3) (2005) 187–194.
- [40] R.C. Joy, S.T. George, A.A. Rajan, M.S.P. Subathra, Detection of ADHD from EEG signals using different entropy measures and ANN, *Clin. EEG Neurosci.* 53 (1) (2022) 12–23.
- [41] W. Bosl, A. Tierney, H. Tager-Flusberg, C. Nelson, EEG complexity as a biomarker for autism spectrum disorder risk, *BMC Med.* 9 (1) (2011) 1–16.
- [42] R.R. Coifman, M.V. Wickerhauser, Entropy-based algorithms for best basis selection, *IEEE Trans. Inform. Theory* 38 (2) (1992) 713–718.
- [43] M.J. Kearns, *The computational complexity of machine learning*, MIT press (1990).
- [44] R.O. Duda, P.E. Hart, *Pattern classification*, John Wiley & Sons, 2006.
- [45] G. Baudat, F. Anouar, Generalized discriminant analysis using a kernel approach, *Neural computation* 12 (10) (2000) 2385–2404.
- [46] C.J. Burges, A tutorial on support vector machines for pattern recognition, *Data Min. Knowl. Discov.* 2 (2) (1998) 121–167.
- [47] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, *J. Mach. Learn. Res.* 10 (2009) 2.
- [48] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [49] S.B. Kotsiantis, I. Zaharakis, P. Pintelas, Supervised machine learning: a review of classification techniques, *Emerg. Artif. Intell. Appl. Comput. Eng.* 160 (1) (2007) 3–24.
- [50] P. Refaelzadeh, L. Tang, H. Liu, Cross-validation, *Encyclopedia of database systems* 5 (2009) 532–538.
- [51] A. Swift, R. Heale, A. Twycross, What are sensitivity and specificity? Evidence-Based Nursing 23 (1) (2020) 2–4.
- [52] T. Alotaibi, F.E.A. El-Samie, S.A. Alshebeili, I. Ahmad, A review of channel selection algorithms for EEG signal processing, *EURASIP J. Adv. Signal Process.* 2015 (2015) 1–21.
- [53] T.H. Cormen, et al., Ch. 16: Greedy algorithms, *Introduction to Algorithms*, MIT press, 2009.
- [54] P.M. Narendra, K. Fukunaga, A branch and bound algorithm for feature subset selection, *IEEE Trans. Comput.* 26 (09) (1977) 917–922.
- [55] I. Foroutan, J. Sklansky, Feature selection for automatic classification of non-Gaussian data, *IEEE Trans. Syst. Man Cybernet.* 17 (2) (1987) 187–198.
- [56] K. Deb, *Multi-objective optimisation using evolutionary algorithms: an introduction*, Springer, London, 2011.
- [57] T. Chugh, K. Sindhya, J. Hakanen, K. Miettinen, A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms, *Soft Computing* 23 (9) (2019) 3137–3166.
- [58] O. Kramer, *Genetic algorithms*, Springer International Publishing, 2017.
- [59] N. Srinivas, K. Deb, Multiobjective optimization using nondominated sorting in genetic algorithms, *Evolut. Computat.* 2 (3) (1994) 221–248.
- [60] K. Deb, A. Pratap, S. Agarwal, T.A.M.T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evolut. Computat.* 6 (2) (2002) 182–197.
- [61] K. Deb, J. Sundar, Reference point based multi-objective optimization using evolutionary algorithms. Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, 2006.
- [62] R.C. Petersen, C.R. Jack, Y.-C. Xu, S.C. Waring, P.C. O'Brien, G.E. Smith, R.J. Ivnik, E.G. Tangalos, B.F. Boeve, E. Kokmen, Memory and MRI-based hippocampal volumes in aging and AD, *Neurology* 54 (3) (2000) 581.
- [63] S.E. Rose, K.L. McMahon, A.L. Janke, B. O'Dowd, G. de Zubizaray, M. W. Strudwick, J.B. Chalk, Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnestic mild cognitive impairment, *J. Neurol. Neurosurg. Psychiatry* 77 (10) (2006) 1122–1128.
- [64] Y. Han, S.u. Lui, W. Kuang, Q.i. Lang, L. Zou, J. Jia, C. Soriano-Mas, Anatomical and functional deficits in patients with amnestic mild cognitive impairment, *PLoS One* 7 (2) (2012) e28664.