



VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

An Integrated MCI Detection Framework Based on Spectral-temporal Analysis

This is the Accepted version of the following publication

Yin, Jiao, Cao, Jinli, Siuly, Siuly and Wang, Hua (2019) An Integrated MCI Detection Framework Based on Spectral-temporal Analysis. *International Journal of Automation and Computing*, 16 (6). pp. 786-799. ISSN 1476-8186

The publisher's official version can be found at
<https://link.springer.com/article/10.1007%2Fs11633-019-1197-4>
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/41033/>

An integrated MCI detection framework based on spectral-temporal analysis

Jiao Yin^{1,2}Jinli Cao¹Siuly Siuly³Hua Wang³¹Department of Computer Science and Information Technology, La Trobe University, Victoria 3083, Australia²School of Software Engineering, Chongqing University of Arts and Sciences, Chongqing 402160, China³Institute for Sustainable Industries & Liveable Cities, VU Research, Victoria University

Abstract: Aiming to differentiate between MCI patients and elderly control subjects, this study proposes an integrated framework based on spectral-temporal analysis for the automatic analysis of resting-state EEG recordings. This framework firstly eliminates noise employing stationary wavelet transformation (SWT). Then, a set of features is extracted through spectral-temporal analysis. Next, a new wrapper algorithm, named three-dimensional (3-D) evaluation algorithm, is proposed to derive an optimal feature subset. Finally, the Support Vector Machine (SVM) algorithm is adopted to identify MCI patients on the optimal feature subset. Decision tree and K-Nearest Neighbors (KNN) algorithms are also used to test the effectiveness of the selected feature subset. Twenty-two subjects are involved in experiments, of which eleven persons were in an MCI condition and the rest were elderly control subjects. Extensive experiments show that our method is able to classify MCI patients and elderly control subjects automatically and effectively, with the accuracy of 96.94% achieved by the SVM classifier. Decision Tree and KNN algorithms also achieved superior results based on the optimal feature subset extracted by the proposed framework. This study is conducive to timely diagnosis and intervention for MCI patients, and therefore to delay cognitive decline and dementia onset.

Keywords: Electroencephalogram (EEG), Dementia early detection, Mild cognitive impairment (MCI), Stationary wavelet transformation (SWT), Support Vector Machine(SVM).

1 Introduction

Background. Alzheimers disease (AD) is the most common form of neurodegenerative dementia accounting for up to 75% of all dementia cases [1]. Despite its prevalence, thus far, no cure exists for AD. To make things worse, the diagnosis of Alzheimers disease is often missed or delayed in clinical practice. The early detection of dementia would provide opportunities for early intervention and symptomatic treatments. Recent studies have demonstrated that AD has a pre-symptomatic phase that can last for years, known as mild cognitive impairment (MCI) [2, 3, 4, 5]. Obviously, detecting MCI is essential and effective for potential patients. However, the symptoms of MCI are easily dismissed as normal consequences of ageing, which makes the medical diagnosis of MCI difficult. The objective of this study is to identify MCI patients and elderly control subjects automatically and efficiently using resting-state electroencephalography (EEG) signals.

Resting-state EEG signals. EEG-based methods have emerged as non-invasive alternative techniques for the detection of MCI. Via multiple electrodes placed on different areas of the scalp, the electrical activities of the brain are recorded in EEG signals, which are in the form of time series of voltage fluctuations [6]. Based on diverse recording conditions, EEG signals can be divided into two groups: event-related potentials (ERPs) and resting-state EEG recordings. The former is recorded in relation to the occurrence of some specific events, while the latter are spontaneous EEG signals recorded without any kind of stimulus. Resting-state EEG recordings are very easy and rapid to carry out

in a clinical environment. Furthermore, it is more comfortable and less stressful for patients, especially for elderly individuals [7]. In this study, resting-state EEG signals are adopted.

Literature review. The EEG signals of each subject contain dozens of channels and each channel consists of a huge amount of data points [8]. Traditional specialist-led approaches are struggling to reach a correct decision efficiently [9, 10]. Therefore, automatic detection methods based on machine learning algorithms are getting more and more attention. A typical MCI detection method consists of four steps, namely data pre-processing, feature extraction, feature selection and classification. The most widely used EEG pre-processing methods include visual inspection, resampling, re-referencing, filtering, smoothing, channel selection and data segmentation. Depending on the purpose and the data acquisition conditions, some of these techniques can be selected to refine the EEG signals. Usually, band-stop filters are a good choice for removing power grid interference (50 or 60 Hz, depending on the region). Band-pass filters can be used to enhance only EEG-related spectral components [11, 12]. Feature extraction is generally performed after data pre-processing. There are many widely used techniques to extract features, such as statistical indices [13], spectral analysis [14, 15] and spectral-temporal analysis [16, 17, 18]. The third step, feature selection based on relevance and redundancy analysis, is optimal, depending on the total amount of epochs and features [19, 20]. In the final step, a classifier is trained and evaluated based on machine learning algorithms to differentiate between MCI patients and elderly control subjects. The most commonly used machine learning algorithms in MCI detection include artificial neural networks (ANNs) [21, 22, 23], k-nearest neighbour [2],

decision trees^[16], support vector machine (SVM)^[24] and neuro-fuzzy inference system^[2, 25]. Although there are a large amount of researches on MCI detection, their performance is still not satisfactory. Furthermore, to our best knowledge, there is no standard procedures are commonly accepted in the area as yet. Most of the existing methods are still in the exploratory stage. So, for specific applications and specific data, experienced data processing scientists and engineers need to investigate further to achieve desirable performance.

Proposed method. In this study, we propose an integrated spectral-temporal analysis based framework for MCI detection using resting-state EEG signals, aiming at improving the accuracy of detection. Compared to existing algorithms, our method has several noteworthy aspects:

1. Removing noise of EEG signals based on the spectral characteristics of raw EEG signals. According to domain knowledge, we eliminate baseline drift and other low-frequency noises by removing 0-0.5Hz components of EEG signals, and also eliminate high-frequency noises including grid interference by removing 32-128 Hz components, so as to denoise the EEG signals.
2. Establishing a three-dimensional discrete feature space, based on stationary wavelet transform and descriptive statistical analysis. SWT decomposed EEG signals into coefficients in the frequency domain and descriptive statistical analysis extract the spectral-temporal features from those coefficients.
3. Proposing a new wrapper algorithm, named 3-D evaluation algorithm, to select an optimal feature subset instead of generating new features based on existing features in the feature selection step. It presents both individual and incremental evaluation on three-dimensional of feature space separately.

2 Methods

2.1 Dataset description

The EEG dataset¹ is an open source dataset, which was collected from subjects who had been admitted to cardiac catheterization units of Sina and Nour Hospitals, Isfahan, Iran^[26]. The data collection was ethically approved by the deputy of research and technology, Isfahan University of Medical Sciences, Isfahan, Iran^[2]. It is a collection of resting-state scalp EEG signals from 27 subjects (16 cognitively healthy subjects and 11 with an MCI) aged from 60 to 77 with elementary or higher education and a history of coronary angiography over the past year. To avoid generating an imbalanced dataset in this study, we picked 11 MCI and 11 cognitively healthy subjects to form a balanced dataset. Subjects with a history of substance misuse, major psychiatric disorders, serious medical disease, head trauma, and dementia were excluded.

¹The dataset can be downloaded from:
<http://www.biosigdata.com/?download=eeg-signals-from-normal-and-mci-cases>

All EEG signals were recorded in the morning for over 30 minutes while the subjects were resting comfortably in a quiet room with their eyes closed but without being drowsy during the procedure. EEG activities were recorded continuously through 19 electrodes positioned on the scalp according to the International 10-20 System, using a 32-channel digital EEG device (Galileo NT, EBneuro, Italy) with 256Hz sampling rate^[2]. The collected EEG signals consist of 19 channels, namely, F_{p1} , F_{p2} , F_7 , F_3 , F_z , F_4 , F_8 , T_3 , C_3 , C_z , C_4 , T_4 , T_5 , P_3 , P_z , P_4 , T_6 , O_1 , O_2 .

In light of Petersons criteria, all subjects underwent a neuropsychiatric interview to diagnose MCI. A mini-mental state examination (MMSE) was utilized to validate the MCI diagnosis, where scores from 21 to 26 indicated MCI and scores more than 26 indicated a cognitively healthy subject. The neuropsychiatry unit cognitive assessment tool (NUCOG) was also used to confirm the diagnosis of MCI^[2].

2.2 Methodology

2.2.1 Architecture

The objective of this work is to identify MCI subjects and elderly control subjects using the resting-state EEG signals. As shown in Fig. 1, the proposed framework consists of 4 steps. The raw EEG signals are cleaned using SWT-based methods in step 1. A hybridized method is proposed to extract spectral-temporal features based on stationary wavelet decomposition and descriptive statistical analysis in step 2. Next, an optimal feature subset is selected through the proposed 3-D evaluation algorithm. Finally, an SVM model is chosen as the classifier in step 4. The subsequent parts of this section describe the implementation of each step in detail.

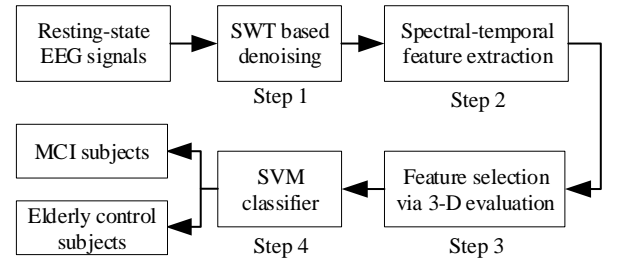


Fig. 1 Work flow of the proposed framework

2.2.2 SWT based denoising

EEG signals contain kinds of noise, such as baseline drift and power line interference. Mixing together with EEG recordings, the large number of artefacts have different time-frequency properties. This study employs wavelet transform to provide information on both the time domain and frequency domain, which makes it possible to preserve the characteristics of EEG signals while minimizing noise.

Previous studies^[27] have shown that the most important frequency bands of EEG signals are between 0.5 Hz and 32 Hz. Therefore, we decompose raw EEG signals into coeffi-

cients with different frequency range using SWT, setting an appropriate decomposition level. Then, the high-frequency (>32 Hz) coefficients and the low-frequency (<0.5 Hz) coefficients are removed as noise [28]. Finally, the cleaned coefficients are reconstructed into time series signals as the denoised EEG signals via inverse stationery wavelet transformation (ISWT).

2.2.3 Spectral-temporal feature extraction

The goal of feature extraction is to obtain features from denoised EEG signals. First of all, in order to form a dataset with a large population, all channels of the denoised EEG signals are divided into small segments synchronously.

Then, each channel in each segment is decomposed using 1-D SWT decomposition into four coefficients corresponding to four frequency bands: f_1 (0.5-4 Hz), f_2 (4-8 Hz), f_3 (8-16 Hz) and f_4 (16-32 Hz). Those coefficients contain information in both time domain and frequency domain, thus is suitable to obtain spectral-temporal features.

After that, the descriptive statistical analysis method is used to extracted features from the decomposed coefficients in the previous step. Nine widely used descriptive statistical features, namely, *median* (*med*), *standard deviation* (*std*), *mean* (*me*), *mode* (*mo*), *interquartile range* (*iqr*), *skewness* (*ske*), *kurtoses* (*kur*), *first quartile* (Q_1) and *third quartile* (Q_3) are extracted from each coefficient. To reduce the impact of individual outliers, the descriptive statistical features *maximum* and *minimum* were not adopted in this work.

Through this way, a discrete feature space with 3 dimensions, namely, channel (*Ch*), frequency bands (*FB*) and descriptive statistical feature (*DSF*), are formed, as shown in Fig. 2.

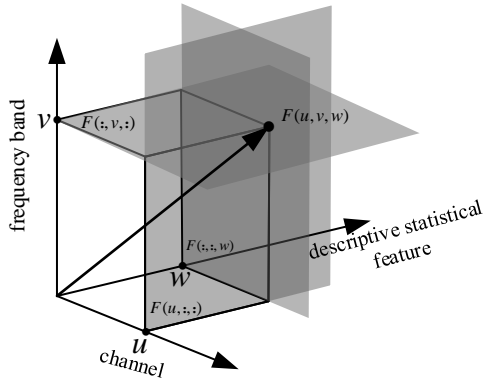


Fig. 2 Three-dimensional discrete feature space

The value ranges of the three dimensions are shown in the following lists,

$$Ch = \{F_{p1}, F_{p2}, F_7, F_3, F_z, F_4, F_8, T_3, C_3, C_z, C_4, T_4, T_5, P_3, P_z, P_4, T_6, O_1, O_2\}, \quad (1)$$

$$FB = \{f_1, f_2, f_3, f_4\}, \quad (2)$$

$$DSF = \{med, std, me, mo, iqr, ske, kur, Q_1, Q_3\}. \quad (3)$$

The discrete feature space is denoted as $F \in \mathbb{R}^{U \times V \times W}$, where U denotes the number of channels; V denotes the number of frequency bands; W denotes the number of descriptive statistical feature, and then

$$n = U \times V \times W \quad (4)$$

denotes the total number of features extracted from each segment. As shown in Fig. 2, $F(u, v, w)$ ($u \in [1, U]$, $v \in [1, V]$, $w \in [1, W]$) is a point or a specific feature in the discrete feature space F , representing the w th statistical feature extracted from the v th frequency band of the u th channel. Fig. 2 also demonstrates that the feature subsets denoted as $F(u, :, :)$, $F(:, v, :)$ and $F(:, :, w)$ are planes in the discrete feature space, where the symbol semicolon ':' stands for all elements in corresponding dimension.

After extracting features from all segments, a dataset $[X|y]$ is generated. The input matrix $X \in \mathbb{R}^{m \times U \times V \times W}$ and the output vector $y \in \mathbb{R}^m$, where m denote the total number of segments. The i th ($i = 1, \dots, m$) sample in this dataset is $[X(i, :, :, :)|y_i]$, where $y_i \in \{0, 1\}$, and '1' means MCI samples, '0' means elderly control samples. $X(i, :, :, :)$ denotes all features extracted from the i th sample (segment). So,

$$X(i, :, :, :) = [X(i, 1, 1, 1), X(i, 1, 1, 2), \dots, X(i, u, v, w), \dots, X(i, U, V, W)]. \quad (5)$$

2.2.4 Feature selection based on 3-D evaluation algorithm

Too many features might lead to bias and over-fitting for MCI classification. Intensive computation and time overheads are other possible problems. Moreover, some of the extracted features might be correlated and therefore provide no new information, thus need to be removed. Compared with those algorithms, such as principal component analysis (PCA) and linear discriminant analysis (LDA), which derive new features from existing features, we proposed a wrapper method, named 3-D evaluation algorithm, to choose an optimal feature subset from the existing feature space F to maintain the interpretability of features.

The basic idea of wrapper algorithm is that the classifier is considered as a black box and its performance is used to select the optimal feature subset. Based on the 3-D discrete feature space F established in Section 2.2.3, the proposed 3-D evaluation algorithm evaluate the elements in three dimensions specified in Equation (1), (2), and (3) individually and incrementally.

More specifically, the pseudocode of individual and incremental assessment on channel dimension is described in **Algorithm 1**. The inputs are the dataset $[X|y]$ formed after feature extraction, the feature space F and a series of scalars, m, U, V, W . After initialization, we conduct individual channel assessment via evaluating the performance of SVM classifier on the feature subset $F(u, :, :)$ ($u \in [1, U]$) (**step 6-10**), then seek out *OptFeaSub*, *MaxAcc*, *MaxSens*, *MaxSpec* (**step 11-14**), where *OptFeaSub* denotes the selected optimal feature subset, and *MaxAcc*, *MaxSens*, *MaxSpec* denote the corresponding accuracy,

Algorithm 1 Individual and incremental evaluation on channel dimension

```

1: Input:  $X \in \mathbb{R}^{m \times U \times V \times W}$ ,  $y$ ,  $F \in \mathbb{R}^{U \times V \times W}$ ,  $m$ ,  $U$ ,  $V$ ,  $W$ 
2: Output:  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   % inationalization
3:  $OptFeaSub = \emptyset$ ,  $MaxAcc=0$ ,  $MaxSens=0$ ,  $MaxSpec=0$ ;
   % feature subset selection on channel dimension
4: for each  $u \in [1, U]$  do
5:   % Individual channel assessment
6:    $X_{eva} = X[:, u, :, :]$ ,  $y_{eva} = y$ ;
7:    $X_{eva} = \text{reshape}(X_{eva}, [m, 1 * V * W])$ 
8:   Randomly separate dataset  $[X_{eva} \mid y_{eva}]$  into three
   parts:  $[X_{train} \mid y_{train}]$ ,  $[X_{val} \mid y_{val}]$ ,  $[X_{test} \mid y_{test}]$  in
   a ratio of 0.6:0.2:0.2;
9:   Train a SVM classifier using  $[X_{train} \mid y_{train}]$ 
10:  Evaluate the trained SVM classifier using  $[X_{val} \mid y_{val}]$ ,
   and calculate  $Acc$ ,  $Sens$ ,  $Spec$ ;
11:  if  $Acc > MaxAcc$  then
12:     $OptFeaSub = F(u, :, :)$ ;
13:     $MaxAcc = Acc$ ,  $MaxSens = Sens$ ,  $MaxSpec = Spec$ ;
14:  end if
   % Incremental channel assessment
15:   $X_{eva} = X[:, 1 : u, :, :]$ ,  $y_{eva} = y$ ;
16:   $X_{eva} = \text{reshape}(X_{eva}, [m, u * V * W])$ ;
17:  Randomly separate dataset  $[X_{eva} \mid y_{eva}]$  into three
   parts:  $[X_{train} \mid y_{train}]$ ,  $[X_{val} \mid y_{val}]$ ,  $[X_{test} \mid y_{test}]$  in
   a ratio of 0.6:0.2:0.2;
18:  Train a SVM classifier using  $[X_{train} \mid y_{train}]$ ;
19:  Evaluate the trained SVM classifier using  $[X_{val} \mid y_{val}]$ ,
   and calculate  $Acc$ ,  $Sens$ ,  $Spec$ ;
20:  if  $Acc > MaxAcc$  then
21:     $OptFeaSub = F(1 : u, :, :)$ ;
22:     $MaxAcc = Acc$ ,  $MaxSens = Sens$ ,  $MaxSpec = Spec$ ;
23:  end if
24: end for
25: return  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ ;

```

sensitivity and specificity achieved on this optimal feature subset. The definitions of accuracy, sensitivity and specificity are specified in Section 2.3. Then, similarly, conduct incremental channel assessment on the feature subset $F(1 : u, :, :)$, where $1 : u$ means all elements between the 1st element to the u th element on the channel dimension. As u changes from 1 to U , the number of features in feature subset $F(1 : u, :, :)$ increase incrementally. At last, $OptFeaSub$, $MaxAcc$, $MaxSens$ and $MaxSpec$ are returned as outputs.

The pseudocode of the whole 3-D evaluation algorithm is specified in **Algorithm 2**. After evaluation in channel dimension, we conduct evaluation in the frequency band (FB) and descriptive statistical feature (DSF) dimensions in order, as described in **step 10-23** of **Algorithm 2**. Individual assessment on feature subset $F(:, v, :)$ ($v \in [1, U]$), $F(:, :, w)$ ($w \in [1, W]$) and also incremental assessment on feature subset $F(:, 1 : v, :)$ and $F(:, :, 1 : w)$ are carried out.

It should be noted that, when evaluating the latter two

dimensions, the initial values of $OptFeaSub$, $MaxAcc$, $MaxSens$, and $MaxSpec$ should be the returned values from the previous dimension evaluation, instead of \emptyset or 0, as described in **step 10** and **17** of **Algorithm 2**.

Algorithm 2 3-D evaluation algorithm

```

1: Input:  $X \in \mathbb{R}^{m \times U \times V \times W}$ ,  $y$ ,  $F \in \mathbb{R}^{U \times V \times W}$ ,  $m$ ,  $U$ ,  $V$ ,  $W$ 
2: Output:  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   % initialization
3:  $OptFeaSub = \emptyset$ ,  $MaxAcc=0$ ,  $MaxSens=0$ ,  $MaxSpec=0$ ,  $seed=1$ ;
   % feature subset selection on channel dimension
4: for each  $u \in [1, U]$  do
5:   Individual channel assessment on feature subset  $F(u, :, :)$ 
6:   Seek out  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   as described in step 11-14 of Algorithm 1
7:   Incremental channel assessment on feature subset  $F(1 : u, :, :)$ 
8:   Seek out  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   as described in step 20-23 of Algrothm 1
9: end for
   % feature subset selection on FB dimension
10: Initiate  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$  as
   the result of step 8;
11: for each  $v \in [1, V]$  do
12:   Individual FB assessment on feature subset  $F(:, v, :)$ 
13:   Seek out  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   similarly as step 6
14:   Incremental FB assessment on feature subset  $F(:, 1 : v, :)$ 
15:   Seek out  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   similarly as step 8
16: end for
   % feature subset selection on DSF dimension
17: Initiate  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$  as
   the result of step 15;
18: for each  $w \in [1, W]$  do
19:   Individual DSF assessment on feature subset  $F(:, :, w)$ 
20:   Seek out  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   similarly as step 6
21:   Incremental DSF assessment on feature subset  $F(:, :, 1 : w)$ 
22:   Seek out  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ 
   similarly as step 8
23: end for
24: return  $OptFeaSub$ ,  $MaxAcc$ ,  $MaxSens$ ,  $MaxSpec$ ;

```

2.2.5 Classification based on SVM

SVM has been widely used in pattern recognition and regression due to its computational efficiency and good generalization performance [29, 30]. The core of the SVM algorithm for binary classification is mapping the input data into a linearly separable space using a kernel function. It also applies a minimization algorithm to minimize the objective function and maximize the margins between two classes at the same time. SVM is stable and effective at dealing with the small or medium scale of data, because

only support vectors are used to construct the separating hyperplane. Considering the scale of our dataset, SVM is chosen as the classification algorithm.

In order to further demonstrate the effectiveness of the extracted and selected optimal feature subset, two well-known machine learning classifiers, Decision Tree and KNN, are also adopted to verify the classification performance of the proposed framework.

2.3 Performance evaluation

To evaluate the proposed algorithm and compare it with other state-of-the-art algorithms, three widely used metrics in this domain based on the confusion matrix are adopted. A confusion matrix C is a square matrix whose size k is equal to the total number of classes to be classified. The element $C(i, j)$ is the count of samples known to be in class i (true condition) and predicted to be in class j (predicted condition), where $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, k$.

The confusion matrix for binary classification is shown in Table 1. The element C_{11} is also known as *true negatives* (TN), which indicates the count of observations predicted to be negative and also known to be negative in the true condition. Similarly, we can get the meaning of false positives (FP), false negatives (FN) and true positives (TP). Obviously,

$$Total\ Population = TN + FP + FN + TP. \quad (6)$$

Table 1 Confusion matrix for binary classification

Total Population	predicted negatives	predicted positives
true negatives	TN	FP
true positives	FN	TP

The three metrics to evaluate the performance of a classifier are defined as Equation (7), (8), (9) respectively^[31]. Sensitivity (*Sens*) is a measure of the capacity to correctly identify *true positives*. Specificity (*Spec*) reflects the capacity to correctly identify *true negatives* and accuracy (*Acc*) is the proportion of correct classified instances.

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

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

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \quad (9)$$

Obviously, accuracy is the average of sensitivity and specificity, so we only take accuracy into account when comparing the performance of two algorithms. Sensitivity and specificity are used as a reference to determine whether an algorithm is biased towards a single category.

3 Results

All the experimental works are simulated and implemented under the MATLAB 2018a software environment. Some built-in functions are called from the ‘Wavelet Toolbox’ and ‘Statistics and Machine Learning Toolbox’. The following parts of this section describe the parameter settings and the experiment results.

3.1 Results of SWT-based denoising

3.1.1 Setting of SWT decomposition

SWT decomposition and reconstruction are implemented by calling the built-in functions ‘swt’ and ‘iswt’ from the ‘Wavelet Toolbox’ of Matlab. The mother wavelet basis function used in this work is *sym9*, which is chosen from the *Symlets* mother wavelet family. Because *sym9* is reported to be suitable for denoising, decomposition, reconstruction, and sub-band feature extraction^[32]. Since the sampling rate of the collected EEG signals is 256 Hz, the decomposition level, following the Nyquist criterion, is set to 8 to obtain the coefficients with the appropriate frequency band. Fig. 3 demonstrates the level-8 1-D SWT decomposition process. SWT applies low-pass and high-pass filters to decompose the input signals and produces two time-series sequences, namely, approximation coefficient A_i and detail coefficient D_i at level i . The two coefficients have the same length as the input signal to be composed. In Fig. 3, SR denotes the sampling rate of the EEG signals; and H_i-D and L_o-D denote the high-pass and low-pass decomposition separately. Level 6 and 7 are omitted.

3.1.2 Results of denoising

As described in Section 2.2.2, in denoising stage, after decomposing via SWT, we keep the components with the frequency of 0.5-32 Hz, and remove components with other frequency.

First, we call ‘swt’ function to decompose each channel of the raw EEG signals to obtain coefficients: A_8 (0-0.5 Hz), D_2 (32-64 Hz) and D_1 (64-128 Hz). Then, we reconstruct them to get the time series signals as the to be removed noise, by calling ‘iswt’ function in Matlab. Finally, the reconstructed noise is subtracted from the original EEG signals to get the denoised EEG signals.

Fig. 4 demonstrates the 19-channel EEG signals of an MCI subject in the time domain. Subplot (a) shows the raw EEG signals, and subplot (b) is the denoised EEG signals.

Fig. 5 illustrates the denoising process in the frequency domain. The signals in Fig. 5 are transformed from time series signals by Fourier transform. The different coloured lines represent different channels. Fig. 5 (a) shows the raw EEG signals in the frequency domain. It is obvious that the frequency components from 0 to 0.5 Hz are extraordinarily large, which is chiefly caused by baseline drift. Furthermore, there is a peak amplitude of 50 Hz, which is mainly caused by power line interference. Fig. 5 (b) shows the low-frequency noise to be removed in 0-0.5 Hz. The partially enlarged detail in Fig. 5 (b) shows that the components with the frequency less than 0.5 Hz will be eliminated. Fig. 5 (c) presents the high-frequency noise to be removed in 32-128 Hz including the obvious 50 Hz power line interference. Since there are no ideal bandpass filters, some components < 32Hz are removed too. However, their amplitudes are too small to interfere with the results, as shown in the partially enlarged detail of Fig. 5 (c). Fig. 5 (d) are the denoised signals after removing signals in (b) and (c) from (a). As a schematic diagram, Fig. 5 is plotted based on a 4-second segment split from an MCI object.

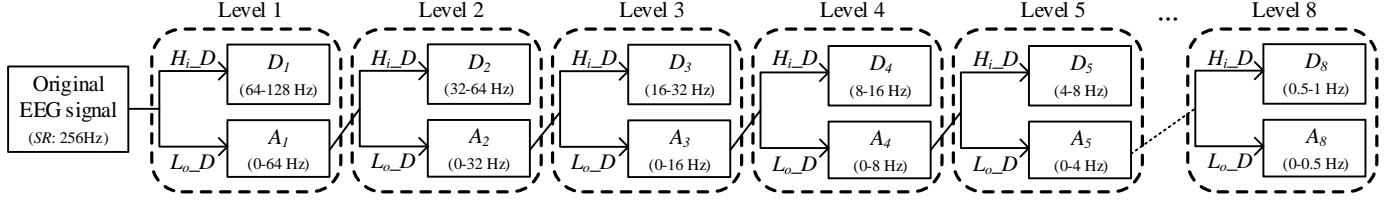


Fig. 3 A level-8 SWT decomposition diagram of EEG signal

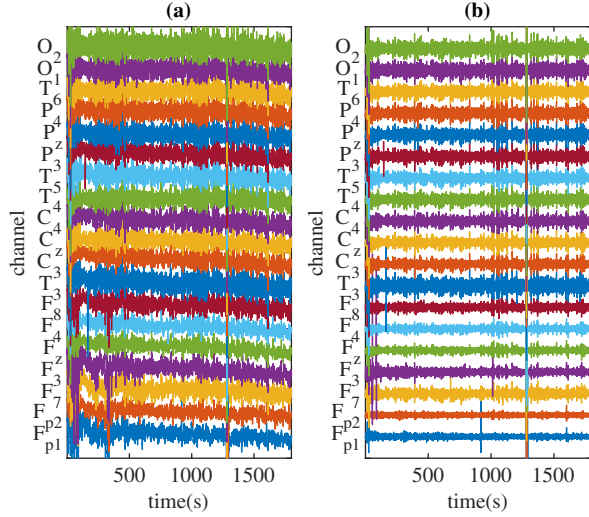


Fig. 4 The 19-channel EEG signals of an MCI subject in time domain. (a) The raw EEG signals; (b) The denoised EEG signals.

3.2 Results of spectral-temporal feature extraction

As mentioned in Section 2.2.3, we do segmentation before feature extraction. The length of the sliding window and the proportion of overlap between two neighbour segments are two factors which need to be considered in the segmentation process. Considering the basic information of our dataset (30 minutes' duration, 22 subjects and a quite large number of features will be extracted), we tried several possible values, i.e., 0.5s, 1s, 2s, 4s, of window length. Finally, the window length is fixed at 2 seconds in this work due to the performance of MCI detection. The overlapping rate is set to 0% to avoid bias result. In this case, the stride of the sliding window is equal to the length of the window. So, the total number of segments (samples) in our experiments is

$$m = \left\lceil \left\lfloor \frac{\text{signal length} - \text{window length}}{\text{stride}} \right\rfloor + 1 \right\rceil \times n_s, \quad (10)$$

where, *signal length* is the total lasting seconds of a signal, which equals to 1800 seconds in our experiments; *window length* means total lasting seconds of a segment, which equals to 2 seconds; *stride*=2 seconds means the in-

terval between two segment starting points and $n_s=22$ denotes the number of subjects. According to Equation (1), $m = 19800$.

For each segment, every channel specified in Equation (1) is decomposed into four coefficients: A_5 (0.5-4Hz), D_5 (4-8Hz), D_4 (8-16Hz) and D_3 (16-32Hz) via 1-D SWT. The mother wavelet basis function is still set to *sym9* and 5 is the appropriate decomposition level to achieve the desired frequency resolution, as shown in Fig. 3.

Fig. 6 demonstrates the spectral-temporal characteristics of Channel O_1 , which is picked from a segment of an MCI subject after denoising. Fig. 6 (a) presents 5 signals in time domain, namely, the denoised channel O_1 and its 4 coefficients of D_3 , D_4 , D_5 , A_5 . While Fig.6 (b) shows those 5 signals in frequency domain.

As shown in Fig. 6, the decomposed coefficients reflect the spectral-temporal characteristics of EEG signals. We employ descriptive statistical analysis to extract features from those coefficients. Specifically, 9 descriptive statistical features listed in Equation (3) are extracted from each coefficient by calling the corresponding built-in functions in Matlab. Therefore, the total number of features extracted from each segment $n=19 \times 4 \times 9=684$ according to Equation (4), where 19 denotes the number of channels, 4 indicates the number of frequency bands, and 9 means the number of descriptive statistical features.

3.3 Results of 3-D evaluation algorithm for feature selection

The proposed 3-D evaluation feature selection method is a kind of wrapper model. An SVM classifier is used as a black box and its performance is used to select the optimal feature subset. We implement the SVM classifier using the built-in functions in 'Statistics and Machine Learning Toolbox' of Matlab. Particularly, we create a 'ClassificationSVM' object as the binary classifier, and train it with the 'fitsvm' function, then use 'predict' function to make prediction with the trained SVM classifier. In order to simplify calculation, the 'KernelFunction' is set as 'polynomial' and 'PolynomialOrder' is set as 2. 'KernelScale' is set as 'auto', 'BoxConstraint' is set as 2, and 'Standardize' is set as true. Other parameters is set as default. The SVM models share the same settings in step 3 (3-D evaluation) and step 4 (classification) of the proposed framework, as shown in Fig. 1.

As described in **Algorithm 1** and **Algorithm 2**, we evaluate the feature subset in channel dimension firstly. Fig.7(a) and the left part of Table 2 show the results of indi-

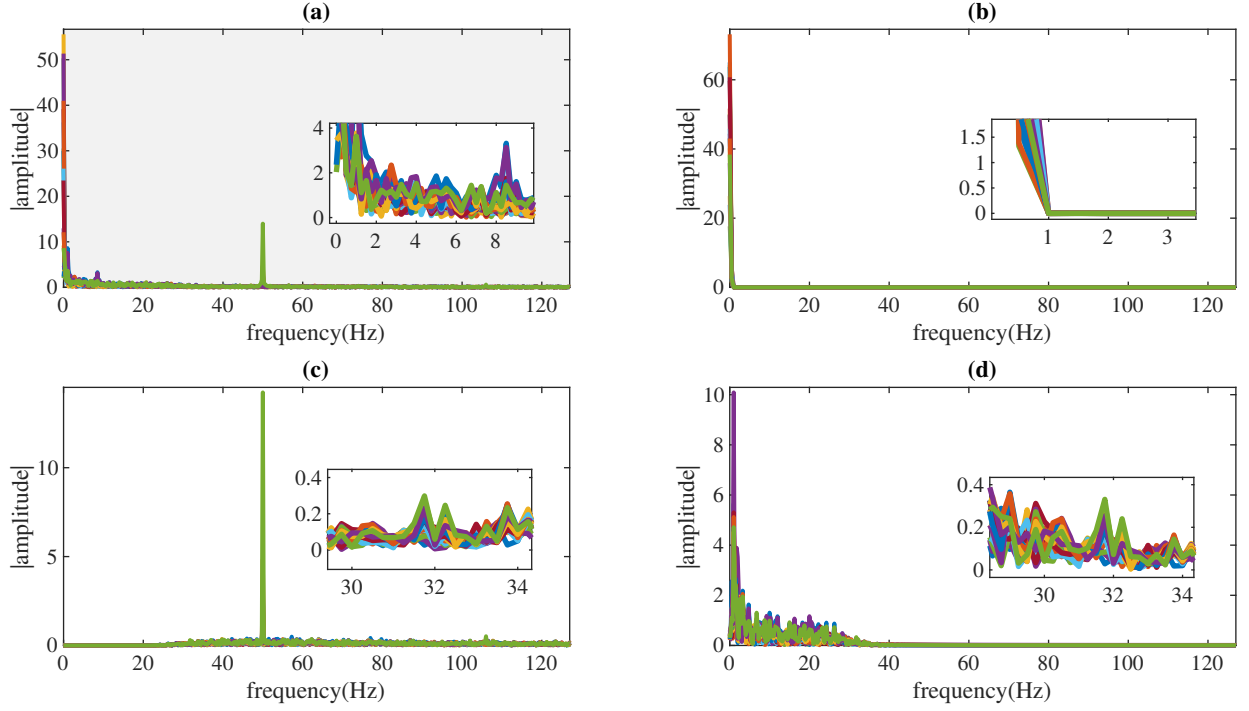


Fig. 5 The frequency domain EEG signals with 19 channels of an MCI subject. (a) The raw EEG; (b) The low frequency noise; (c) The high frequency noise; (d) The denoised signals.

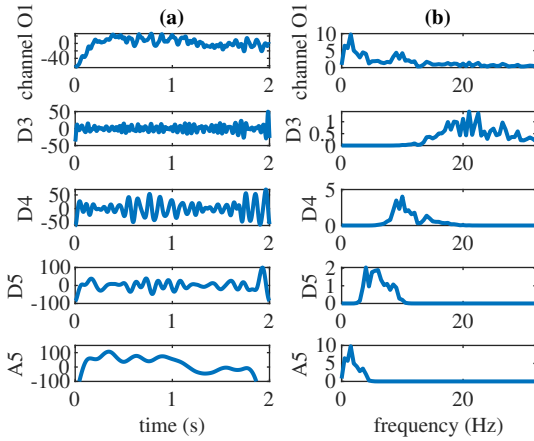


Fig. 6 Channel O_1 and its four coefficients. (a) time domain signals of channel O_1 and D_3 , D_4 , D_5 , A_5 ; (b) frequency domain signals of channel O_1 and D_3 , D_4 , D_5 , A_5 .

vidual assessment. The solid lines in Fig.7 gives the performance on training set while the dash lines is on validation set. As emphasized with bold text in Table 2, after individual evaluation, $OptFeaSub = F(13, :, :)$ and $MaxAcc = 75.28\%$ on the corresponding validation set. Obviously, using features extracted from any single channel can not achieve good MCI detection performance. When evaluating incrementally on channel dimension, as marked by bold text in the right part of Table 2, we get the best performance on $F(1 : 19, :, :)$, which means that $OptFeaSub = F(1 : 19, :, :)$ and $MaxAcc = 94.74\%$. Since we have 19 channel in total, the evaluation results on channel dimension indicates that the best performance is achieved on the whole feature set.

The evaluation results on frequency band dimension are show in Fig. 8 and Table 3. We emphasized the best performance of individual assessment and incremental assessment with bold-type, and both of them are no better than $MaxAcc = 94.74\%$. According to the description in **Algorithm 2**, still $OptFeaSub = F(1 : 19, :, :)$ and $MaxAcc = 94.74\%$.

As for the evaluation results on the descriptive statistical feature dimension, we demonstrated in Fig. 9 and Table 4. The best performance of individual evaluation is marked in bold-type, and it is the best one so far. So, $OptFeaSub = F(:, :, 5)$ and $MaxAcc = 97.45\%$. The best performance of incremental evaluation is also marked in bold-type, and it is deprecated because it is less than 97.45%.

As a conclusion, the final optimal feature subset selected is $OptFeaSub = F(:, :, 5)$, which means the 5th descriptive

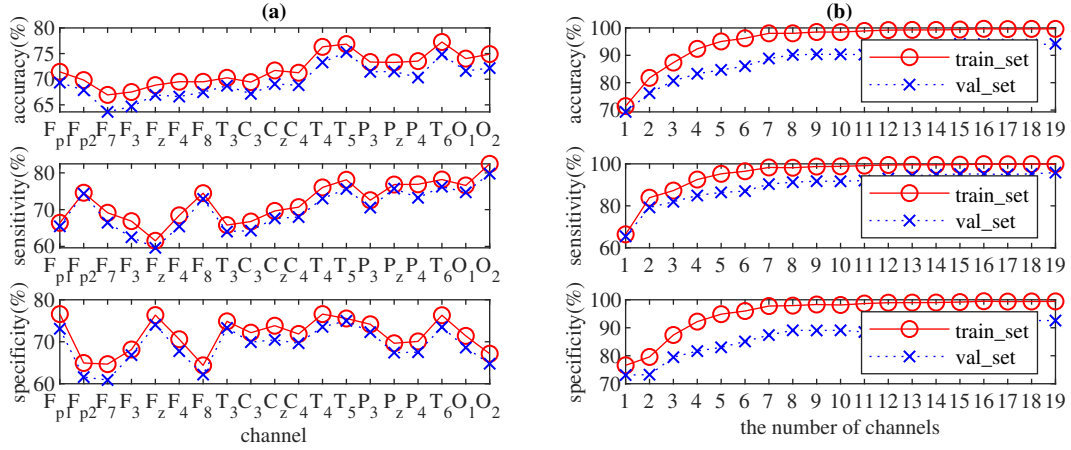


Fig. 7 Evaluation results on channel dimension. (a) Individual assessment (b) Incremental assessment.

Table 2 Numerical results of feature subset evaluation on channel dimension (performance on validation set)

Individual evaluation				Incremental evaluation			
feature subset	accuracy	sensitivity	specificity	feature subset	accuracy	sensitivity	specificity
$F(1, :, :)$	69.44%	66.14%	72.65%	$F(1, :, :)$	69.44%	66.14%	72.65%
$F(2, :, :)$	68.10%	73.06%	63.28%	$F(1 : 2, :, :)$	76.24%	78.91%	73.64%
$F(3, :, :)$	63.52%	66.75%	60.39%	$F(1 : 3, :, :)$	80.13%	80.91%	79.37%
$F(4, :, :)$	64.86%	63.62%	66.07%	$F(1 : 4, :, :)$	83.37%	84.61%	82.16%
$F(5, :, :)$	66.48%	58.95%	73.79%	$F(1 : 5, :, :)$	85.06%	86.40%	83.76%
$F(6, :, :)$	66.96%	65.67%	68.21%	$F(1 : 6, :, :)$	85.97%	86.92%	85.05%
$F(7, :, :)$	67.42%	73.27%	61.73%	$F(1 : 7, :, :)$	88.52%	89.99%	87.10%
$F(8, :, :)$	68.20%	64.19%	72.10%	$F(1 : 8, :, :)$	90.34%	91.59%	89.14%
$F(9, :, :)$	67.52%	64.70%	70.25%	$F(1 : 9, :, :)$	90.34%	91.59%	89.14%
$F(10, :, :)$	68.63%	67.16%	70.05%	$F(1 : 10, :, :)$	90.60%	91.94%	89.29%
$F(11, :, :)$	68.63%	67.98%	69.26%	$F(1 : 11, :, :)$	90.65%	92.30%	89.04%
$F(12, :, :)$	73.66%	73.32%	73.99%	$F(1 : 12, :, :)$	91.96%	93.28%	90.68%
$F(13, :, :)$	75.28%	74.81%	75.73%	$F(1 : 13, :, :)$	93.91%	94.61%	93.22%
$F(14, :, :)$	71.54%	69.98%	73.04%	$F(1 : 14, :, :)$	93.58%	94.61%	92.58%
$F(15, :, :)$	71.71%	75.63%	67.91%	$F(1 : 15, :, :)$	93.98%	94.92%	93.07%
$F(16, :, :)$	69.89%	73.22%	66.67%	$F(1 : 16, :, :)$	94.19%	94.97%	93.42%
$F(17, :, :)$	74.60%	75.89%	73.34%	$F(1 : 17, :, :)$	94.14%	95.02%	93.27%
$F(18, :, :)$	71.99%	75.68%	68.41%	$F(1 : 18, :, :)$	94.08%	95.18%	93.02%
$F(1 : 19, :, :)$	72.65%	80.71%	64.82%	$F(1 : 19, :, :)$	94.74%	95.90%	93.62%

Table 3 Numerical results of feature subset evaluation on frequency band dimension (performance on evaluation set)

Individual evaluation				Incremental evaluation			
feature subset	accuracy	sensitivity	specificity	feature subset	accuracy	sensitivity	specificity
$F(:, 1, :)$	75.53 %	80.66 %	70.55 %	$F(:, 1, :)$	75.53 %	80.66 %	70.55 %
$F(:, 2, :)$	79.90 %	84.09 %	75.83 %	$F(:, 1 : 2, :)$	82.86 %	85.79 %	80.02 %
$F(:, 3, :)$	83.37 %	86.51 %	80.32 %	$F(:, 1 : 3, :)$	89.13 %	91.64 %	86.70 %
$F(:, 4, :)$	88.62 %	89.12 %	88.14 %	$F(:, 1 : 4, :)$	94.62 %	95.84 %	93.42 %

Table 4 Numerical results of feature subset evaluation on descriptive statistical feature dimension

Individual evaluation				Incremental evaluation			
feature subset	accuracy	sensitivity	specificity	feature subset	accuracy	sensitivity	specificity
$F(:, :, 1)$	68.30 %	76.40 %	60.44 %	$F(:, :, 1)$	68.30 %	76.40 %	60.44 %
$F(:, :, 2)$	96.94 %	97.43 %	96.46 %	$F(:, :, 1 : 2)$	89.79 %	92.00 %	87.64 %
$F(:, :, 3)$	64.61 %	77.37 %	52.22 %	$F(:, :, 1 : 3)$	88.47 %	90.30 %	86.70 %
$F(:, :, 4)$	91.94 %	93.89 %	90.03 %	$F(:, :, 1 : 4)$	90.34 %	92.25 %	88.49 %
$F(:, :, 5)$	97.45 %	97.74 %	97.16 %	$F(:, :, 1 : 5)$	94.41 %	95.43 %	93.42 %
$F(:, :, 6)$	66.56 %	70.81 %	62.43 %	$F(:, :, 1 : 6)$	93.20 %	94.56 %	91.88 %
$F(:, :, 7)$	64.11 %	68.55 %	59.79 %	$F(:, :, 1 : 7)$	92.75 %	94.30 %	91.23 %
$F(:, :, 8)$	97.14 %	97.38 %	96.91 %	$F(:, :, 1 : 8)$	93.63 %	94.97 %	92.33 %
$F(:, :, 9)$	96.79 %	97.38 %	96.21 %	$F(:, :, 1 : 9)$	94.69 %	95.84 %	93.57 %

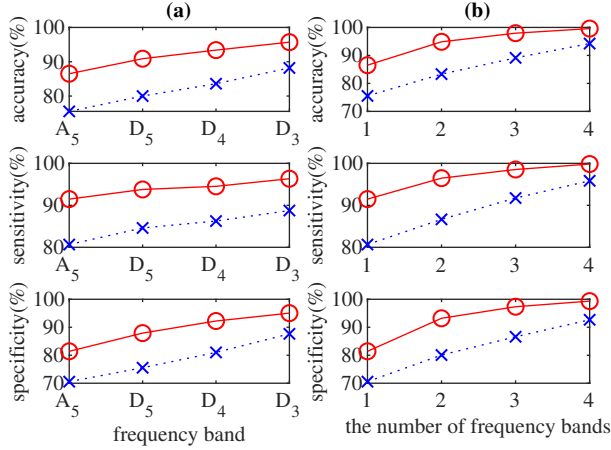


Fig. 8 Evaluation results on frequency band dimension. (a) Individual assessment. (b) Incremental assessment.

statistical feature, i.e., *interquartile range*, extracted from all frequency bands listed in Equation (2) and all channels listed in Equation (1). The total number of features in the extracted optimal feature subset is,

$$n_{opt} = U \times V \times 1 = 19 \times 4 \times 1 = 72. \quad (11)$$

Considering the total number of features $n=684$ calculated by (4), the suppression ratio of the proposed 3-D evaluation feature selection algorithm is

$$\text{suppression ratio} = \frac{n_{opt}}{n} \times 100\% = \frac{72}{684} \times 100\% = 11.11\%. \quad (12)$$

3.4 Results of classification

Finally, we test the performance of the proposed framework on the test set and compare it with other MCI detection algorithms based on EEG signals.

As described in section 3.3, $F(:, :, 5)$ is the final optimal feature subset. So, in this stage, we only picked the data with feature subset $F(:, :, 5)$ to test the classification performance. In order to get an unbiased results, we test the performance on the test set $[X_{test} | y_{test}]$, which has not been seen by the classifier when training.

Besides SVM, we also implement two well-known s^[33] (Decision Tree and KNN) to verify the effectiveness of the selected feature subset in the classification stage. All those three classifiers are implemented by the built-in functions in ‘Statistics and Machine Learning Toolbox’ of Matlab. The SVM classifier keeps the same setting described in Section 3.3. We use the built-in function *fitcknn* to optimize the hyperparameters automatically for KNN classifier. As a result, the optimised *NumNeighbors* is set to 5 and the *Distance* is set as *seuclidean*. Other parameters keep the default setting. Similarly, we use *fitctree* to implement and optimize the Decision Tree classifier and the parameter *MinLeafSize* is set to 9.

The classification results are shown in the last three rows of Table 5. We can see that, with the right features *OptFeaSub* selected by the proposed framework, all the three classifiers can achieve superior performance compared with other similar works^[2, 34, 35, 36, 37] reported recently, especially the work in^[2], which uses the same dataset with us. Among that three classifiers, the SVM classifier is with a narrow lead.

4 Discussion

4.1 Effectiveness of the spectral-temporal feature extraction method

The spectral-temporal characteristics of each channel is reflected on the four decomposed coefficients, i.e. D_3 , D_4 , D_5 , A_5 . The present study extract descriptive statistical features from those coefficients, so the extracted features contain information in both time domain and frequency domin.

As shown in the last row of the right parts of Table 2, 3 and 4, $F(1 : 19, :, :) = F(:, 1 : 4, :) = F(:, :, 1 : 9)$ denotes the whole feature space F . The accuracy is located in the range of [94.62%, 94.74%] on the whole feature space F , which is outperformed other algorithms listed in Table 5. Thus, the proposed spectral-temporal feature extraction strategy is quite effective.

There is no large deviation between the sensitivity and specificity of the proposed framework, which indicates that the strategy of choosing a balanced data set is effective in reducing the inconsistency caused by the data structure. Moreover, the accuracy achieved on the whole feature set F is slightly different, ranging from 94.62% to 94.74% as

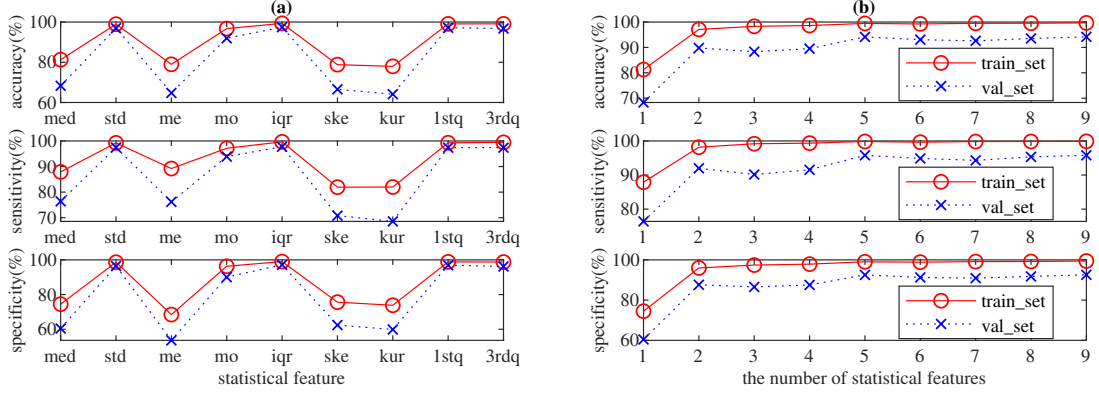


Fig. 9 Evaluation results on the statistical feature dimension. (a) Individual assessment (b) Incremental assessment.

Table 5 Performance comparison

algorithms	accuracy	sensitivity	specificity	classification
NF-KNN [2]	88.89%	100%	83.33%	MCI vs. healthy control
Christoph Lehmann et al. [34]	88.5%	89%	88%	MCI vs. healthy control
Joseph C.McBride et al. [35]	92.59%	100%	84.61%	MCI vs. healthy control
Paolo M.Rossini et al. [36]	93.46%	95.87%	91.06%	MCI vs. healthy control
wavelet+SVM [37]	91.7%	91.7%	91.7%	MCI vs. healthy control
<i>OptFeaSub</i> + SVM	96.94%	96.89%	96.99%	MCI vs. healthy control
Decision Tree	95.47%	95.38%	95.55%	MCI vs. healthy control
KNN	96.89%	97.25%	96.54%	MCI vs. healthy control

shown in Table 2, 3 and 4. The reason behind is that, when evaluating on different dimensions, the order we incrementally add into the feature subset is different, thus the initialization state of SVM classifier is different. However, this nuance does not affect the consistency of results.

4.2 Discussion on 3-D evaluation algorithm

Ranking all channels in descending order of accuracy based on the individual evaluation results in the channel dimension, the result is: $T_5, T_6, T_4, O_2, O_1, P_z, P_3, P_4, F_{p1}, C_z, C_4, T_3, F_{p2}, C_3, F_8, F_4, F_z, F_3, F_7$. Considering the location of the 19 channels on the scalp as shown in Fig. 10, we discover that the temporal and occipital areas are more effective for MCI detection than the frontal and central areas. Similarly, the frequency band sequence in descending order is D_3, D_4, D_5, A_5 , which means coefficients of higher frequency bands are more effective than lower ones in the frequency scope of 0.5-32 Hz for MCI detection. The descending order of the descriptive statistical feature sequence is *iqr*, *1stq*, *std*, *3rdq*, *mode*, *median*, *ske*, *mean* and *kur* in the statistical feature dimension. Fig. 9 (a) and the left part of Table 4 further demonstrates that Q_1 , *std* and Q_3 are also quite effective in differentiating subjects with an MCI and those who are cognitively healthy.

The results of the individual evaluation on channel and frequency band dimensions show that a single channel or a single frequency band can hardly achieve good performance in MCI detection. As for descriptive statistical feature dimension, the optimal feature subset is $F(:, :, 5)$ which contains only one descriptive statistical feature, i.e. *iqr*, but

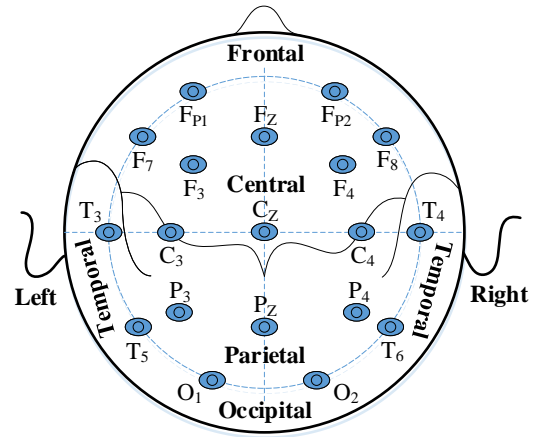


Fig. 10 The international 10-20 system of electrode placement.

includes all channels and frequency bands. So, we can draw a conclusion that every channel and frequency band has unique information for MCI detection, whereas the statistical features are highly redundant and *igr* is the properly selected descriptive statistical feature in this problem.

The presented 3-D evaluation algorithm is efficient and effective for feature selection, because with a suppression ratio of 11.11%, the selected feature subset $F(:, :, 5)$ can obtain the best performance, which is even better than that on the full feature space. Conversely, with the same number of inappropriate features, for example, the feature subset of $F(:, :, 7)$, the accuracy is only 64.11%, as shown in Table 4 marked with bold-type.

4.3 Limitations and perspectives

Although the aforementioned advantages, this work also suffers from a problem of the limited dataset. The total number of subjects involved in the experiments is 22, which makes the trained model can hardly be used in non-patient specific scenes. In future work, we plan to collect more data and try the automatic feature extraction method based on deep learning algorithms. Multi-class classification between MCI, healthy control and AD patients will also be involved.

5 Conclusions

A systematic framework is proposed to identify MCI patients and elderly control subjects using resting-state EEG signals. The proposed scheme can efficiently eliminate the baseline drift and power line interference from the raw EEG signals. It also takes advantage of extracting information from both time domain and frequency domain, and a set of highly representative spectral-temporal features are extracted. Moreover, an effective feature subset is selected through the proposed 3-D evaluation algorithm. Extensive experiments were conducted based on clinical data. The results show that, compared with other similar works, our method achieves a better performance.

Acknowledgements

The first author was supported by a La Trobe University Postgraduate Research Scholarship and a La Trobe University Full Fee Research Scholarship, she was also supported by the Science Foundation of Chongqing University of Arts and Sciences Chongqing, China (No.Z2016RJ15).

References

- [1] N. Houmani, F. Vialatte, E. Gallego-Jutglà, G. Dreyfus, V.-H. Nguyen-Michel, J. Mariani, and K. Kinugawa, "Diagnosis of alzheimers disease with electroencephalography in a differential framework," *PloS one*, vol. 13, no. 3, p. e0193607, 2018.
- [2] M. Kashefpoor, H. Rabbani, and M. Barekatin, "Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features," *Journal of medical signals and sensors*, vol. 6, no. 1, p. 25, 2016.
- [3] S. Khatun, B. I. Morshed, and G. M. Bidelman, "Single channel eeg time-frequency features to detect mild cognitive impairment," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Rochester, USA: IEEE, May 2017, pp. 437–442.
- [4] V. Bibina, U. Chakraborty, R. M. Lourde, and A. Kumar, "Time-frequency methods for diagnosing alzheimer's disease using eeg: A technical review," in *Proceedings of the 6th International Conference on Bioinformatics and Biomedical Science*. New York, USA: ACM, June 2017, pp. 49–54.
- [5] N. K. Al-Qazzaz, S. H. B. Ali, S. A. Ahmad, K. Chellappan, M. Islam, J. Escudero *et al.*, "Role of eeg as biomarker in the early detection and classification of dementia," *The Scientific World Journal*, vol. 2014, pp. 1–16, 2014.
- [6] Z.-J. Yao, J. Bi, and Y.-X. Chen, "Applying deep learning to individual and community health monitoring data: A survey," *International Journal of Automation and Computing*, vol. 15, no. 6, pp. 643–655, 2018.
- [7] F. Vecchio, C. Babiloni, R. Lizio, F. D. V. Fallani, K. Bli-nowska, G. Verrienti, G. Frisoni, and P. M. Rossini, "Resting state cortical eeg rhythms in alzheimer's disease: toward eeg markers for clinical applications: a review," in *Supplements to Clinical neurophysiology*. Elsevier, 2013, vol. 62, pp. 223–236.
- [8] V. Bajaj, S. Taran, and A. Sengur, "Emotion classification using flexible analytic wavelet transform for electroencephalogram signals," *Health information science and systems*, vol. 6, no. 12, pp. 1–7, 2018.
- [9] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "An efficient framework for the analysis of big brain signals data," in *Australasian Database Conference*. Gold Coast, Australia: Springer, May 2018, pp. 199–207.
- [10] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "Eeg sleep stages analysis and classification based on weighed complex network features," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018.
- [11] P. A. Kanda, E. F. Oliveira, and F. J. Fraga, "Eeg epochs with less alpha rhythm improve discrimination of mild alzheimer's," *computer methods and programs in biomedicine*, vol. 138, pp. 13–22, 2017.
- [12] H. Garn, C. Coronel, M. Waser, G. Caravias, and G. Ransmayr, "Differential diagnosis between patients with probable alzheimers disease, parkinsons disease dementia, or dementia with lewy bodies and frontotemporal dementia, behavioral variant, using quantitative electroencephalographic features," *Journal of Neural Transmission*, vol. 124, no. 5, pp. 569–581, 2017.
- [13] S. Siuly, E. Kabir, H. Wang, and Y. Zhang, "Exploring sampling in the detection of multicategory eeg signals," *Computational and mathematical methods in medicine*, vol. 2015, pp. 1–12, 2015.
- [14] M. Buscema, E. Grossi, M. Capriotti, C. Babiloni, and P. Rossini, "The ifast model allows the prediction of conversion to alzheimer disease in patients with mild cognitive impairment with high degree of accuracy," *Current Alzheimer Research*, vol. 7, no. 2, pp. 173–187, 2010.
- [15] E. Barzegaran, B. van Damme, R. Meuli, and M. G. Knyazeva, "Perception-related eeg is more sensitive to alzheimer's disease effects than resting eeg," *Neurobiology of aging*, vol. 43, pp. 129–139, 2016.
- [16] P. Ghorbanian, D. M. Devilbiss, A. Verma, A. Bernstein, T. Hess, A. J. Simon, and H. Ashrafuon, "Identification of resting and active state eeg features of alzheimers disease using discrete wavelet transform," *Annals of biomedical engineering*, vol. 41, no. 6, pp. 1243–1257, 2013.
- [17] S.-S. Poil, W. De Haan, W. M. van der Flier, H. D. Mansvelder, P. Scheltens, and K. Linkenkaer-Hansen, "Integrative eeg biomarkers predict progression to alzheimer's disease at the mci stage," *Frontiers in aging neuroscience*, vol. 5, p. 58, 2013.

- [18] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, and Y. Zhang, "Arrhythmias classification by integrating stacked bidirectional lstm and two-dimensional cnn," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Macau, China: Springer, April 2019, pp. 136–149.
- [19] L. Trambaiolli, N. Spolaôr, A. Lorena, R. Anghinah, and J. Sato, "Feature selection before eeg classification supports the diagnosis of alzheimers disease," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2058–2067, 2017.
- [20] R. Wang, J. Wang, S. Li, H. Yu, B. Deng, and X. Wei, "Multiple feature extraction and classification of electroencephalograph signal for alzheimers' with spectrum and bispectrum," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 1, p. 013110, 2015.
- [21] A. I. Triggiani, V. Bevilacqua, A. Brunetti, R. Lizio, G. Tattoli, F. Cassano, A. Soricelli, R. Ferri, F. Nobili, L. Gesualdo *et al.*, "Classification of healthy subjects and alzheimer's disease patients with dementia from cortical sources of resting state eeg rhythms: A study using artificial neural networks," *Frontiers in neuroscience*, vol. 10, p. 604, 2017.
- [22] F. Bertè, G. Lamponi, R. S. Calabrò, and P. Bramanti, "Elman neural network for the early identification of cognitive impairment in alzheimers disease," *Functional neurology*, vol. 29, no. 1, p. 57, 2014.
- [23] S. Afrakhteh, M.-R. Mosavi, M. Khishe, and A. Ayatollahi, "Accurate classification of eeg signals using neural networks trained by hybrid population-physic-based algorithm," *International Journal of Automation and Computing*, pp. 1–15, 2018.
- [24] H. Aghajani, E. Zahedi, M. Jalili, A. Keikhosravi, and B. V. Vahdat, "Diagnosis of early alzheimer's disease based on eeg source localization and a standardized realistic head model," *IEEE journal of biomedical and health informatics*, vol. 17, no. 6, pp. 1039–1045, 2013.
- [25] I. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of eeg signals using wavelet coefficients," *Journal of neuroscience methods*, vol. 148, no. 2, pp. 113–121, 2005.
- [26] M. B. Masoud Kashefpoor, Hossein Rabbani, "EEG Signals From Normal and MCI (Mild Cognitive Impairment) Cases- MISP Datasets," 2016. [Online]. Available: <http://www.biosigdata.com/?download=eeg-signals-from-normal-and-mci-cases>
- [27] J. V. Morant, "Neurotherapies and alzheimer: a protocol oriented review," *NeuroRegulation*, vol. 4, no. 2, p. 79, 2017.
- [28] B.-T. Zhang, X.-P. Wang, Y. Shen, and T. Lei, "Dual-modal physiological feature fusion-based sleep recognition using cfs and rf algorithm," *International Journal of Automation and Computing*, vol. 16, no. 3, pp. 286–296, 2019.
- [29] S. M. Hosni, M. E. Gadallah, S. F. Bahgat, and M. S. AbdelWahab, "Classification of eeg signals using different feature extraction techniques for mental-task bci," in *2007 International Conference on Computer Engineering & Systems*. Cairo, Egypt: IEEE, Nov 2007, pp. 220–226.
- [30] F. Liu, X. Zhou, Z. Wang, J. Cao, H. Wang, and Y. Zhang, "Unobtrusive mattress-based identification of hypertension by integrating classification and association rule mining," *Sensors*, vol. 19, no. 7, p. 1489, 2019.
- [31] D. Pandey, X. Yin, H. Wang, and Y. Zhang, "Accurate vessel segmentation using maximum entropy incorporating line detection and phase-preserving denoising," *Computer Vision and Image Understanding*, vol. 155, pp. 162–172, 2017.
- [32] N. K. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Selection of mother wavelet functions for multi-channel eeg signal analysis during a working memory task," *Sensors*, vol. 15, no. 11, pp. 29 015–29 035, 2015.
- [33] S. Siuly, V. Bajaj, A. Sengur, and Y. Zhang, "An advanced analysis system for identifying alcoholic brain state through eeg signals," *International Journal of Automation and Computing*, pp. 1–11, 2019.
- [34] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L.-O. Wahlund, Y. Dodge, and T. Dierks, "Application and comparison of classification algorithms for recognition of alzheimer's disease in electrical brain activity (eeg)," *Journal of neuroscience methods*, vol. 161, no. 2, pp. 342–350, 2007.
- [35] J. C. McBride, X. Zhao, N. B. Munro, C. D. Smith, G. A. Jicha, L. Hively, L. S. Broster, F. A. Schmitt, R. J. Kryscio, and Y. Jiang, "Spectral and complexity analysis of scalp eeg characteristics for mild cognitive impairment and early alzheimer's disease," *Computer methods and programs in biomedicine*, vol. 114, no. 2, pp. 153–163, 2014.
- [36] P. M. Rossini, M. Buscema, M. Capriotti, E. Grossi, G. Rodriguez, C. Del Percio, and C. Babiloni, "Is it possible to automatically distinguish resting eeg data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy?" *Clinical Neurophysiology*, vol. 119, no. 7, pp. 1534–1545, 2008.
- [37] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. De Salvo, A. Bramanti, P. Bramanti, and M. C. De Cola, "Combining eeg signal processing with supervised methods for alzheimers patients classification," *BMC medical informatics and decision making*, vol. 18, no. 1, pp. 1–10, 2018.



Jiao Yin is a Ph.D. student at Department of Computer Science and Information Technology, La Trobe University, Australia. She received a masters degree in Vehicle Operation Engineering from Sun Yat-sen University, China, in 2013 and a bachelors degree in Automation from Northeastern University, China, in 2010. Her research areas include machine learning algorithms, data-based smart health monitoring

and disease detection.

E-mail: j.yin@latrobe.edu.au (Corresponding author)

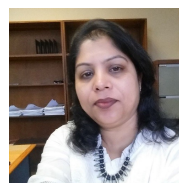
ORCID iD: 0000-0002-0269-2624



Jinli Cao is a senior lecturer at Department of Computer Science and Information Technology, La Trobe University, Australia. Dr Cao is an active researcher in the evolutionary fields of data engineering and information systems including Big data analytics, Cloud computing, Query Mining and Decision support systems. Dr Cao has published over 100 research papers in international conferences and journals such as IEEE Trans. on

Distributed and Parallel Processing, IEEE Trans. on Knowledge and Data Engineering, Journal of Future Generation Computer Systems, and the top conferences such as WWW, WISE, CAISE and DASFAA, etc. Dr Cao has 9 graduated PhD students and is supervising 2 PhD students currently.

E-mail: j.cao@latrobe.edu.au



Siuly Siuly received Ph.D. degree in Biomedical Engineering from the University of Southern Queensland, Australia, on September 2012. Currently, she is a Research Fellow with the Centre for Applied Informatics, College of Engineering and Science, Victoria University, Australia. Her current research includes Biomedical Signal Processing, Analysis and Classification, Detection and Prediction of Neurological Abnormality from Brain Signal Data (e.g.

EEG data), Brain-Computer Interface, Machine Learning, Pattern Recognition, Artificial Intelligence and Medical Data Mining. She already developed some breakthrough methods in the mentioned areas. She made significant contributions to the mentioned research areas since the beginning of her PhD in July 2008. Most recently, Dr Siuly has authored a book, titled: EEG Signal Analysis and Classification: Techniques and Applications published by Springer in December 2016. Currently she is serving an Associate Editor of a prestigious journal IEEE Transactions on Neural Systems of Rehabilitation Engineering and also as the managing editor of Health Information Science and Systems.

E-mail: Siuly.Siuly@vu.edu.au



Hua Wang received his Ph.D. degree from the University of Southern Queens-

land, Australia. He is now a full-time Professor at Victoria University. He was a professor at the University of Southern Queensland before he joined Victoria University. Hua has more than ten years teaching and working experience in Applied Informatics at both enterprise and university.

He has expertise in electronic commerce, business process modelling and enterprise architecture. As a Chief Investigator, three Australian Research Council (ARC) Discovery grants have been awarded since 2006, and 200 peer-reviewed scholar papers have been published. Six Ph.D. students have already graduated under his principal supervision.

E-mail: Hua.Wang@vu.edu.au