A New Framework for Automatic Detection of Patients With Mild Cognitive Impairment Using Resting-State EEG Signals

Siuly Siuly, *Member, IEEE*, Ömer Faruk Alçin, Enamul Kabir, Abdulkadir Şengür, Hua Wang, *Member, IEEE*, Yanchun Zhang, *Member, IEEE*, and Frank Whittaker

***Abstract*— Mild cognitive impairment (MCI) can be an indicator representing the early stage of Alzheimier's disease (AD). AD, which is the most common form of dementia, is a major public health problem worldwide. Efficient detection of MCI is essential to identify the risks of AD and dementia. Currently Electroencephalography (EEG) is the most popular tool to investigate the presenence of MCI biomarkers. This study aims to develop a new framework that can use EEG data to automatically distinguish MCI patients from healthy control subjects. The proposed framework consists of noise removal (baseline drift and power line interference noises), segmentation, data compression, feature extraction, classification, and performance evaluation. This study introduces Piecewise Aggregate Approximation (PAA) for compressing massive volumes of EEG data for reliable analysis. Permutation entropy (PE) and auto-regressive (AR) model features are investigated to explore whether the changes in EEG signals can effectively distinguish MCI from healthy control subjects. Finally, three models are developed based on three modern machine learning techniques: Extreme Learning Machine (ELM); Support Vector Machine (SVM) and K-Nearest Neighbours (KNN) for the obtained feature sets. Our developed models are tested on a publicly available MCI EEG database and the robustness of our models is evaluated by using a 10-fold cross validation method. The results show that the proposed ELM based method achieves the highest classification accuracy (98.78%) with lower execution time (0.281 seconds) and also outperforms the existing methods. The experimental results suggest that our proposed framework could provide a robust biomarker for efficient detection of MCI patients.**

***Index Terms*— Mild cognitive impairment (MCI), Alzheimer's disease (AD), electroencephalogram (EEG), piecewise aggregate approximation (PAA), auto-regressive (AR) model, permutation entropy (PE), extreme learning machine (ELM).**

I. Introduction

MILD cognitive impairment (MCI) is a degenerative neurological disorder characterized by cognitive decline which is greater than expected for an individual's age, but does not necessarily interfere with daily activities [1]. Approximately 15 to 20 percent of people aged 65 or older have MCI [2]. MCI is often an indicator of Alzheimier's disease (AD), which is characterized by progressive memory loss, diminished vocabulary, and lowered ability to execute precise motor movements, together with impaired activities of daily living [1]. AD is the most common form of dementia and accounts for 60–80% of dementia cases worldwide [3] making it a global major public health challenge. In 2019, Alzheimer's Disease International (ADI) estimated that there were over 50 million people living with dementia globally, a figure set to increase to 152 million by 2050. Someone develops dementia every three seconds and the current annual cost of dementia is estimated at US $1trillion, a figure set to double by 2030 [4]. Therefore, preventing this disease is of great importance for better health care as well as for the national economy. As there is no cure for AD and no treatment to stop or reverse its progression [5], early diagnosis and an accurate characterization of the disease progression is important to improve the quality of life of AD patients and their caregivers.

As MCI is the first symptomatic stage of AD, discovery of MCI is a dynamic arena of research. Detection of MCI is challenging because many of the symptoms overlap with those of normal ageing related decline [6]. There are various types of tests to investigate MCI such as Psychological tests (e.g. Mini-Mental State Examinations (MMSE)), blood tests, spinal fluid, neurological examination, and magnetic resonance imaging, positron emission tomography and Electroencephalography (EEG) [6]–[8]. This study explores the application of EEG

Manuscript received December 20, 2019; revised May 31, 2020 and July 2, 2020; accepted July 5, 2020. Date of publication July 31, 2020; date of current version September 7, 2020. This work was supported by Australian Research Council Linkage Project (Project ID: LP170100934). *(Corresponding author: Siuly Siuly.)*

Siuly Siuly, Hua Wang, and Yanchun Zhang are with the Institute for Sustainable Industries & Liveable Cities, Victoria University, Melbourne, VIC 3011, Australia (e-mail: siuly.siuly@vu.edu.au; hua.wang@vu.edu.au; yanchun.zhang@vu.edu.au).

Ömer Faruk Alçin is with the Department of Electrical-Electronics Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, 44210 Battalgazi, Turkey (e-mail: omer.alcin@ozal.edu.tr).

Enamul Kabir is with the School of Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: enamul.kabir@usq.edu.au).

Abdulkadir Şengür is with the Department of Electrical and Electronics Engineering, Faculty of Technology, Firat University, 23119 Elazig, Turkey (e-mail: ksengur@firat.edu.tr).

Frank Whittaker is with the Industry Partner, Victoria University, Melbourne, VIC 3011, Australia (e-mail: frank@nexusonline.com.au).

Digital Object Identifier 10.1109/TNSRE.2020.3013429

signal data to identify MCI patients from healthy control subjects because currently the EEG technique is the most popular as it is relatively inexpensive, non-invasive and portable. EEG signals representing the electrical activity of the brain at the time of recording; frequency and amplitude content vary according to the subject's biological state, mental state, age and disease process etc. [6], [9], [10].

In the last few years, many research studies have been undertaken to investigate the effects and progression of MCI and AD on EEG signals. In [7], Khatun *et al.* reported a single-channel EEG-based MCI detection methodology for regular home-based patient monitoring. They evaluated the cognitive state of the subjects by using the Montreal cognitive assessment test and extracted 590 features achieving 87.9% classification accuracy. Yin *et al.* [11] developed an integrated spectral-temporal analysis-based scheme for MCI detection using resting-state EEG signals. A three-dimensional discrete feature space was introduced with stationary wavelet transform and descriptive statistical analysis, with maximum accuracy rate of 96.94% achieved by the SVM classifier. In [8], Kashefpoor, *et al.* proposed a machine learning based methodology to discriminate MCI versus normal cases using simple spectral EEG features. Their method obtained accuracy of around 88% with frequency band features. Sharma *et al.* [12] investigated eight EEG biomarkers: power spectral density, skewness, kurtosis, spectral skewness, spectral kurtosis, spectral crest factor, spectral entropy and fractal dimension, for diagnosis of MCI patients. Control vs MCI signal classification reached accuracy ranges between 73.2% and 89.8% In [1], McBride *et al.* performed spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early AD. An average of 79.2% accuracy was achieved in MCI classification for resting eyes closed protocol. Durongbhan *et al.* [6] designed a supervised classification framework that used EEG signals to classify healthy controls and AD participants, using the K-nearest neighbor (KNN) method. From the literature, it may be seen that their performance results are not satisfactory and not sufficient for practical implementation. Most of the methods have a complex structure and are time consuming as they did not use a data efficient compression method for reliable and speedy analysis. In addition, most of the methods used to detect MCI patients from heathy control subjects are manual.

Addressing the problems, the key aim of this study is to introduce a new framework that can use EEG data to automatically detect MCI patients from healthy control subjects with higher classification accuracy and less execution time. The proposed framework consists of several tasks such as: (1) design a strategy for noise removal; (2) segment data/making sliding window; (3) introduce a scheme for EEG data compression; (4) extract and aggregate suitable features; and (5) investigate an appropriate classification model for detection of MCI patients from healthy control subjects. In this study, firstly we designed a stationary wavelet transformation (SWT) method to remove low frequency (including baseline drift) and high frequency (including power line interferences) noises. Secondly, we divided each channel data into a 2 second sliding window length (512 samples) in non-overlapping fashion. Thirdly, we introduced a Piecewise Aggregate Approximation (PAA) technique on each segmented data group to reduce data size and obtain the most important samples from the signals as EEG data which has a huge volume. Fourthly we extracted representative characteristics of the EEG signals using time dynamics-based measures such as permutation entropy (PE) and auto-regressive (AR) model features for identifying MCI biomarkers. Then finally we explored three popular machine learning models: Extreme Learning Machine (ELM); Support Vector Machine (SVM) and K-Nearest Neighbours (KNN) with the obtained features for classification of MCI patients.

The main contributions of this study are summarized as: (1) Design and validate a new framework for automatic identification of MCI patients from healthy control subjects (2) Introduce and implement the PAA approach for reducing the massive size EEG data (3) Explore efficient biomarkers of EEG to detect MCI (4) Investigate a sustainable classification model for the proposed features to differentiate the subject groups (5) Improve classification accuracy compared to existing methods (6) Build a low cost time model. To the best of our knowledge, this is the first work to apply the PAA technique to MCI EEG data to reduce data size and acquire representative points from original data.

The remainder of this paper is organised as follows. Section II presents the proposed methodology framework with a description of analysed data. Section III describes the experimental design and set up. Section IV discusses the results along with comparative analysis with existing state-of-art methods. Finally, Section V provides concluding remarks and indicates potential future work.

II. Proposed Framework

The key purpose of this study is to develop a scheme that can automatically identify MCI patients from healthy control subjects using resting-state EEG signal data. This section presents our proposed framework for dealing with EEG signal classification tasks as illustrated in Fig.1, which involves several components discussed in the following sub-sections.

A. Data Acquisition

This study uses an EEG dataset of MCI and healthy control subjects for testing our proposed framework. These data were collected from the subjects who had been admitted to cardiac catheterization units of Sina and Nour Hospitals, Isfahan, Iran [8], [13]. This data set consisted of 27 subjects' resting-state scalp EEG signals (16 cognitively healthy subjects and 11 MCI patients) aged from 60 to 77 years. Subjects with a history of substance misuse, major psychiatric disorders, serious medical disease, head trauma, and dementia were excluded.

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 [8]. 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. The collected EEG signals consists of 19 channels,

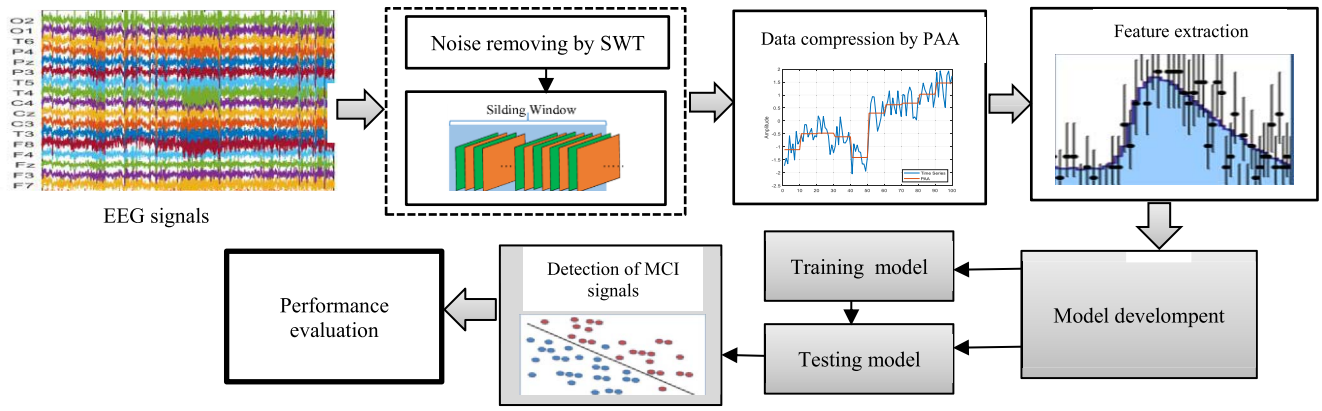

Fig. 1. Proposed framework for automatic detection of MCI patients from EEG signal data.

TABLE I
DEMOGRAPHIC CHARACTERISTICS OF MCI AND HC SUBJECTS

Characteristics	MCI	HC
Age (years) (mean± SD)	66.4±4.6	65.3±3.9
Education level (years) (mean± SD)	10.3±3.8	11.1±3.0
MMSE scores (mean± SD)	27.6±0.9	29.0±0.8
NUCOG score	82.4±3.6	91.1±3.0

*SD= standard deviation

namely, Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2. Neuropsychiatric interviews considering Peterson's criteria for MCI were completed for all subjects. MMSE scores from 21 to 26 were utilized for validation of MCI diagnosis and scores more than 26, were considered normal controls. The Neuropsychiatry unit cognitive assessment tool (NUCOG) has been used to confirm the diagnosis of MCI [8]. Table I reports demographic information of the MCI and HC subjects.

B. Removing Baseline Drift and Power Line Interference With Other Low & High Frequency Noises

EEG signals are contaminated by various types of unwanted signals (called noises) such as "baseline drift" and "power line interference" which may bias the analysis, and may lead to wrong conclusions. In this study, we employed stationary wavelet transformation (SWT) to eliminate baseline drift and other low-frequency noises by removing 0-0.5 Hz components of EEG signals, and also eliminate high-frequency noises including grid interference by removing 32 Hz–128 Hz components. Thus, this study retains the components with the frequency of 0.5-32 Hz. The reason for considering the SWT in this study is that this method is able to perfectly deconstruct complex and non-stationary EEG signals into original signals of finite bandwith, and then reconstructing them again with very little loss of information compared to conventional filters such as Infinite impulse response (IIR) or Finite impulse response (FIR) filters.

In this study, we applied the SWT decomposition and reconstruction for denoising EEG signals. We used the "swt" function for decomposing raw EEG signals with different frequency range and "iswt" function for reconstructing the cleaned coefficients as the denoised EEG signals in Matlab (using "Wavelet Toolbox"). We considered the mother wavelet basis function: "sym9" which is chosen from the "symlets" mother wavelet family. We set the decomposition level as 8 to acquire the coefficients with the appropriate frequency band as the sampling rate of the collected EEG signals is 256Hz.

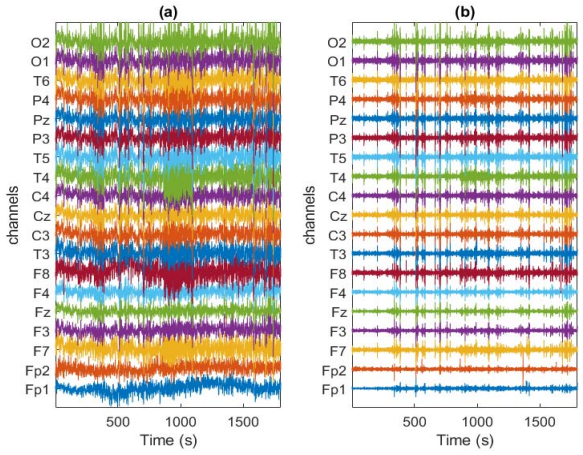

Fig. 2. An illustration of EEG signals obtained from a MCI patient (a) raw signals, (b) the denoised signals (from subject 1).

In order to show the transformation of the raw EEG signals to the denoised EEG siganls, Fig. 2 displays an example of 19-channel EEG signals of an MCI patient in the time domain for the raw EEG signals (Fig. 2(a)), and the denoised EEG signals (Fig. 2(b)).

C. Segmentation of the Denoised Signals

In this study, we segmented each channel's denoised data in each subject into some non-overlapping sliding windows based on a particular time period. The reason for segmentation is to properly account for possible stationarities having representative information from a specific time period as EEG signals are aperiodic and non-stationary and the magnitude of the signals change over time. In this work, we considered a 2 second sliding window length after testing a few values i.e., 0.5 second, 1 second, 2 second, 4 second, of window length

based on the performance of MCI patients' classification results. We set an overlapping rate of 0% to avoid bias result.

D. Data Compression Using Piecewise Aggregate Approximation

As EEG data is huge in size, the goal of this section is to reduce data size in each segment of a subject. For this purpose, this study introduces a new technique for data compression called "Piecewise Aggregate Approximation (PAA)" which is used for the first time in EEG time series data. The PAA is a lossy data compression method which is a powerful and efficient way of reducing the number of data points in a time series. The lossy compression technique is beneficial, especially for signal data and image data as this method significantly reduces the size of the data having a close match of the original data without significant loss of information. In this study, we apply the concept of the PAA technique to reduce the length of each channel EEG data within a segment, keeping the same shape of the original signals. In the EEG data compression, the PAA method eliminates some amounts of data that are not noticeable and thus diminishes the redundancy of the representation. The motivation for using the PAA algorithm in this study is that it reduces the dimensionality of huge EEG signal data preserving important information about the original data. The algorithm also keeps the same shape of the original data while filtering methods (e.g. low-pass filter) fail to recover the original signal with the same shape if the filter is not chosen properly.

Suppose we have a time series of N dimension in a segment, $X = [x_1, x_2, ..,x_N]$. Let us say that this time series contains N data points, and we want to reduce the time series to n data points, such that $n< N$. In order to condense the data from N dimensions to n dimensions, the data is divided into n equivalent sized 'frames'. The mean value of the data falling within a 'frame' is calculated and a vector of these values becomes the data reduced representation. In the PAA, a time series X of length N is represented in n space by a vector $\bar{X} = [\bar{x_1}, \bar{x_2}, \ldots, \bar{x_n}]$. The i^{th} element of $\bar{X}$ is computed by using Equation 1 [14]–[16].

$$\bar{x_i} = \frac{n}{N} \sum_{j=\frac{N}{n}(i-1)+1}^{\frac{N}{n}i} x_j \tag{1}$$

Here $\bar{x_i}$ symbolizes the mean value of the i$^{\text{th}}$ subsequence. In order to provide a clear idea about the PAA approach, we give an example of the PAA representations of a sequence in Fig.3. Fig. 3 illustrates a small example of how the PAA method divides an EEG time series ($N = 100$ samples) reduced to $n = 10$ samples. It is seen in Fig.3 that the number of data points has been reduced while still keeping the shape of the original curve for the time series.

E. Feature Extraction Using the Auto-Regressive Model and Permutation Entropy

Features represent task-relevant information transforming raw EEG signals into a small number of relevant values. This study aims to design a feature extraction structure for handling the chaotic and non-stationary nature of EEG data.

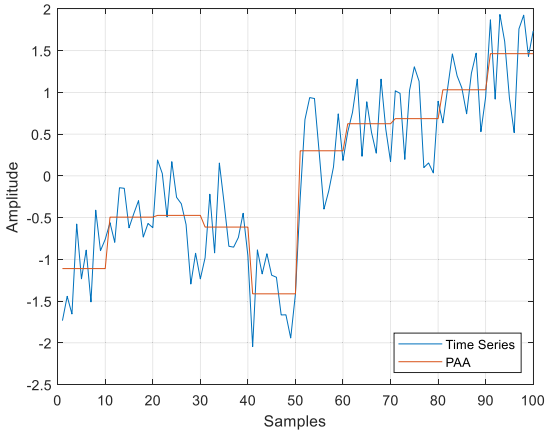

Fig. 3. An illustration of the PAA transformation.

Thus, this study considers the auto-regressive (AR) model and permutation entropy (PE) technique for feature extraction because the AR component can capture the time-dependent dynamics from the EEG signals and the PE can identify dynamic behavior of a time series which can give a quantitative complexity measure of dynamic EEG signals to distinguish MCI patients from healthy subjects.

1) Auto-Regressive (AR) Model: AR modelling is one of the prominent parametric methods for time series. AR models provide compact, computationally efficient representations of EEG signals. Furthermore, the parameters of AR models are invariant to scaling changes in the data that can arise from inter-subject variations. An AR model of order p for a single channel can be written as [17]–[19]:

$$x_n = \sum_{i=1}^{p} \alpha_i x_{n-i} + e_n \tag{2}$$

where the n^{th} value of x_n can be predicted by its previous p values: $x_{n-1}, x_{n-2}, \ldots, x_{n-p}$ and $\alpha_i (i = 1, 2, \ldots p)$ are model coefficients and e_n is the fitting error for x_n. The goal of the AR model is to estimate model coefficients that can fit compressed data as much as possible through the optimization process. In this study, we consider a 4$^{\text{th}}$ order AR model after empirical analysis which was executed through the Yule-Walker method in Matlab.

2) Permutation Entropy (PE): Permutation entropy (PE) is a quantitative complexity measure that explores the local order structure of chaotic and random behavior of a dynamic time series [17], [20], [21]. It transforms a given time series into a series of ordinal patterns, each describing the order relations between the present and a fixed number of equidistant past values at a given time [22]. Suppose, x_j ($j = 1, 2, \ldots, s$), is an EEG time series where s is the length of the time series. The time series can be formed into a symbolic sequence by using embedding theorems and delays [23], [24]. The process creates s-d+1 vectors $\{x(1), x(2), \ldots, x(s\text{-}d+1)\}$, where d denotes an embedded parameter which determines how much information is contained in each vector. Then it calculates a histogram of distribution of the ordinal patterns [17], [20], [21], [24]. For the time series, the number of any permutation pattern appearing is symbolized as $Q(\pi_i)$; thus relative frequency $p(\pi_i)$

is defined;

$$p(\pi_i) = \frac{Q(\pi_i)}{s - d + 1} \tag{3}$$

The PE is defined as for d ≥ 2;

$$H(d) = -\sum_{i=1}^{d!} p(\pi_i) \ \ln \ p(\pi_i) \tag{4}$$

This value is the measure of the amount of information contained in comparing d consecutive signal samples over some time interval. H(d) is generally normalized as;

$$H_n(d) = \frac{H(d)}{d - 1} \tag{5}$$

where the embedding parameter should be small enough to capture the deterministic dynamics of the signal. Additionally, if d is too small, PE may not work well. Bandt and Pompe proposed a value between 3 and 7 for the embedding parameter to provide the above situations [20], [17], [24].

In this study, we consider d value as 3 after several empirical evaluations. In Matlab, we use 'pec' function for the PE analysis and obtain two outcomes: 'permutation entropy' and the 'histogram' for the order distribution'; they are both used as features in this study. We consider the six bins of the histogram as six features. Finally we create a final feature vector set combining 'AR components', 'permutation entropy' and the 'histogram for the order distribution' features which is used to generate training sets and testing sets through the 10 fold cross-validation process for classification MCI patients from healthy control subjects discussed in the below section.

F. Classification

The aim of this section is to investigate which machine learning technique is suitable to generate enhanced results for the obtained feature vector set. In this study, we consider three popular classification methods: 'Extreme Learning Machine (ELM)'; 'Support Vector Machine (SVM)' and 'K-Nearest Neighbours (KNN)' and develop three models for these individual three methods with the same feature set, called the 'ELM based model', 'SVM based model' and 'KNN based model'.

1) ELM Based Model: ELM is a comparatively new machine learning method which can provide better generalization performance at a much faster learning speed and can obtain the global optimal solution [24], [25]. ELM was originally developed for Single-hidden layer feed forward networks (SLFNs) and then extended to the "generalized" SLFNs [26], [27]. In ELM theory, the input weights and hidden layer biases have been randomly initiated and all processes remain constant. The output weights are calculated by exploiting the Least Squares solution. Accordingly, the SLFN's training process is performed in this way. For a given training data $\{x_i, y_i\}$ with N samples, where x_i $(x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^{\mathrm{T}} \epsilon \mathrm{R}^{\mathrm{n}} \mathrm{i} = 1, 2, \ldots, \mathrm{N})$ represents the inputs and y_i $(y_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^{\mathrm{T}} \epsilon \mathrm{R}^{\mathrm{m}})$ represents the target data, the ELM model which has zero training error with M hidden neurons can be written as;

$$\sum_{i=1}^{M} \beta_i g\left(w_i . x_j + b_i\right) = y_j, \quad j = 1, ..., N \tag{6}$$

where g(x) is activation function, w_{ij} is input weights, b_i is hidden layer bias vector, β_{i} is output weights, o_j is ELM outputs and $(w_i . x_j)$ shows the inner product. The ELM model is assumed to have zero training error in Eq. (6). Eq. (6) can be expressed in matrix form;

$$Y = H\beta \tag{7}$$

where H is hidden layer output matrix and H, Y and β are as follows;

$$H = \begin{bmatrix} g(w_1 . x_1 + b_1) & \cdots & g(w_M . x_1 + b_M) \\ \vdots & \cdots & \vdots \\ g(w_1 . x_N + b_1) & \cdots & g(w_M . x_N + b_M) \end{bmatrix}_{NxM},$$
$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{Mxm} \text{ and } Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{Nxm} \tag{8}$$

The least-squares solution of Eq. (7) is;

$$\beta = H^{\dagger} Y \tag{9}$$

where $H^{\dagger}$ represents the Moore–Penrose generalized H. For ELM, the number of layers is an important parameter. In this study, after several experimental trials we selected the ELM hidden layer neuron number as 50 in order to provide a better performance. We used the activation function '*hardlim*' in Matlab.

2) SVM Based Model: SVM is the most popular machine learning tool that can classify data separated by non-linear and linear boundaries, originated from Vapnik's statistical learning theory [28]. The main concept in the SVM algorithm is to first transform the input data into a higher dimensional space and then construct an optimal separating hyper-plane between the two classes in the transformed space [28]–[30]. In order to solve nonlinear problems, SVM utilizes kernel function [29], [31], which allows a better fit of the hyper plane to more general data sets. The detailed description of SVM is available in references [28]–[30], [31]. In this study, the linear kernel is selected as an optimal kernel function for the SVM model after testing all the kernels (e.g. linear, polynomial, radial basis kernel) as this kernel produced better performance compared to others.

3) KNN Based Model: KNN is a another familiar supervised learning approach, which classifies a given data point according to the majority of its neighbors [31], [32]. It chooses nearest k samples from the training set, then takes the majority vote of their class where k should be an odd number to avoid ambiguity [33], [34]. To find the neighbor, it makes use of distance metrics like 'euclidean distance'. This study used the default distance metric 'euclidean distance' for distance measure and K = 1 in the KNN model after several experimental evaluations. In the experiments, we considered a range of K values in between 1 and 30, and picked an appropriate K value that generated the lowest error rate as the lowest error rate refers to the best model. In the experimental results, we obtained the lowest error rate for K = 1. We used "*fitcknn*" function in Matlab for the KNN model. A detailed discussion of this method is available in references [31]–[34].

G. Criteria of Performance Evaluation

In this study, we assess the stability of the performance of the proposed models on the tesing data set through most of the criteria that are usually used in biomedical research as depicted below [10], [35]–[39]. These criteria allow to estimate the behaviour of the classifiers on the extracted feature data.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$\boldsymbol{Sensitivity\,(Recall)} = \frac{TP}{TP + FN} \times 100$$

$$\boldsymbol{Specificity} = \frac{TN}{TN + FP} \times 100$$

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \times 100$$

$$\boldsymbol{FalseAlramrate} = \frac{FP}{FP + TN} \times 100$$

$$\boldsymbol{F1score} = 2 * \frac{precision * recall}{precision + recall} \times 100$$

where

True Positive (TP): number of MCI samples classified correctly by the proposed model as MCI;

False Positive (FP): number of healthy control (HC) sample classified incorrectly by the proposed model as MCI;

True Negative (TN): number of HC samples classified correctly

by the proposed model as HC;

False Negative (FN): number of MCI samples classified incorrectly by the proposed model as HC.

1) Receiver Operating Characteristic (ROC) Curve Area: area under the ROC curve provides a measure of overall performance of the classifier. The ROC curve displays the plots of True Positive Rate (sensitivity) versus false positive rates [29].

2) Kapa Coefficient: Kappa is a chance-corrected measure of agreement between the classifications and the true classes [29], [39]. It is calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement.

III. Experimental Set Up

The experiments were carried out in a MATLAB (2018b) environment in a computer with an Intel Core i7-4810 CPU and 32 GB memory. In the experimental setting, firstly we selected the parameters of the mentioned classifiers: ELM, SVM and KNN because the classification performances of a classifier depend on the values of the parameters as discussed in Section II (F). As mentioned before, the data set that is used in this study contains a total of 27 subjects, including 16 cognitively healthy subjects and 11 MCI patients. The size of the original data was 19×460.800 for one MCI patient and it was the same size for one healthy control (HC) subject. After the compression of data size by the PAA technique, the size of original data reduced to $19 \times 256 \times 899$ for one MCI patient and it was the same size for one HC subject. Finally we obtained a feature vector set which was 24, 273×209 together for all 27 MCI and HC subject. After that we developed a training data set to train the proposed framework and testing dataset to test the performance of the proposed framework.

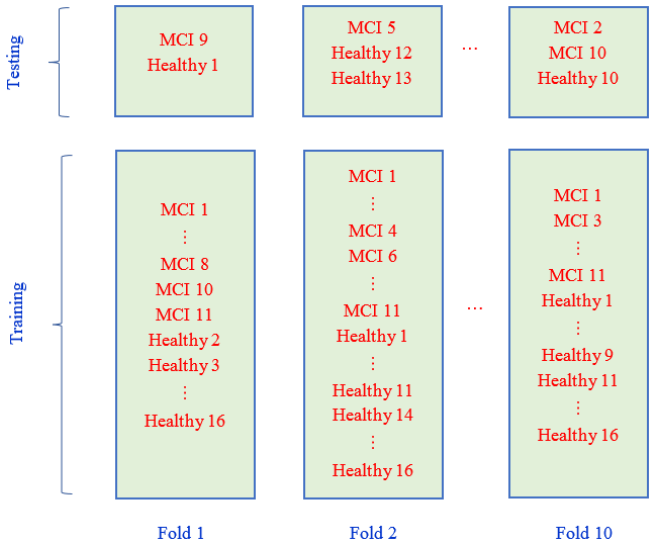

Fig. 4. An example of subject-wise data partition in the 10-fold cross validation.

To reduce bias, we proposed to employ a k-fold cross-validation technique [37], [38], [40] considering k = 10 for getting the training set and the testing set. This 10-fold cross-validation design was prepared based on subject-wise data. In this design, all 27 subjects' data were randomly divided into 10 folds containing approximately equal size (without repetition) as shown in Fig. 4. One of the 10-folds was used as a testing set and the other 9 folds were used to form a training set for the proposed model. This procedure was repeated 10 times, and then the average performance (e.g. accuracy, sensitivity, specificity) across all 10 repetitions of the testing set was computed for consideration. In this procedure, a subject's data are either in the train or test sets, but not both, in each of the folds. Fig.4 shows an example of the subject-wise data partition in the experimental set up. As 27 (no. of subject) is not equally dividable by 10, thus the folds did not contain equal number of subjects' data. As an example, shown in Fig. 4, in Fold 1, the data of MCI 9 and Healthy 1 subjects were used as the testing set, the rest of the subjects' data were considered as the training set, while in Fold 2, the data of MCI 5, Healthy 12 and Healthy 13 were used as the testing set, the rest of the subjects' data were considered as the training set. Thus, the training data size in Fold 1, is 22475×209 and testing data size is 1798×209. In Fold 2, the training data size is 21576×209 and testing data size is 2697×209.

IV. Results and Discussion

This section firstly explores the representative features that are considered in this study. Secondly this section reports the resultant classification performance of the proposed models on the testing dataset. Finally, this section provides a comparison between the proposed method and some existing methods.

A. Exploration of Representative Features

An exemplary EEG signal pattern for the MCI patient and Healthy control (HC) subject is illustrated in Fig. 5. It is

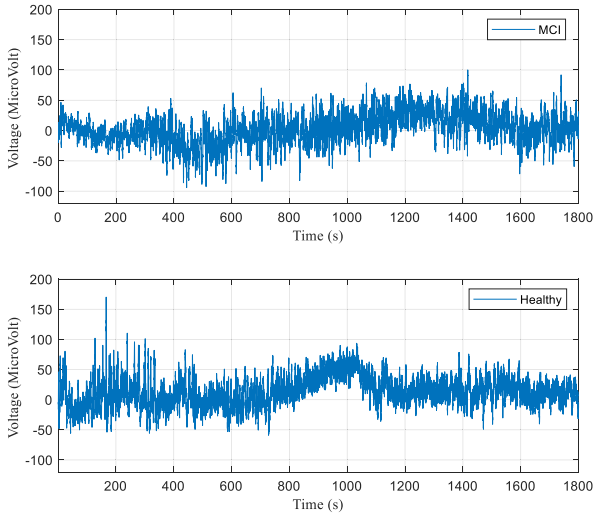

Fig. 5. Exemplary EEG Signal pattern for MCI and healthy control subject (Subject 1: FP1 channel).

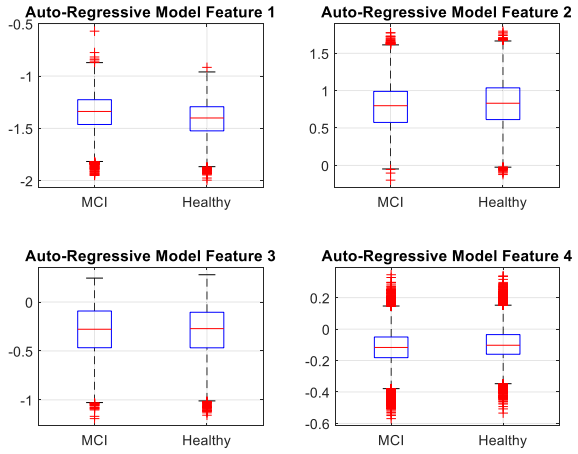

Fig. 6. Boxplots for the AR model features on 19 channels grouped by MCI and healthy subject.

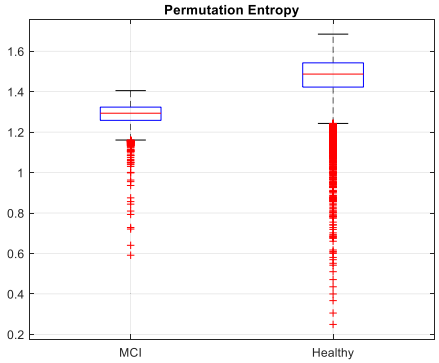

Fig. 7. Boxplots for the PE feature on 19 channels grouped by MCI and healthy subjects.

observed from Fig. 5 that the voltage amplitude of the EEG signal is lower for the MCI patient compared to the HC subject.

In order to provide the concealed information of the obtained features, we drew three boxplot diagrams: Fig. 6, Fig. 7 and Fig. 8 based on three feature values including AR coefficients, PE and Histogram features, respectively. These three figures (Fig.6-Fig.8) present the distribution of the features for two categories of MCI and healthy control subjects. The box plot is a standardized way of displaying the distribution of data based on the five-number summary: minimum, first quartile (Q1), median, third quartile (Q3), and maximum. The body of the boxplot consists of a "box". The top of the box represents the Q3 and the bottom of the box represents the Q1, and the line in the middle of the box represents the median (also called centre). The "whiskers" are the two lines outside the box that extend to the highest and lowest observations. The whiskers go from each quartile to the minimum or maximum. In the box plot, an outlier is defined as a data point that is located outside the whiskers of the box plot represented by the circles or stars beyond the whiskers. Outliers are either $3\times IQR$ or more above the Q3 or $3\times IQR$ or more below the Q1 and Suspected outliers are either $1.5\times IQR$ or more above the Q3 or $1.5\times IQR$ or more below the Q1.

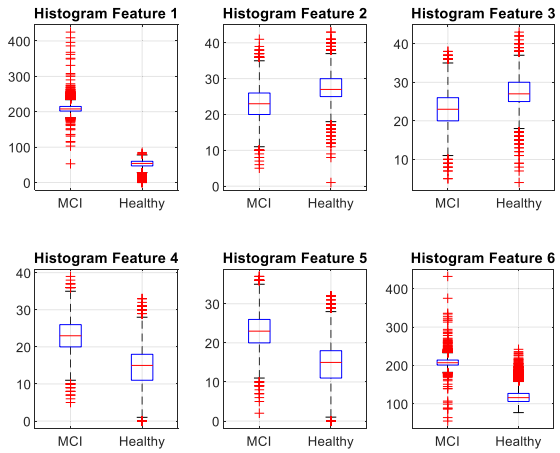

Fig. 8. Boxplots for the histograms features on 19 channels grouped by MCI and healthy subjects.

The distributions of four AR models (4^{th} order AR model) in the MCI group and the healthy group are shown in Fig.6. It is apparent from this figure that in all of the four AR models, the shape of the distributions in both MCI and healthy group are symmetric and there are some outliers in each diagram. The centre value of MCI group is slightly higher than in the healthy group and most of the cases' lower spread is seen in the healthy group. It is noticeable that there is a difference in each of the distributions of the AR models between the MCI and Healthy groups.

Fig.7 displays the distribution of PE values in MCI and healthy groups. It is seen that the shape of the distribution in both groups are symmetric but there is a significant difference in the central value of both groups. It is also observed from Fig.8 that there is a significant difference in each of the histogram distributions into two categories: MCI and healthy groups. The boxplots in Fig.6-Fig.8 clearly demonstrate that there is a clear significant difference in the values of the feature set in two groups that help in the efficient classification.

TABLE II
PERFORMANCE OF PROPOSED MODELS FOR MCI VS. HEALTHY CONTROL SUBJECTS

Models	Accuracy (%)	Sensitivity (recall) (%)	Specificity (%)	Precision (%)	F1 score (%)	False alarm rate	Kappa Coefficient
ELM based model	98.78	98.32	99.66	99.69	98.95	0.003	0.974
SVM based model	97.41	95.11	100.00	100.00	97.42	0.000	0.946
KNN based model	98.19	100.00	95.72	96.76	98.34	0.043	0.961

TABLE III
PERFORMANCE OF THE PROPOSED ELM BASED MODEL IN THE FOLDS OF 10-FOLD CROSS VALIDATION PROCEDURE

Folds	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	False alarm rate	Kappa Coefficient
Fold 1	90.36	85.71	99.67	99.81	92.22	0.003	0.797
Fold2	99.96	100.00	99.89	99.94	99.97	0.001	0.999
Fold3	99.33	98.89	99.56	99.11	99.00	0.004	0.985
Fold4	99.61	99.22	100.00	100.00	99.61	0.000	0.992
Fold5	99.85	99.83	99.89	99.94	99.89	0.001	0.997
Fold6	99.44	100.00	98.89	98.90	99.45	0.011	0.989
Fold7	100.00	100.00	100.00	100.00	100.00	0.000	1.000
Fold8	99.83	100.00	99.67	99.67	99.83	0.003	0.997
Fold9	99.74	99.61	100.00	100.00	99.80	0.000	0.994
Fold10	99.63	99.94	99.00	99.50	99.72	0.010	0.992
Average	98.78	98.32	99.66	99.69	98.95	0.003	0.974

B. Experimental Results

Table II provides overall classification performances for the proposed models using 10-fold cross validation in terms of accuracy, sensitivity (recall), specificity, precision, false alarm rate, F1 score (F-measure) and kappa coefficient. As mentioned before, the obtained feature set is modelled on three modern machine learning techniques: ELM, SVM and KNN. As shown in Table II, the ELM based model yields overall 98.78% of classification accuracy (with sensitivity 98.32% and specificity 99.66%) while this value is 97.41% (with sensitivity 95.11% and specificity 100%) for the SVM based model and 98.19% for the KNN based model (with sensitivity 100% and specificity 95.72%). The results show a 1.37% improvement in overall accuracy for the ELM based model compared to the SVM based model and 0.68% over the KNN based model for the same inputs. We can also see from Table II that the highest kappa value (0.974∼97%) and F1 score (98.95%) is obtained by the ELM based model. The SVM based model produces highest precision value (100%) and the lowest false alarm rate (0.00∼0%) compared to the other two models. The results demonstrate that the ELM based model is superior compared to the SVM and the KNN based models for the same features.

As the ELM based model has proved to be the best model for our proposed framework, we have provided more detailed results for the ELM model to have clear information about the precision, false alarm rate, F1 score (F-measure) and kappa through each of the 10 repetitions in 10-fold cross validation procedures in terms of accuracy, sensitivity (recall), specificity, and coefficient reported in Table III. It can be seen from this table that there are no significant variations in the performances in the 10 repetitions. In Table III, in all of the 10 fold cases, the classification accuracies are more than 99% except fold-1, which confirms the reliability of the proposed ELM based method in distinguishing MCI patients from healthy subjects.

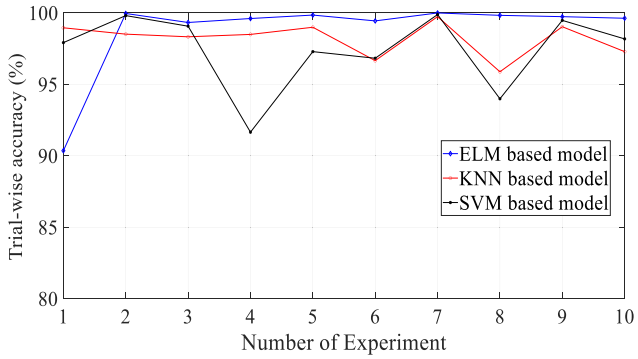

Fig. 9. Trail-wise classification accuracy scores for the three proposed models in 10 experiements.

Fig. 9 shows the fluctuations of the classification accuracy for each of the three models varying across different experimental trials with the same input feature set. It can be seen from the figure that overall test accuracy is relatively higher in the all of the 10 experiments for the ELM based model compared to the SVM based model and KNN based model except for the experiment 1. It proves that discriminative features extracted from the brain signals are effective for the ELM classifier.

In addition, we utilized ROC curves to further assess the MCI detection performance for the three models with the same feature set. The Area Under the Curve (AUC) is the value of area under ROC curve, that belongs between 0 and 1. The performance of MCI detection is better when the AUC value is closer to 1 and the standard deviation value is lower [5]. Fig.10 illustrates the ROC curves of three MCI detection models with the same input feature set. As can be seen

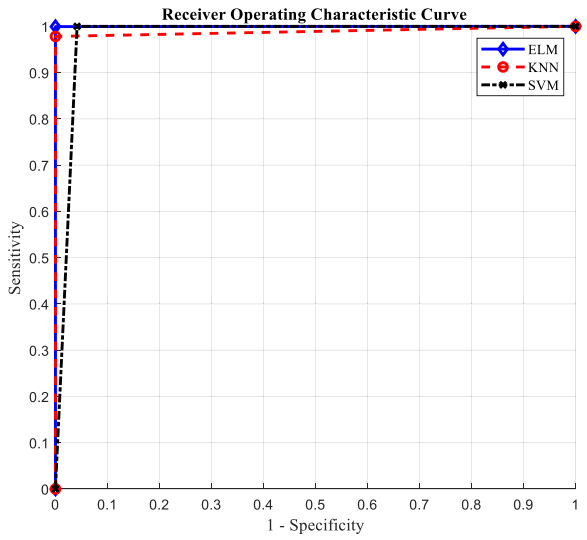

Fig. 10. The ROC curves of three classification models with same input data.

TABLE IV
AREA UNDER CURVE (AUC) VALUES OF THE THREE MCI DETECTION MODELS WITH THEIR STANDARD DEVIATION

Models	Area Under Curve (AUC) values	Stanadrd deviation
ELM based model	0.97981	0.04441
SVM based model	0.95106	0.05431
KNN based model	0.95723	0.02540

TABLE V
TIME COMPARISON FOR THE THREE CLASSIFIERS

Classifier	Training Time (in second)	Testing Time (in second)
ELM	0.915	0.281
SVM	23.031	0.188
KNN	0.141	159.453

from Table IV, the AUC of the ELM based model is higher (0.97981) compared to the SVM based model and the KNN based model with the lower standard deviation (0.04441). This result proves that the ELM based model has better capability to generate a higher performance compared to the SVM based model and the KNN based model.

To evaluate the computational efficiency of the proposed models, the execution (running) time of the ELM classifier is compared with the other two classifiers: SVM & KNN in the training part and testing part. Table V shows that the ELM classifier took only 0.281 seconds (2^{nd} the lowest time) in the testing part and 0.915 seconds in the training part (2^{nd} the lowest time). The KNN classifier took the highest time (159.453 seconds) for the testing part but took the lowest time, 0.141 seconds for the training section. On the other hand, the SVM took the lowest time, 0.188 seconds for the testing part and took 23.031 seconds in the training part.

In order to further evaluate the performance of our proposed models, we also checked the methods subject-wise applying the Leave-one-out cross-validation (LOOCV) method. As we have a total of 27 subjects, one of the 27 subjects was used as a testing set and the rest of the 26 subjects were used as a training set and the same procedure was repeated 27 times to ensure that each subject was used as test sample once. The proposed algorithm was applied once for each case, using all other cases as a training set and using the selected case as a single-item test set.

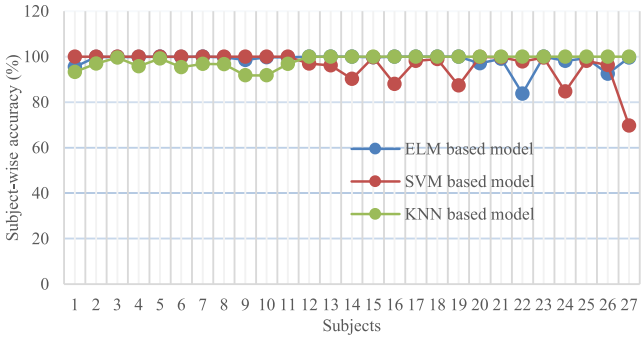

Fig. 11. Subject-wise classification accuracy scores for the three proposed models.

Fig.11 presents a subject-wise accuracy score for the proposed three models. This figure shows the experimental results of the LOOCV method. For each of 27 subjects, the obtained accuracy scores (in percentage) for the testing set were provided in Fig.11. It is apparent from this figure that in most of the cases (except subject no. 22), the proposed ELM based model performed better on subject-wise experiments compared to the other two models. These results show similar consistency to previous results (trial-basis) where the ELM based model was considered as the best model for identifying MCI signals from HC signals. We can conclude that using trial-basis and subject-wise experiments in both ways, the ELM based model yielded better results with the obtained feature set compared to the other two models."

The rationale for the better performance of the ELM is that the ELM is fully automatically implemented without iterative tuning, and no intervention is required from users while other machine learning methods such as SVM and KNN face difficulties in manually tuning control parameters. The ELM does not need to tune parameters for its hidden layers, resulting in a faster training speed and an improved generalization performance while the SVM and KNN methods need to tune the hyperplane parameters which usually are time consuming and often give overfitting results.

C. Comparison With Other Methods

The objective of this section is to compare our proposed model with the existing methods which were used in the same dataset: the MCI EEG data set [8] that was used in this study. Until now, we have found only two research papers [8], [11] which used the same MCI EEG data set that we used in this study. Table VI shows the performance comparison of the proposed method with these two published methods [8], [11]. In [8], Kashefpoor *et al.* achieved 88.89% of classification accuracy. The sensitivity and specificity scores of their method were 100% and 83.33%, respectively. Yin *et al.*

TABLE VI
COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS

Authors	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
Kashefpoor et al. [8]	Spectral features with NF-KNN	88.89	100	83.33
Yin et al. [11]	Spectral-temporal based features with SVM	96.94	96.89	96.99
Proposed Method	AR and PE based features with ELM	98.78	98.32	99.66

[11] obtainedoverall classification accuracy of 96.94% and sensitivity and specificity scores were 96.89% and 96.99%, respectively. As seen in Table VI, our proposed ELM based model produced higher performance scores (accuracy 98.78%; sensitivity 98.32% and specificity 99.66%) compared to the existing two methods. The produced accuracy score of our proposed ELM based model is 9.89% better than Kashefpoor *et al.* [8] accuracy score and 1.84% better than Yin *et al.* [11] accuracy score.

The results show that our proposed ELM based method is significantly better than the other methods in the literature. The improved classification performance may be yielded due to the effectiveness of the features considered to be appropriate. Our proposed combined features (AR and PE) have high ability to represent and model the characteristics and information inside the signals for extracting brain information relevant to MCI, offering better frequency resolution and smooth power spectra.

D. Limitations of the Study

This study has some limitations which are planned as the focus for our future research. The main limitation is small data size. We used an MCI based EEG dataset, which is small in size (total 27 subjects, including 16 HC subjects and 11 MCI patients). However, a large dataset can validate the robustness of the proposed method for classification work. In the near future, we will extend the application of this method to clinical big datasets. Another limitation is: this study focuses on a two-class classification problem (MCI vs HC), although we plan to extend the proposed algorithm to multiclass situations in the near future such as the classification of MCI, HC, and AD subjects.

V. CONCLUSION

This study aimed to develop an efficient framework in extracting discriminating features from EEG signals by means of the AR model and PE method for the automatic detection of MCI patients from control subjects. The performance of our proposed framwork was evaluated in two ways: trial-wise and subject-wise. In both ways, the research findings indicate that the considered features have the capability to extract important information for identifying MCI patients from healthy control subjects. The results showed that our proposed ELM based model with the extracted features could significantly enhance the classification accuracy (overall 98.78%) compared to the other two models with a lower executive time (0.281 seconds). Finally, we compared our proposed model with other state-of-the-art methods in the literature, which used same dataset that we used in this study. The results demonstrated that our model significantly outperforms the other existing methods in the literature. In the future, this study can be expanded to real-time application developing a software/demo/App that will help the experts to automatically and effieciently identy MCI from EEG signal data.

REFERENCES

[1] J. C. McBride *et al.*, "Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease," *Comput. Methods Programs Biomed.*, vol. 114, no. 2, pp. 153–163, Apr. 2014.

[2] K. Weir, "Spotting the signs of mild cognitive impairment," *Amer. Psychol. Assoc.*, vol. 50, no. 8, p. 40, Oct. 2019.

[3] Alzheimer's Association Report, "Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 11, no. 3, pp. 332–384, Mar. 2015.

[4] Alzheimer's Disease International, London, U.K. (Sep. 2019). *World Alzheimer Report 2019 Attitudes to Dementia*. [Online]. Available: https://www.alz.co.uk/research/WorldAlzheimerReport2019.pdf

[5] R. Ju, C. Hu, P. Zhou, and Q. Li, "Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 1, pp. 244–257, Jan./Feb. 2019.

[6] P. Durongbhan *et al.*, "A dementia classification framework using frequency and time-frequency features based on EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 826–835, May 2019.

[7] S. Khatun, B. I. Morshed, and G. M. Bidelman, "A single-channel EEG-based approach to detect mild cognitive impairment via speech-evoked brain responses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1063–1070, May 2019.

[8] M. Kashefpoor, H. Rabbani, and M. Barekatain, "Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features," *J. Med. Signals Sens.*, vol. 6, no. 1, pp. 25–32, Jan./Mar. 2016.

[9] R. Zarei, J. He, S. Siuly, G. Huang, and Y. Zhang, "Exploring Douglas–Peucker algorithm in the detection of epileptic seizure from multicategory EEG signals," *BioMed Res. Int.*, vol. 2019, pp. 1–19, Jul. 2019.

[10] S. Siuly, V. Bajaj, A. Sengur, and Y. Zhang, "An advanced analysis system for identifying alcoholic brain state through EEG signals," *Int. J. Autom. Comput.*, vol. 16, no. 6, pp. 737–747, Dec. 2019, doi: 10.1007/s11633-019-1178-7.

[11] J. Yin, J. Cao, S. Siuly, and H. Wang, "An integrated MCI detection framework based on spectral-temporal analysis," *Int. J. Autom. Comput.*, vol. 16, no. 6, pp. 786–799, Dec. 2019.

[12] N. Sharma, M. H. Kolekar, K. Jha, and Y. Kumar, "EEG and cognitive biomarkers based mild cognitive impairment diagnosis," *IRBM*, vol. 40, no. 2, pp. 113–121, 2019.

[13] *EEG Signals from Normal and MCI (Mild Cognitive Impairment) Cases*. [Online]. Available: http://ww25.biosigdata.com/?download=eeg-signals-from-normal-and-mci-cases#comment-2909

[14] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowl. Inf. Syst.*, vol. 3, no. 3, pp. 263–286, 2001.

[15] H. Ren, X. Liao, Z. Li, and A. AI-Ahmari, "Anomaly detection using piecewise aggregate approximation in the amplitude domain," *Int. J. Speech Technol.*, vol. 48, no. 5, pp. 1097–1110, May 2018.

[16] B. K. Yi and C. Faloutsos, "Fast time sequence indexing for arbitrary Lp norms," in *Proc. 26th Int. Conf. Very Large Data Bases*, 2000, pp. 385–394.

[17] D. Li, Z. Liang, Y. Wang, S. Hagihira, J. W. Sleigh, and X. Li, "Parameter selection in permutation entropy for an electroencephalographic measure of isoflurane anesthetic drug effect," *J. Clin. Monitor. Comput.*, vol. 27, no. 2, pp. 113–123, Apr. 2013.

[18] V. Lawhern, W. D. Hairston, K. McDowell, M. Westerfield, and K. Robbins, "Detection and classification of subject-generated artifacts in EEG signals using autoregressive models," *J. Neurosci. Methods*, vol. 208, no. 2, pp. 181–189, Jul. 2012.

[19] M. Wang, S. Abdelfattah, N. Moustafa, and J. Hu, "Deep Gaussian mixture-hidden Markov model for classification of EEG signals," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 4, pp. 278–287, Aug. 2018.

[20] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.

[21] L. Xiao-Feng and W. Yue, "Fine-grained permutation entropy as a measure of natural complexity for time series," *Chin. Phys. B*, vol. 18, no. 7, p. 2690, 2009.

[22] G. Ouyang, J. Li, X. Liu, and X. Li, "Dynamic characteristics of absence EEG recordings with multiscale permutation entropy analysis," *Epilepsy Res.*, vol. 104, no. 3, pp. 246–252, May 2013.

[23] P. Bhuvaneswari and J. S. Kumar, "Total variation based multi feature model for epilepsy detection using support vector machine," *IETE J. Res.*, vol. 62, no. 6, pp. 822–832, 2016.

[24] S. Siuly, O. F. Alcin, V. Bajaj, A. Sengur, and Y. Zhang, "Exploring Hermite transformation in brain signal analysis for the detection of epileptic seizure," *IET Sci., Meas. Technol.*, vol. 13, no. 1, pp. 35–41, 2018.

[25] Ö. F. Alçin, S. Siuly, V. Bajaj, Y. Guo, A. Şengur, and Y. Zhang, "Multi-category EEG signal classification developing time-frequency texture features based Fisher Vector encoding method," *Neurocomputing*, vol. 218, pp. 251–258, Dec. 2016.

[26] L. Duan, M. Bao, J. Miao, Y. Xu, and J. Chen, "Classification based on multilayer extreme learning machine for motor imagery task from EEG signals," *Procedia Comput. Sci.*, vol. 88, pp. 176–184, 2016.

[27] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[28] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 2000.

[29] S. Siuly, E. Kabir, H. Wang, and Y. Zhang, "Exploring sampling in the detection of multicategory EEG signals," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–12, Apr. 2015, Art. no. 576437.

[30] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, "Classification of motor imagery EEG signals with support vector machines and particle swarm optimization," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–8, May 2016, Art. no. 4941235.

[31] S. Supriya, S. Siuly, H. Wang, J. Cao, and Y. Zhang, "Weighted visibility graph with complex network features in the detection of epilepsy," *IEEE Access*, vol. 4, pp. 6554–6566, 2016.

[32] A. Bablani, D. R. Edla, and S. Dodia, "Classification of EEG data using k-nearest neighbor approach for concealed information test," in *Proc. 8th Int. Conf. Adv. Comput. Commun. (ICACC)*, vol. 143, 2018, pp. 242–249.

[33] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[34] H. R. Al Ghayab, Y. Li, S. Siuly, and S. Abdulla, "Epileptic EEG signal classification using optimum allocation based power spectral density estimation," *IET Signal Process.*, vol. 12, no. 6, pp. 738–747, Aug. 2018.

[35] S. Siuly and Y. Li, "A novel statistical algorithm for multiclass EEG signal classification," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 154–167, Sep. 2014.

[36] S. Faul and W. Marnane, "Dynamic, location-based channel selection for power consumption reduction in EEG analysis," *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 1206–1215, 2012.

[37] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "EEG sleep stages analysis and classification based on weighed complex network features," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Nov. 5, 2018, doi: 10.1109/TETCI.2018.2876529.

[38] S. Siuly and Y. Li, "Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification," *Comput. Methods Programs Biomed.*, vol. 119, no. 1, pp. 29–42, Apr. 2015.

[39] Siuly, X. Yin, S. Hadjiloucas, and Y. Zhang, "Classification of THz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers," *Comput. Methods Programs Biomed.*, vol. 127, pp. 64–82, Apr. 2016.

[40] S. Siuly and Y. Li, "Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, Jul. 2012.