Biomedical Signal Processing and Control 70 (2021) 103049

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Diagnosis of mild Alzheimer's disease by EEG and ERP signals using linear and nonlinear classifiers

Elias Mazrooei Rad, Mahdi Azarnoosh *, Majid Ghoshuni, Mohammad Mahdi Khalilzadeh

Department of Biomedical Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

ARTICLE INFO

Keywords:
Alzheimer's disease
EEG brain signal
ERP
Variance analysis
LDA
Elman neural network
Convolutional neural network

ABSTRACT

A new method for the diagnosis of Alzheimer's disease in the mild stage is presented due to its association with the characteristics of the EEG brain signal. In the proposed method, three channels Pz, Cz, and Fz of brain signals were recorded from 40 participants in the four modes of closed eyes, open eyes, recall, and stimulation. After preprocessing, wide various features of EEG signals were extracted. Results have shown that after stimulation, the amplitude and the latency of the P300 component change with increasing age and the stages of Alzheimer's disease. In addition, the first changes in brain signal in mild Alzheimer's patients are increased activity in the theta band and decreased activity in the beta band, which is associated with a decrease in alpha-band activity. After selecting the proper features, linear discriminant analysis, and Elman and convolutional neural networks were used to classify the participants' features. The results showed that the Pz channel among three EEG signals and the stimulation mode among the four recording steps had greater accuracy than the others. By using features of the Pz channel, the accuracy of LDA was 59.4% and 66.4% in the recall and excitation modes, accuracy of Elman neural network was 92.3% and 94.1%, and CNN was 97.5% and 99% respectively. Results have shown that extracting appropriate linear and nonlinear features has increased the accuracy of the classifiers. In addition, due to the dynamic nature of the brain signal, the convolutional neural network had better performance than LDA and Elman.

1. Introduction

Alzheimer's disease is a progressive disease of the intellect that is commonly seen in the elderly. Significant symptoms include loss of memory, judgment, and reasoning, and in addition significant behavioral changes in the individual [1]. The disease is caused by the loss of synapses of neurons in some areas of the brain, the necrosis of brain cells in different areas of the nervous system, the creation of spherical-like protein structures called aging plaques outside the neurons of certain brain regions, and filamentous protein structures called clots. Alzheimer's disease (AD) is the most common form of dementia and the primitive stage of AD is known as Mild Cognitive Impairment (MCI) [2]. The prevalence of Alzheimer's disease is increasing rapidly [3]. The number of Alzheimer's patients in Iran has almost doubled over the past 13 years, according to the Alzheimer's Association of Iran. On the other hand, the medical costs, as well as the care and nursing of these patients, are very high and difficult. It causes various mental disorders in the patient. It usually takes several years from the first symptoms to the acute stages of the disease, where most of the brain cells are destroyed. Unless timely diagnosed, new and up-to-date therapies will not work. The solution is to find the mechanism of this disease and its effect on brain signals, which is difficult to do due to the dynamic nature of the brain signals, and complex nature of the disease [4,5].

Early detection of AD by the EEG signal has generated widespread interest in the field due to the nonlinear changes and the dynamic nature of the signal [6,7]. Due to the dynamics and range of available EEG data, how to learn abstract representations of EEG for better generalization is still difficult [8,9]. In addition, most studies suggest that the machine learning model may not be able to adapt to the extracted features. In fact, EEG changes in high-dimensional space are nonlinear and complex [10,11]. In this study, the performance of three classifiers is compared. Due to the nonlinear and dynamic nature of the EEG signal in return networks, especially CNN, better results can be observed and the frequency and spectral properties of the EEG signal in opinions have been defined [12-14]. Characteristics of frequency bands are used as tools for data processing [15]. Comparison of brain signal abnormalities in the group of patients with mild Alzheimer's disease with healthy individuals and dementia with Lewy bodies (DLBMCI) shows that the alpha

* Corresponding author.
E-mail address: M_Azarnoosh@mshdiau.ac.ir (M. Azarnoosh).

https://doi.org/10.1016/j.bspc.2021.103049
Received 27 March 2021; Received in revised form 16 July 2021; Accepted 8 August 2021
Available online 18 August 2021
1746-8094/

frequency band in this group is reduced compared to healthy elderly, while the results are different in the delta frequency band. The amplitude of this frequency band has increased for mildly ill people compared to healthy people [16]. However, CNN's ability to learn signal and image has made it possible to use the power spectrum as input for classification. On the other hand, with the help of the Fourier feature [17,18], a 2D or 3D network can be designed [19,20]. Amplitude and signal strength are considered as network input [21]. One of the advantages of neural networks is that they automatically learn complex features from large amounts of data, and this is the idea of the end-to-end learning method [22]. One method of the learning algorithms is dictionary learning, which is examined in a large data set, and a separate learning dictionary is defined for each class. However, error determination is required for a large number of classes in this method [23]. Although deep learning methods are new statistical algorithms, they are still not useful for single-channel data and cannot be compared and evaluated. On the other hand, if only the data of one subject is used, it is necessary to use the same data for long-term classification [24]. Dictionary pairing is a way to learn a hybrid dictionary in which the data is not separated by the classifier separately and the data is separated faster. The strategy is to use linear multiplication instead of nonlinear scattering [25].

ERPs are produced in conjunction with an event and can be extracted from EEG signals by the synchronous averaging method. Major ERPs are the responses that the brain produces when experiencing sensory stimulation. Usually, ERPs break down into sub-components according to different criteria. The endogenous components of ERP, including P100, N100, P200, N200, P300, and N400, it has been shown to have a direct relationship with some cognitive activity, but among these components, P300 is the most valid and potent locus for activity. The amplitude and latency indices of this ERP component depend on the amplitude of the excitation signal, the frequency of the excitation signal, the excitation intervals, and especially the probability of occurrence of two excitation categories. It is shown that if the frequency of target stimuli decreases or the intervals between stimuli increase, the amplitude of this component increases in healthy individuals. In addition the amplitude and delay of this component change with age and the stages of Alzheimer's disease and even forms of dementia [26].

As a result, the aim of this study is to identify the best and most effective indicator for detecting the disease and realized how it is related to brain signal characteristics in the mild state. Although there are several imaging modalities now available to detect this disease, but the high cost of these imaging modalities is one of the disadvantages of imaging modalities. Therefore, it is necessary to find a low-cost method with proper accuracy. Therefore, in addition to extracting proper linear and nonlinear features of EEG signal especially in alpha frequency subbands another aim of this study is to identify the ability of combining extracted features of both ERP and EEG signals in increasing the accuracy of mild Alzheimer's disease detection by using a proper classifier.

2. Materials and methods

In this study, brain signals were recorded from 40 healthy, mild, and severely ill participants. First, they should be evaluated by a physician using criteria such as clinical tests.

2.1. Clinical diagnosis of candidates

One of the criteria for classifying the participants is the MMSE memory test. There are other tests, such as the DRS, which aims to test Alzheimer's and Dementia differentiation. However, the long duration of this test considering the proper performance of the MMSE test is not helpful in measuring memory status. The label assigned according to the score earned by the subject in the test and given the threshold. The following groups are taken: 1- Healthy 2- Mild Patient 3- Serious Patient. When performing the test, it is important to note that according to the questions raised, the difference between literate and illiterate taken into account. The range for both groups is shown in Table 1.

2.2. Brain signal recording

The Powerlab SP device with two amplifiers was used to record the signal. The four channels of this device were used to record three channels of brain signal and one EOG signal. In addition, an external exciting audio signal used to solve the problem of synchronization of the excitation signal and the ERP [27]. The recording was in 4 channels and according to the standard 10–20 with a sampling rate of 1000 Hz and 16 bits for each sample. Three unipolar channels (Fz, Pz, and Cz) were used to record brain activity. A bipolar signal was recorded to evaluate EOG activity and its effect on brain signals. Recording the subject's brain signals involves the following steps: 1- Training; 2- Recording with closed eyes for 1 min; 3- Recording with open eyes for 1 min; 4- Recording while performing the tasks assigned to the subject which includes two parts: A- Remembering the displayed shapes; B- Responding to target and non-target sounds in the auditory stimulation assessment test. In the first stage, after diagnosing the condition of the participants by the physician in three groups, including healthy people and patients of the mild and severe types, the description of all steps of signal recording and training of target and non-target sounds is performed. After preparing the participants, at the second and third steps the EEG signals recorded for 1 min in the closed and opened eye modes respectively. After that, participants were shown Fig. 1 images for 1 min, after which they were asked to close their eyes and review the images in mind. Meanwhile, brain signals were recorded for 1 min. Then they wanted to open their eyes and name the shapes one by one loudly.

Finally, two sounds with different frequencies of 1 and 1.5 kHz are played, one of which is the target sound and the other is the non-target sound. Participants were asked to press the correct key as soon as they heard the sounds. They must click the left key for the target stimulus and the right key for the non-target ones. The total number of stimuli was 120, of which 25% were target sounds. The interval between auditory stimuli was 2 s and each played randomly for 300 ms. For a total of four steps, the total signal recording time was approximately 10 min for each participant.

2.3. ERP signal

The P300 is a specific type of ERP, or in other words, a component of ERP that appears in certain circumstances. According to research done, when the brain processes a new (unusual) stimulus during processing a series of routine stimuli, a P300 wave appears in the recorded brain signal. However, for the production of the P300, it is also necessary for the individual to have a specific task that can only be performed in response to the stimulation of the target, such as asking the participant to count the number of these stimuli [28]. Physically, this component has a positive polarity and has a range of about 10 to 15 μV. For acoustic stimulation, the average P300 wave delay is about 300 ms on average, due to its positive polarity and 300 ms delay, but for other stimuli, such as visual stimulation, this time may be as high as 1000 ms. but on average it has a delay of between 300 and 1000 ms. The location of the signal is recorded from the three channels in the center-line of the head, namely Fz, Cz, Pz. Research has shown that compared to these three channels, most cases of P300 have the highest amplitude in the parietal

Table 1
The MMSE threshold for illiterate and illiterate people.

Condition	Literate	Illiterate
Healthy	$23 \leq MMSE < 30$	$22 \leq MMSE < 30$
Alzheimer's patient	$0 < MMSE < 23$	$0 < MMSE < 22$
Mild AD	$20 \leq MMSE < 23$	$19 \leq MMSE < 22$
Severe AD	$0 < MMSE < 20$	$0 < MMSE < 19$

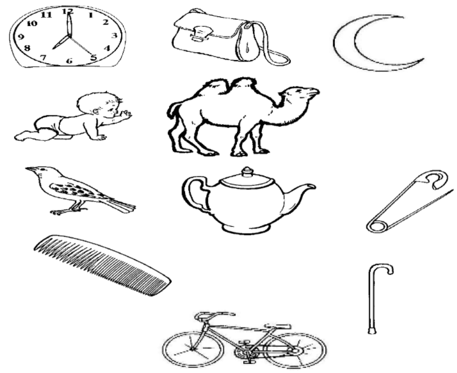


Fig. 1. Images displayed for each participant.

area (Pz) and the lowest amplitude in the frontal area (Fz) (Pz > Cz > Fz). Of course, it is important to note that age, gender, and many other psychological characteristics such as intelligence and personality, etc., affect the amplitude and delay of this wave.

In addition, variations in the magnitude of this wave also depend on the amount of information provided by the stimulus, meaning that the greater the intellectual task required of the individual on unusual stimuli, the greater the P300 wave amplitude extracted. The P3 component amplitude and delay indices of the ERP signal are dependent on the amplitude of the excitation signal, the excitation signal frequency, the excitation intervals, and in particular the probability of occurrence of two excitation bands. P3 amplitude is increased in healthy people. Another important point is that the amplitude and latency of the P3 component changes with increasing age and stages of Alzheimer's disease and even dementia. On the other hand, the range of this component is directly related to how memory operates, meaning that the more a person enjoys a more appropriate memory state, the greater the scope of that component. The delay of this component has been reported approximately 300 ms in various references, which can vary depending on the subject and the record of this component, but another important point is that the shorter latency from 300 ms indicates mental performance better [26]. Fig. 2 shows how the amplitude and delay rate of the ERP signal change in the target stimulation, decreasing the amplitude of the ERP peak component and increasing the delay rate. These changes can be detected in the Pz channel and compared to the healthy group. This signal is manifested by decreasing the sampling rate better [29].

2.4. Research participants

In this study, 40 participants aged 60 to 88 years (mean age 68.43 with a standard deviation of 8.86) were used to record brain signals. All of them were right-handed and there was an equal number of participants in each gender. There were 19 healthy participants with MMSE scores from 23 to 30 with a mean of 27.57 and a standard deviation of 2.29. Eleven participants were in the group of mild patients with an MMSE score range from 19 to 22 with a mean of 20.71 and a standard deviation of 0.95. The last 10 participants were in the severely ill group with an MMSE score range of 3 to 18 with a mean of 13 and a standard deviation of 6.09.

2.5. Brain signal preprocessing

Electroencephalogram (EEG) plays an important role in identifying brain activity and behavior. However, the electrical activity of the brain will always be contaminated with artifacts and types of noise, affecting EEG signal analysis. Therefore, it is necessary to identify effective and accurate methods to eliminate such destructive signals [30]. To process the EEG brain signals, the first step is to remove background noise and interference from the signals. The five steps are intended to remove deviations from the baseline, eliminate high and low-frequency

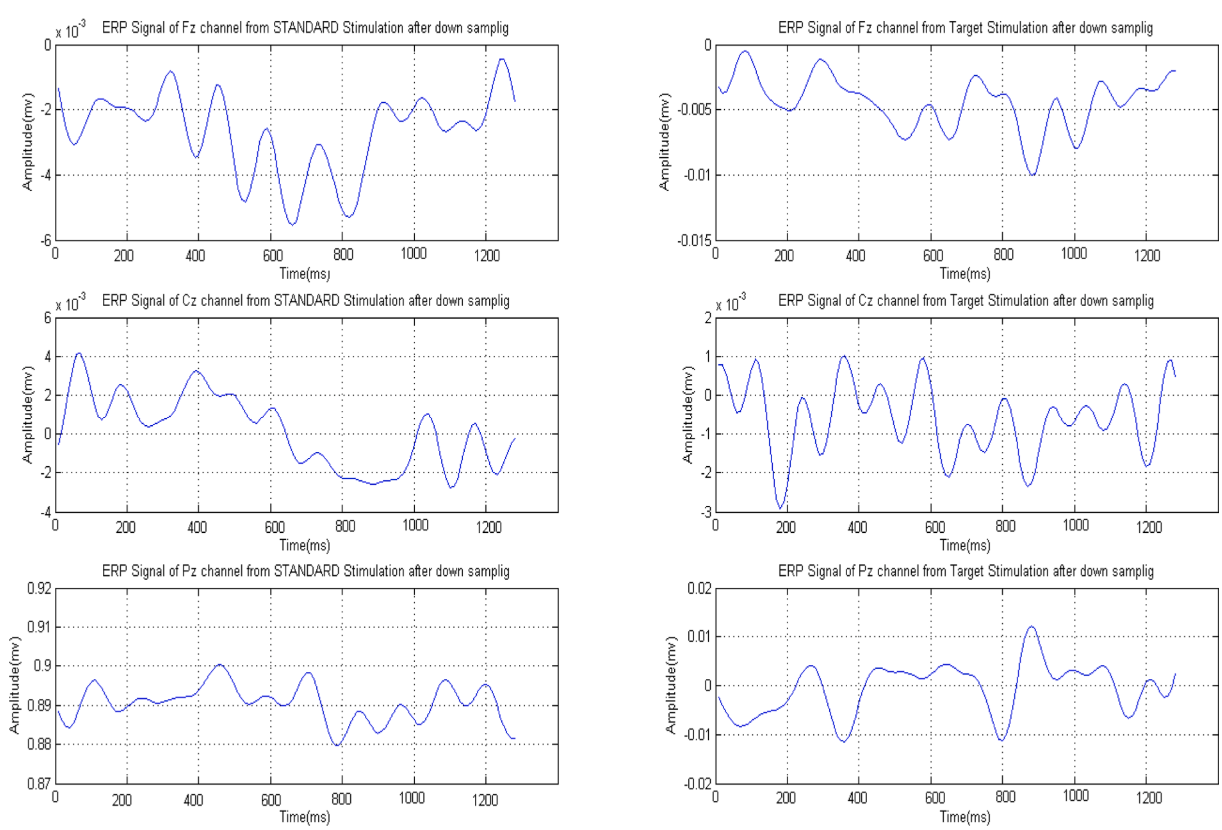


Fig. 2. ERP in the three channels displayed in target and non-target stimulation (standard).

artifacts, eliminate city noise, and reduce sampling rates. Clearly, removing redundant information from the EEG signal improves the quality and accuracy of subsequent processing. The elimination of rapid signal changes plus its baseline fluctuations greatly enhances the processing quality. The recorded EEG signal has a set of artifacts and noises. The artifacts in the EEG signal can be divided into two main parts: high-frequency artifacts caused by EMG of the head and neck muscles and low-frequency artifacts due to electrode movement and transpiration. To eliminate these artifacts and noise from the city, a pass filter with cutoff frequencies of 0.5 to 45 Hz was used [31]. Fig. 3 shows the EEG signal in the Fz, Cz, Pz channels as a frequency spectrum after eliminating city noise and other temporal and frequency interactions.

2.6. Frequency features

Due to the importance of spectral changes in the EEG signal in the frequency bands of Delta, Theta, Alpha, Low Beta, Middle Beta, High Beta, and Gamma, this processing method is used. The usual EEG signal bandwidth is between 0.5 Hz and 45 Hz. Quantitative analyses of EEG signal rhythms are very low cost and contain information relevant to the study of neuronal disorders. This type of analysis involves estimating the power intensity in frequency bands [32]. By comparing the EEG signals of healthy participants and Alzheimer's patients, we see a slow EEG signal. Increases in delta bandwidth (0.5–4 Hz) and theta (4–8 Hz) are seen along with a decrease in alpha bandwidth (8–13 Hz). The slow EEG signal rhythms are related to the early stages of the disease and its progression from the early stages. The neuronal mechanism of alpha-high rhythm is the result of a series of excitatory and inhibitory post-synaptic potentials in the dendrites of pyramidal neurons of the brain. The concurrent activity of a large number of neurons increases the alpha rhythm amplitude and decreases the peak frequency of the alpha power amplitude. The alpha band is considered to be from 8 to 13 Hz, which is due to age-related changes in the frequency window [33]. Frequency evaluation of the EEG signal actually indicates, with respect to various pathologies, an increase or decrease in the alpha band frequency seen, resulting in an alpha peak shift to higher or lower frequencies. One of the important results of EEG signal analysis is the identification of alpha-band frequency sub-bands that represent different cognitive functions. Regarding the diagnosis of mild Alzheimer's patients, it is possible to distinguish features from this group of healthy individuals and severe patients. Based on the power spectrum and frequency analysis, we can see how the power spectrum changes in the early stages of the disease with increasing power. Theta and beta power depletion, along with a decrease in alpha frequency and in more severe stages of the disease, we observed a decrease in alpha activity and an increase in delta power. In later stages, there will be a greater decrease in relative alpha strength and an increase in relative delta strength [34]. Since the alpha frequency band contains more unique information, we divide this band into sub alpha1, alpha2, and alpha3 frequency bands. We first compare the power of different bands for the 3 channels of Fz, Cz, and Pz for the healthy participants, mild patients, and severe patients. Then, for better evaluation, we compare the power of the different bands of these three groups. Figs. 4 to 6 show the variations of the EEG signal power spectrum at different frequency bands for healthy participants, mild and severe patients [35,36].

Fig. 7 compares the power of different Pz channel bands between three groups of healthy subjects, mild patients and severe patients. As it can be seen, the power spectrum increases with the severity of Alzheimer's disease in the lower bands, and sharply decreases in the middle bands, and slightly increases in the upper bands. But the best band for comparison is alpha2 because of significant differences in frequency characteristics. In the alpha frequency bands (alpha1, alpha2, alpha3) with increasing the severity of the disease, the power spectrum is significantly reduced.

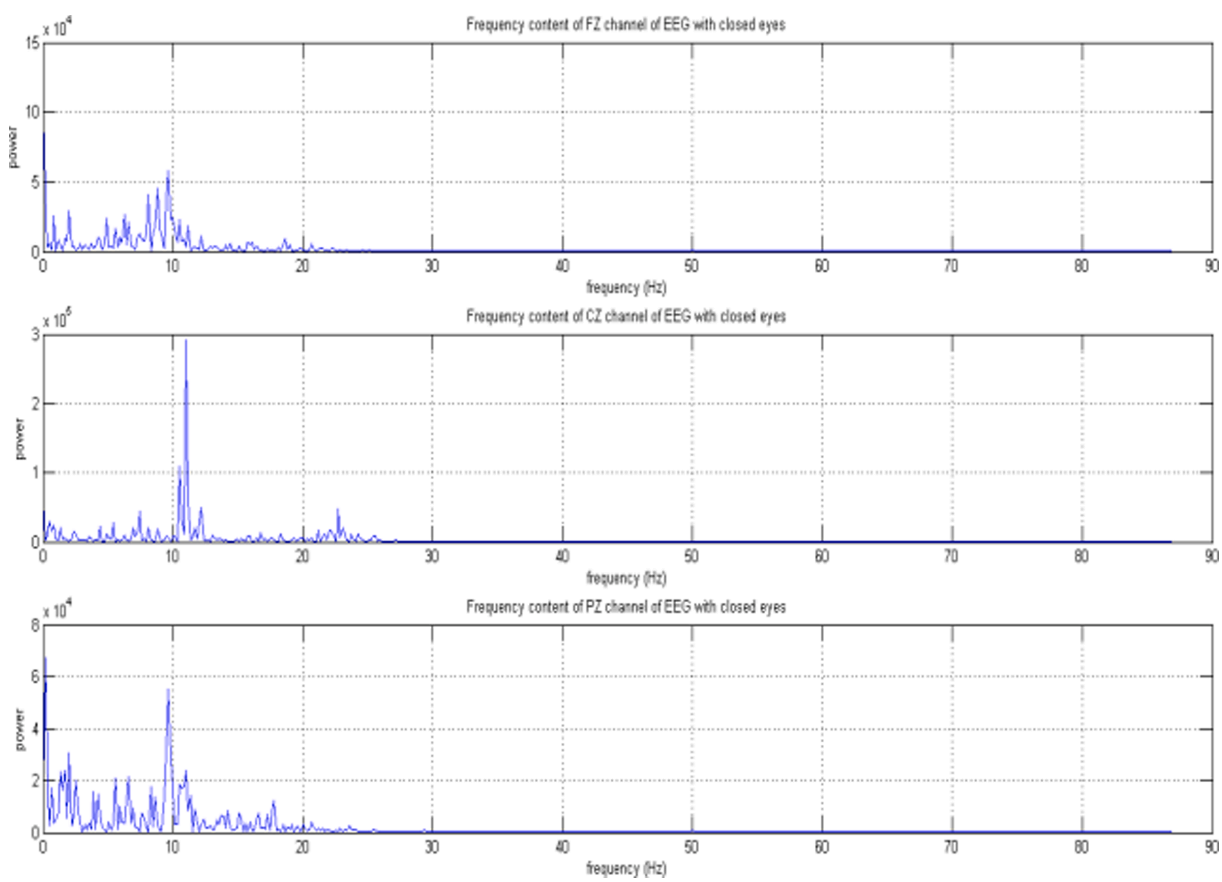
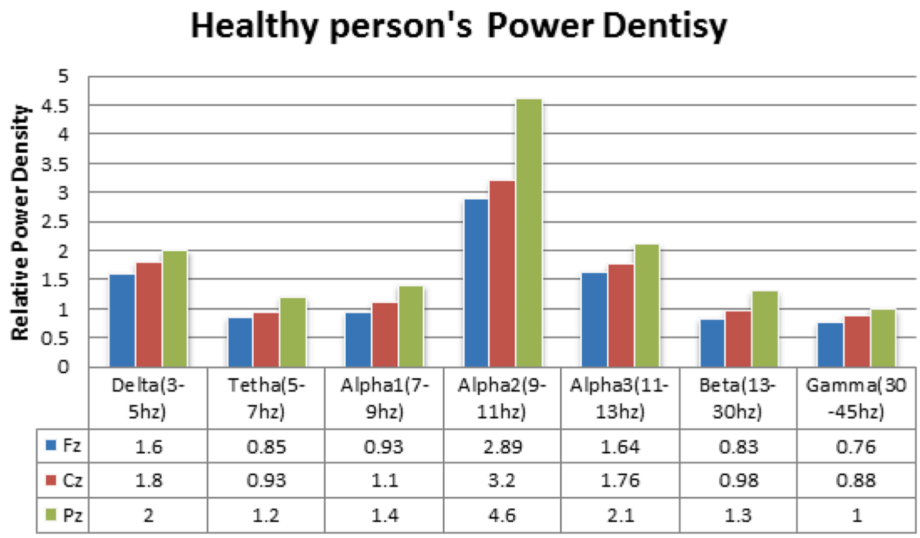
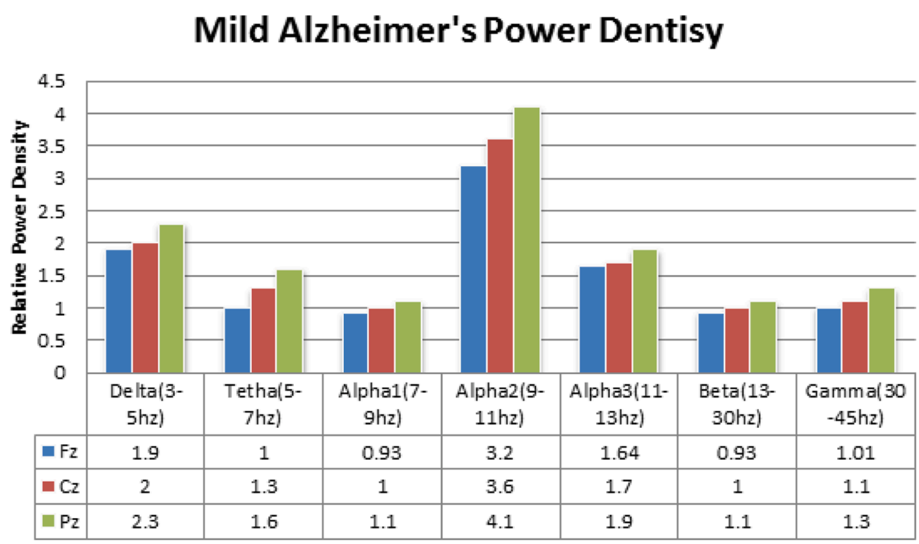


Fig. 3. Display of the frequency spectrum of the signal after pre-processing in closed eye mode.



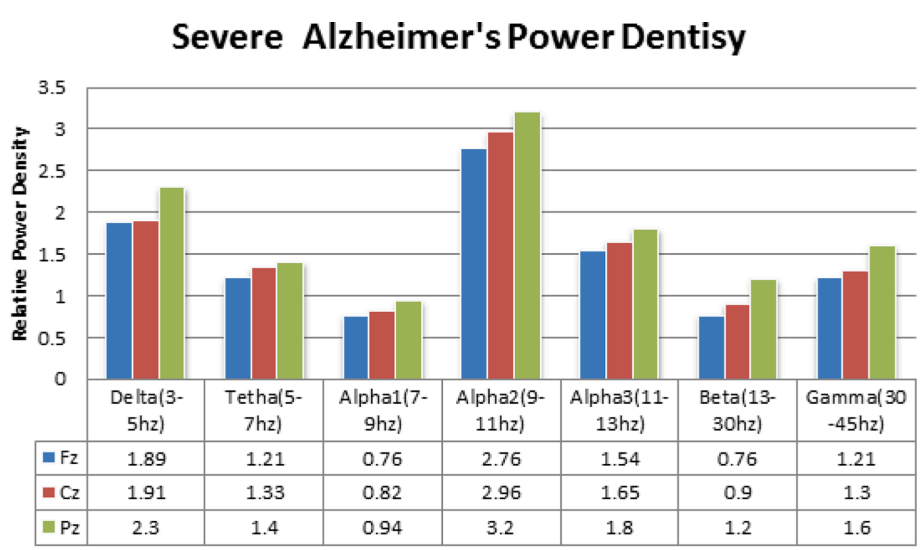
	Delta(3-5hz)	Tetha(5-7hz)	Alpha1(7-9hz)	Alpha2(9-11hz)	Alpha3(11-13hz)	Beta(13-30hz)	Gamma(30-45hz)
Fz	1.6	0.85	0.93	2.89	1.64	0.83	0.76
Cz	1.8	0.93	1.1	3.2	1.76	0.98	0.88
Pz	2	1.2	1.4	4.6	2.1	1.3	1

Fig. 4. Properties of the three channels' power spectrum for the healthy participants.



	Delta(3-5hz)	Tetha(5-7hz)	Alpha1(7-9hz)	Alpha2(9-11hz)	Alpha3(11-13hz)	Beta(13-30hz)	Gamma(30-45hz)
Fz	1.9	1	0.93	3.2	1.64	0.93	1.01
Cz	2	1.3	1	3.6	1.7	1	1.1
Pz	2.3	1.6	1.1	4.1	1.9	1.1	1.3

Fig. 5. Three channel's power spectrum characteristics for the mild patients.



	Delta(3-5hz)	Tetha(5-7hz)	Alpha1(7-9hz)	Alpha2(9-11hz)	Alpha3(11-13hz)	Beta(13-30hz)	Gamma(30-45hz)
Fz	1.89	1.21	0.76	2.76	1.54	0.76	1.21
Cz	1.91	1.33	0.82	2.96	1.65	0.9	1.3
Pz	2.3	1.4	0.94	3.2	1.8	1.2	1.6

Fig. 6. Characteristics of the three-band power spectrum for the severely ill group.

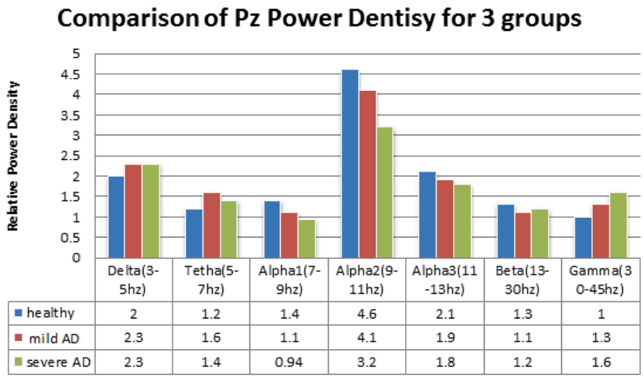


Fig. 7. Comparison of Pz channel power spectrum characteristics for three groups of healthy participants, mild patients and severe patients.

2.7. Feature selection

After appropriate signal processing, statistical properties such as mean, median, mode, maximum, minimum, standard deviation, variance extracted. In addition, frequency characteristics such as the power of delta, theta, alpha, beta, and gamma bands, frequency band shift, and nonlinear properties such as Lyapunov exponent, the correlation dimension, and dynamic changes in brain signal are identified. After extracting the various features, the goal is to calculate and select the optimal features. Analysis of variance (ANOVA) was used to examine the features or channels that have the most impact on the classification. According to the 3 classes of healthy group, mild patient and severe patient, the optimal characteristics can be determined according to the normal distribution of data. First, analysis of variance analyzes features that contain significant differences between the three classes . In this study, 3 EEG channels Fz, Cz, Pz were examined in each period, and 37 features in the closed eye mode, 37 features in the open-eye mode, 37 features in the recall period, and 45 features in the stimulation period were extracted. Analysis of these features and the obtained results expressed in continue. Fig. 8 shows the ANOVA test evaluation for the extracted characteristics of three channels Fz, Cz, and Pz between three groups of healthy, mild and severe patients.

2.8. Analysis of variance

This method is a way to investigate and evaluate the averages of three or more statistical populations using a simple analysis of variance. It is important to note that if we use multiple T-tests instead of this analysis, we will increase the error rate exponentially with the number of tests. Since the paper contains three classes of healthy, mildly ill and severely ill, this analysis was used to derive the optimal feature. To apply this analysis, it is necessary to consider several assumptions: 1- The distribution of the statistical population is normal; 2- The samples (features) in the statistical population are independent of each other; 3- The sample differences in the statistical population should be categorized as one. In this method, three types of variance are defined: 1- intergroup variance, 2- Intra-group variance 3- Total variance. Also, due to the importance of the normality of the data, firstly, by using the Kolmogorov-Smirnov test and P-P graph in SPSS, the normal distribution of the features is investigated. Figs. 9 and 10 show the result of these tests on a number of features [37].

2.9. Classification methods

The ultimate goal in any pattern identification problem is to separate a set of samples into two or more different groups. The aim of this study was to distinguish brain signals in three cases of healthy participants, mild patients, and severe patients [38]. In this study, two classifier types such as linear discriminant analysis, and the Elman and Convolutional neural networks have been used, aiming at comparing static and dynamic classifiers [39]. Linear separation methods, similar to non-

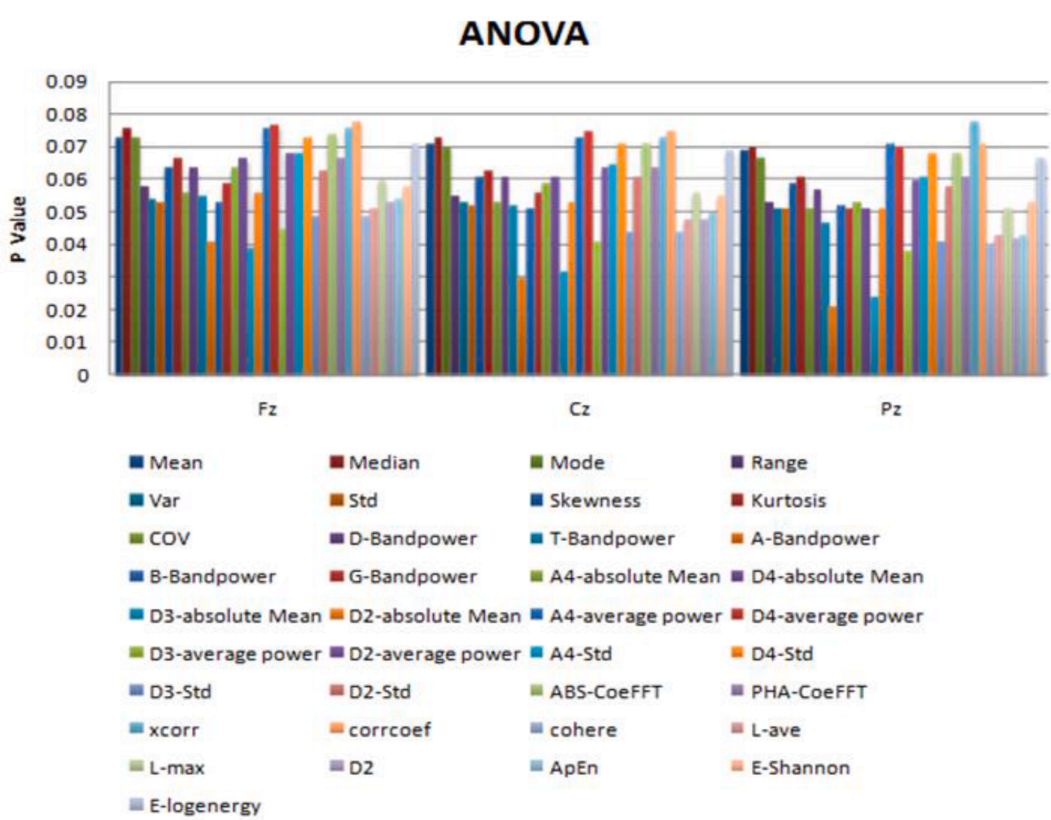


Fig. 8. Evaluation ANOVA test for extracted characteristics of three channels Fz, Cz, Pz between three classes of healthy, mild and severe patients.

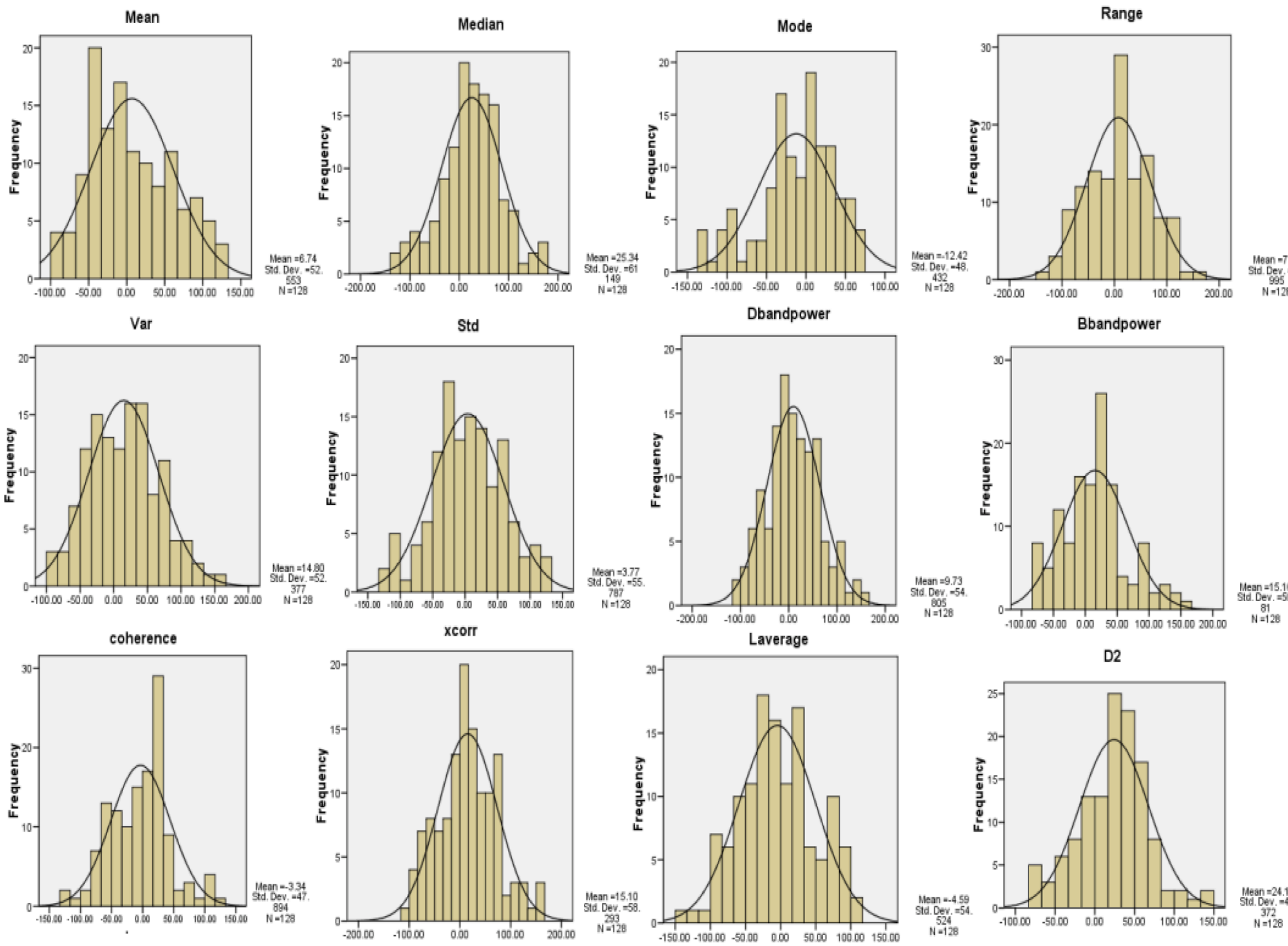


Fig. 9. Examination of the normal distribution of features by Kolmogorov.

parametric methods, do not require knowledge of the distribution function, and by assuming the appropriate form of the separation function as well as training data, the parameters of this method are estimated. In linear separation methods, the combination of feature vector components with a series of specific linear weights is assumed. The decision levels that separate classes are not always linear, and sometimes the complexity of these weights becomes so great that more nonlinear levels are needed. As the linear separator for 3 classes contains poor accuracy results, three classifiers were used for the three modes: healthy versus mild, healthy versus severe, and mild versus severe. In this study, a 2-layer Elman neural network was used which has 8 neurons in the latent layer and 1 neuron in the output layer [40]. The number of input nodes in the input layer was proportional to the number of optimal features selected. Various experiments have been performed with different numbers of hidden layer neurons to reach the best results. In the hidden and output layers, the sigmoid activation function is used, because this function is nonlinear and derivative. There are many training functions for teaching the Elman network, wherein this study the Levenberg-Marquardt error propagation algorithm was used due to higher convergence than other training functions, and the error rate of 0.001 is considered to stop training. At last, the specific architectures of deep learning were examined. The most common form of architectural design, CNN, consists of alternating layers of convolution with pool layers (usually the maximum pool layers). The main features of CNN design are the number of convolution layers and the type of final classifier, the final classification is usually a number of fully connected layers. MLPs are a combination of a number of dense layers and include another type of deep learning algorithm. The next group of architectures in relation to the total number of studies was the RNN, which consists of a number of recurring layers (each layer contains a specific number of studies of the recurring units) followed by a number of fully connected layers. MLPNN, which has the main feature of designing the number of hidden layers as its only examined feature. The final group of studies used SAE, which has the total number of fully connected layers, followed by a fully connected layer in all cases [41]. The processing system we used here to make the classification among categories is based on a deep learning framework for multivariate time-series classification named multi-channels deep convolutional neural networks (MC-DCNN) [42]. Traditional convolutional neural networks (CNN) usually include two parts [43,44]. One is a feature extractor module, which is able to learn features from raw data automatically. The second module is a trainable fully connected MLP, which performs classification based on the learned features from the previous part. Eventually, a fine-tuning supervised step of the whole processing structure can be executed. Generally, the feature extractor is composed of multiple similar stages, and each stage includes three cascading layers: filter (convolutional) layer, activation layer, and pooling layer. The input and output vectors of each layer are referred to as feature maps. In previous works on CNN, the feature extractor usually contains one, two or three such three-layer stages. There are two popular pooling layers, referred to as “average” and “max” pooling; we decided to use the max pooling approach since it is considered to improve the performance. In this work, the CNN input is generated by a preprocessing step of the EEG time recordings [7].

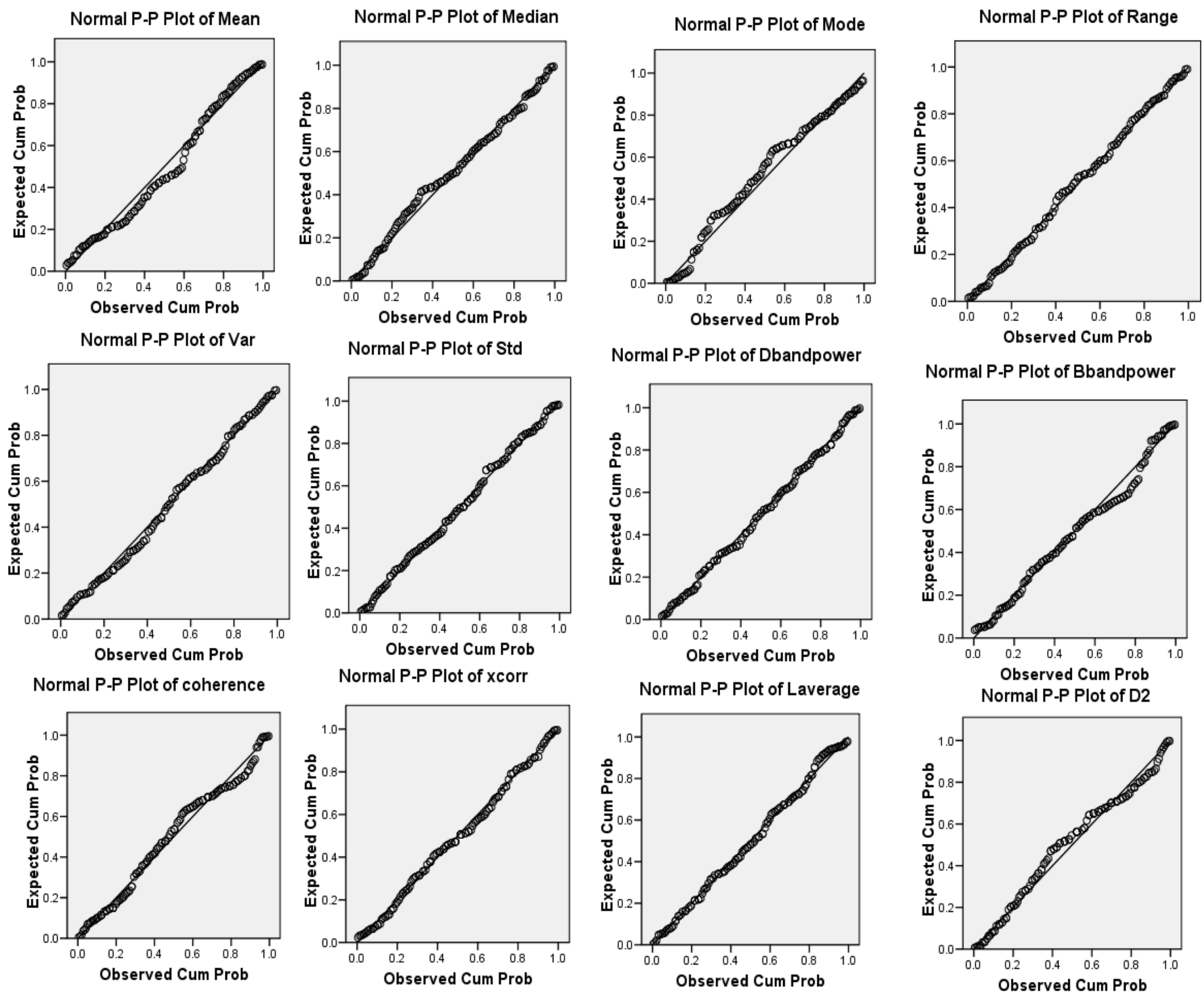


Fig. 10. P-P verification of the normal distribution of features.

3. Results

Selected optimal features by ANOVA method in four modes of the closed eye, open eye, recall, and stimulation are shown in Figs. 11 to 14 to compare the number and types of optimum features between Fz, Cz, and Pz channels.

Then the linear discriminant is applied to optimize the results obtained by ANOVA to separate the three classes. Fig. 15 shows the accuracy of the LDA results on the three Fz, Cz, Pz channels in four different modes. To process and extract the attribute, we call the extracted attributes of each section by labeling it with its own label name. Examples include closed eye section features with index c, open eye section features with index o, periodic recall features with index r, and stimulation period features with index s, which itself includes target stimulation features with st indexes and features Non-target stimulation (standard) with index ss. On the other hand, according to recording in Fz, Cz, Pz channels, F, C, P indices represent the desired channel. The index indicates the position of the signal channel after the index of the different segments. For example, the feature corresponding to the recall period and the Pz channel has the Pzr index and the open eye mode and the Cz channel has the Czo index. Fig. 16 shows the evaluation of closed eye, open eye, recall, and stimulation modes for Fz, Cz, and Pz channels in ANOVA mode.

The Elman neural network confusion matrix for both recall and excitation modes are shown in Figs. 17 and 18.

For evaluating classification results of Elman neural network in recall and excitation modes, different quantitative metrics are calculated and shown in tables 2 and 3. In addition, Fig. 19 illustrates the accuracy of the three-channel results using the optimal ANOVA features in each mode separately.

At last, each EEG channel is divided into non-overlapping windows of 10 s and after separating the frequency bands of Delta, Theta, Alpha, Beta, and Gamma and recording three channels of brain signals in CNN input, it consists of 50x3 = 150 inputs. The CNN we propose here has two hidden convolutional layers. The input vector convolves with 3 masks of 50 elements (one per channel) generating 3 features for each channel. The activation layer with nonlinear logistic sigmoid activation functions has a length of 60. Then, the pooling layer generates a subsampled representation of 3 elements (each one is generated taking the maximum of over 3 entries). The second convolutional layer is generated through the first (compressing) layer of an auto-encoder MLP. In particular, the auto-encoder MLP compresses the 3 outputs of the previous layer in 50 elements that form the input of the final classification NN (50–7-3). The classification NN is trained by supervised learning (backpropagation). The output of the network is a couple of bits (actually, a soft-max nonlinearity is used, thus the sum of the two

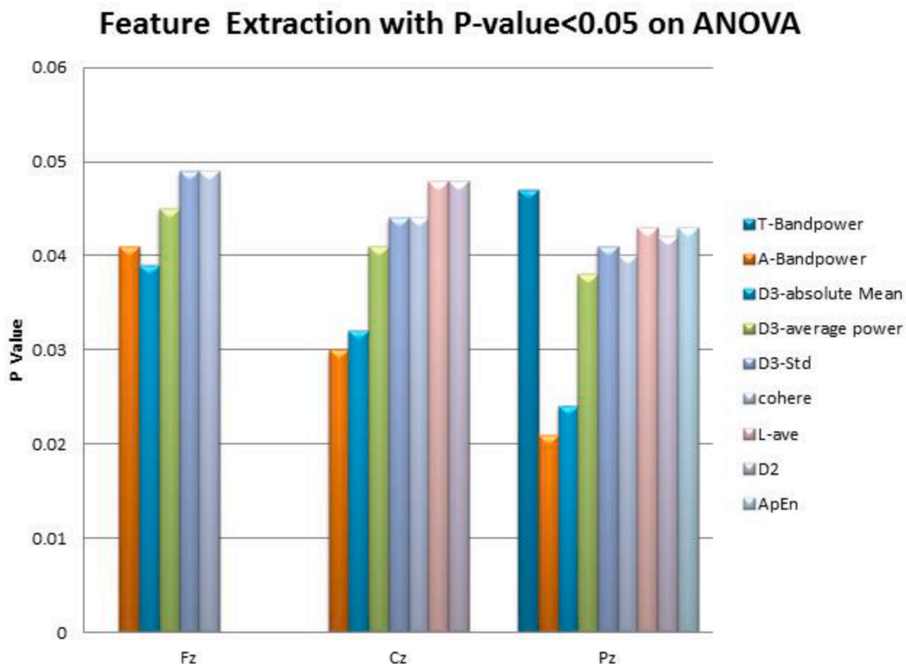


Fig. 11. Optimal properties selected by variance analysis of the three channels of Fz, Cz, and Pz in the closed eye mode.

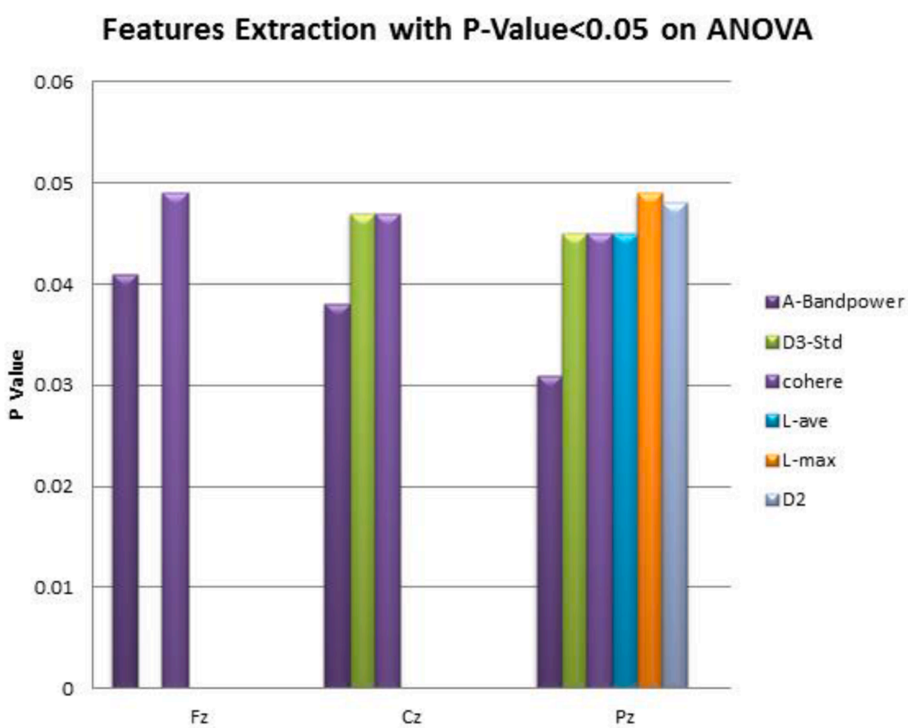


Fig. 12. Optimal properties selected by variance analysis for the three channels of Fz, Cz, Pz in the opened-eye mode.

outputs is constrained to unity), that correspond to the MCI or AD label. 10% of data is used for testing, and 20% for validation, and 70% for train. To re-evaluate the performance accuracy, we can also use the predicate function to calculate the matrix configuration, and before that, the evaluation is determined for the trained data. Loss diagrams and performance are shown in Figs. 20 to 21.

The convolutional neural network confusion matrix for recall modes is shown in Fig. 22. In addition, the classifier's evaluating criterions such as precision, sensitivity, specificity, and F-score are shown in Table 4.

4. Discussion

Alzheimer is a disease of the nervous tissues that has special characteristics associated with intellectual and cognitive impairment. In fact, the EEG signal has been the diagnostic tool for decades. Abnormal signals in this signal include shifting the power spectrum to lower frequencies and decreasing correlation in fast rhythms. These abnormal features are related to the lack of connection in the cortical regions of the brain, which results in the death of cortical neurons, axonal damage, and

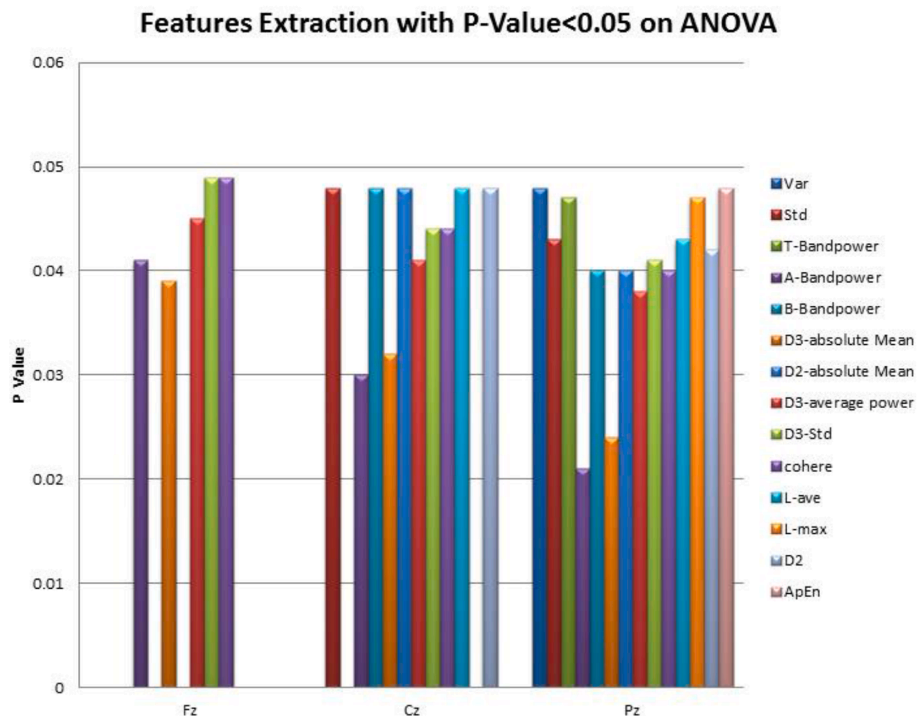


Fig. 13. Variance analysis selected by variance analysis for the three channels of Fz, Cz, Pz in the recall mode.

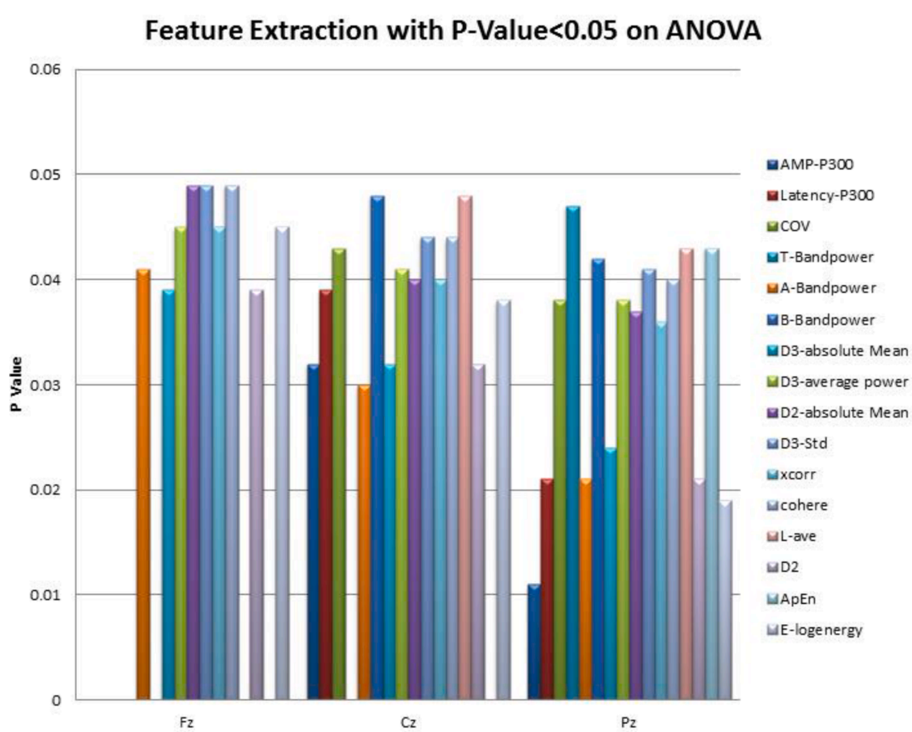


Fig. 14. Optimal properties selected by variance analysis for the three channels of Fz, Cz, and Pz in the excitation mode.

dysfunction of acetylcholine-containing nerve fibers. Nonlinear changes in the brain signal of Alzheimer's patients are an important and influential issue. The reduction in the complexity of the brain signal pattern is the result of reduced information transmitted between cortical regions. The brain signal in Alzheimer's disease has other characteristics such as reduced mean frequencies and reduced complex and correlated activities. Abnormal symptoms of the disease affect the EEG signal, including decreased rhythms (slow signal) and reduced dependence between different areas of the brain.

The use of EEG signal is important for several reasons. The first reason is that Alzheimer's disease is a cortical disease whose abnormalities is well visible in this brain signal and has been comparable to the normal signal. Abnormal states of Alzheimer's disease in the brain signal directly reflect the functional and anatomical defects of the damaged cerebral cortex. Therefore, research on EEG signal dynamics in relation to neuronal injury has been very important. Another reason is

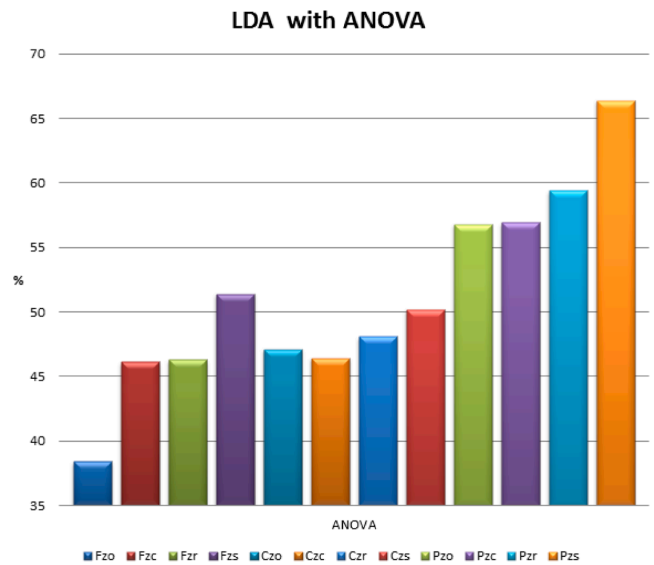


Fig. 15. LDA accuracy of three-channel Fz, Cz, Pz signal separation results using selected ANOVA features.

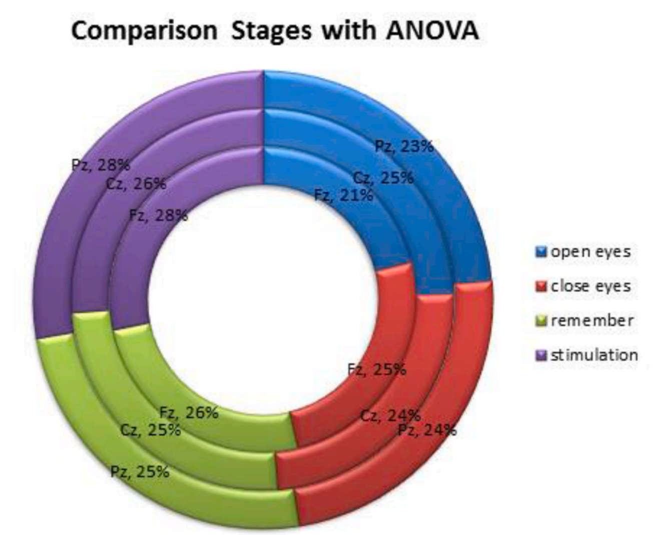


Fig. 16. Evaluation of closed eye, open eye, recall and excitation modes for Fz, Cz, and Pz channels with ANOVA.

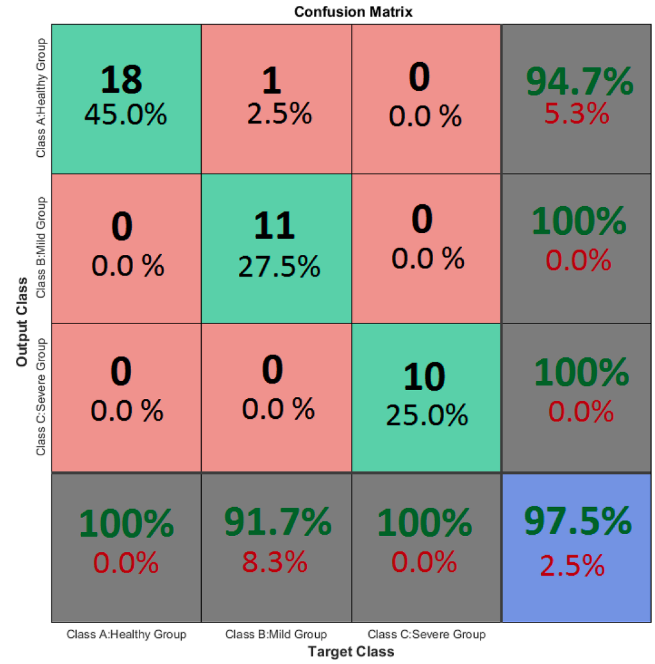


Fig. 17. Elman neural network confusion matrix in recall mode.

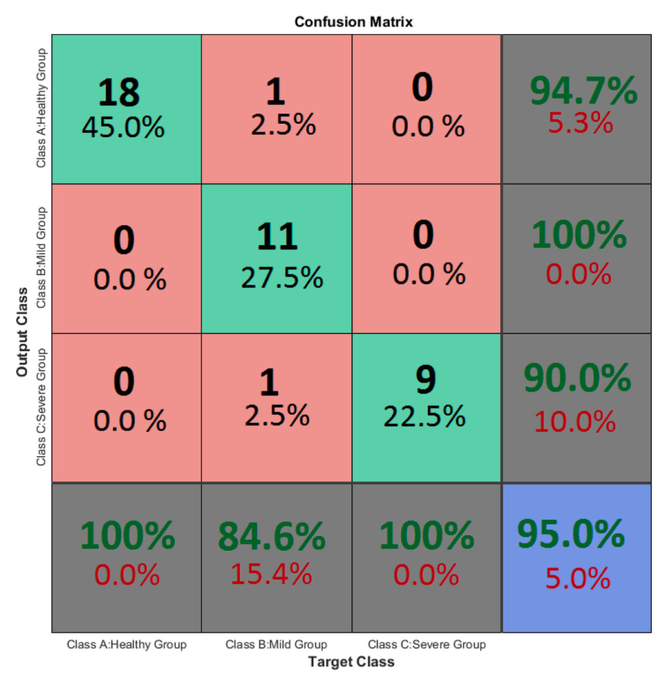


Fig. 18. Elman neural network confusion Matrix in excitation mode.

that non-invasiveness can be considered a privilege according to signal analysis and investigation of the mechanism of the disease. In this way, the cortical synaptic function is evaluated. Flexible states of synapses are considered a crisis for brain function, especially in learning and memory. The interference of synaptic connections is the cause of many neurological diseases. The study of nonlinear dynamics of this signal in Alzheimer's disease indicates nonlinear brain activity at different stages of the disease. Nonlinear dynamic analysis of this signal shows a decrease in the complexity of the brain signal pattern and a decrease in connections due to a decrease in the nonlinear cell dynamics between cortical regions.

Frequency features indicate that the first changes in the brain signal are an increase in theta activity and a decrease in beta activity, which is accompanied by a decrease in alpha activity. Increased delta activity occurs in the later stages of the disease. In patients with higher stages and severe progression of the disease, we see decreased alpha activity and increased delta activity. In the middle stages, a decrease in beta activity and an increase in theta activity in the brain signal can be observed. The lack and decline of functional interactions in the cerebral cortex is a possibility for abnormal cognitive activity in Alzheimer's disease. The P3 component of the ERP signal delay depends on the

Table 2
The values of precision, sensitivity and precision in different groups in recall mode for Elman neural network.

Group	Precision	Sensitivity	Specificity	F-Score
Healthy	100%	89.5%	90.9%	94.45%
Mild	83.3%	90.9%	100%	86.93%
Severe	90.9%	100%	89.5%	95.23%

Table 3
The values of precision, sensitivity and precision in different groups in excitation mode for Elman neural network.

Group	Precision	Sensitivity	Specificity	F-Score
Healthy	100%	94.7%	100%	97.27%
Mild	84.5%	100%	90%	91.65%
Severe	100%	90%	94.7%	94.73%

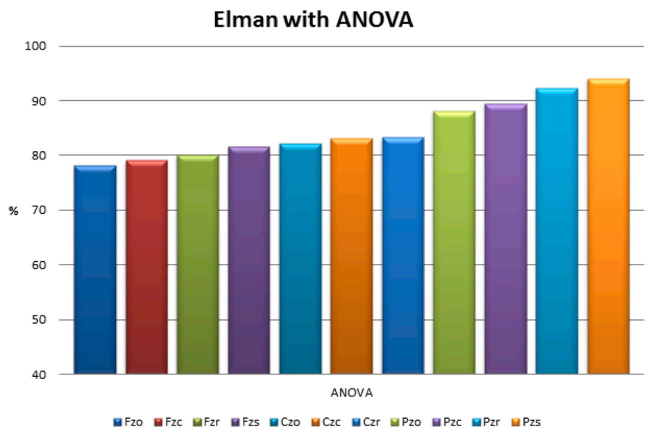


Fig. 19. Classification accuracy results of Elman Neural Network with selected ANOVA features of three channels Fz, Cz, and Pz in different modes.

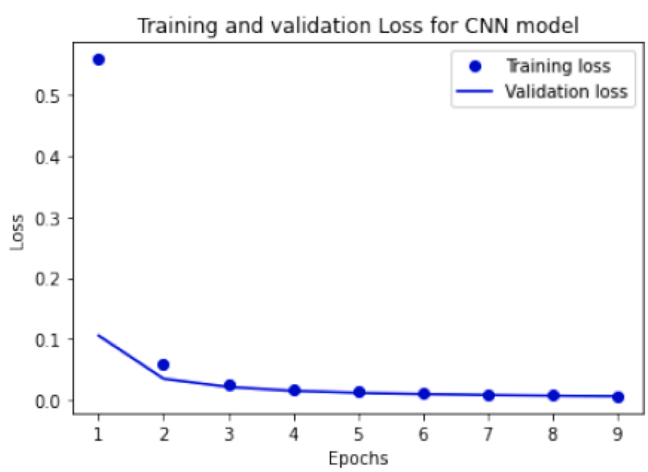


Fig. 20. Training and validation loss for CNN model.

amplitude of the excitation signal, the frequency of the excitation signal, the excitation interval, and in particular the probability of occurrence of two excitation bands. However, in the past research, only the P3 component domain index has been investigated [45]. If the frequency of target stimuli decreases or the intervals between stimuli increases, the amplitude of the P3 component in healthy individuals increases. The delay of this component has been reported approximately 300 ms in various references, which can vary depending on the subject and the record of this component, but another important point is that the shorter latency from 300 ms indicates mental performance better. Another important point is that the amplitude and latency of the P3 component changes with increasing age and stages of Alzheimer's disease and even dementia. On the other hand, the domain of this component is directly related to how memory operates, meaning that the more a person has a more appropriate memory state, the more the domain of this component increases. The subjects studied modes are closed eye, open eye, recall, and stimulation, and among these four periods, the best period for evaluating and extracting attributes is the recall and stimulation periods,

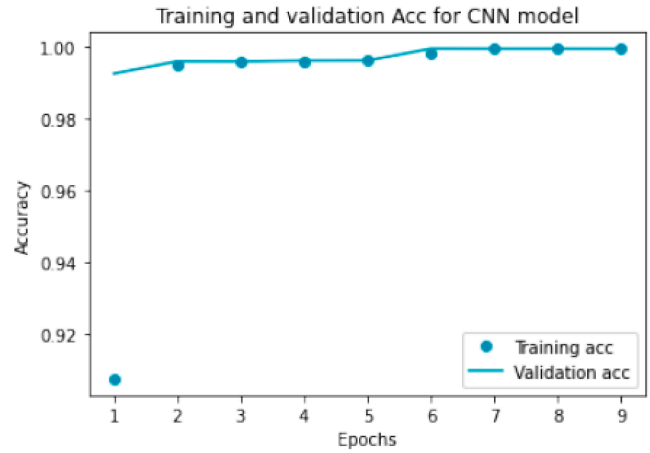


Fig. 21. Training and validation ACC for CNN model.

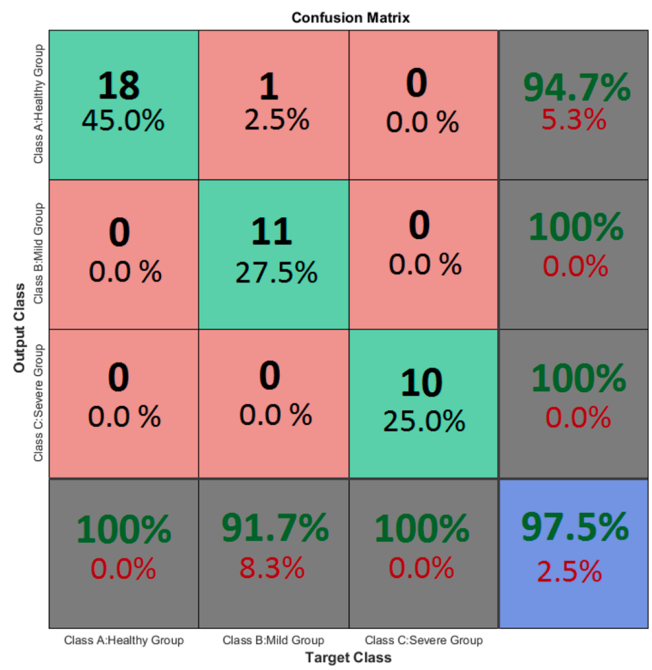


Fig. 22. Convolutional neural network confusion matrix for recall mode.

Table 4
The values of precision, sensitivity and precision in different groups in recall mode for convolutional neural network.

Group	Precision	Sensitivity	Specificity	F-Score
Healthy	100%	94.7%	100%	97.27%
Mild	91.7%	100%	100%	95.6%
Severe	100%	100%	94.7%	100%

since using evaluation by variance analysis methods the number of optimal features in these two periods is higher than the others. On the other hand, among the three channels, Fz, Cz, and Pz, for brain signal recording, the Pz channel exhibits better features for the diagnosis of Alzheimer's disease, because of the number of optimal features in this channel is higher than the other two channels. The Pz channel in the recall period by variance analysis has 14 optimal features. While the number of these features in other channels is <2 to 7 features. Among the processing methods proposed to classify three classes of healthy, mildly ill and severely ill, the nonlinear methods have higher accuracy than the linear separation method. The accuracy of the Pz channel results in recall and stimulation with selected features of linear analysis of variance was 59.4% and 66.4% respectively and in the neural network, they have 92.5% and 95% accuracy and with the CNN 97.5% an 99% respectively.

To compare the results of the proposed method to other research an available EEG database [16] of two groups of mild cognitive impairment due to Alzheimer's diseases and healthy elderly participants was evaluated. Comparison between the group of healthy participants and mild patients shows that the alpha bandwidths (alpha1, alpha2, alpha3) in mild Alzheimer's patients has decreased in comparison to healthy individuals, and also this trend has been seen for the delta-band. The highest classification accuracy between these two groups was in the alpha-2 sub-band and it was about 86.7%. However, based on the proposed method, the accuracy of the available database classification by using the Elman classifier was 97.3%, 95%, and 74.1% for the extracted features of Pz, Cz, and Fz channels respectively. In addition, by using CNN the accuracy of the classifier was 100%, 76.6%, and 66.6% respectively.

Based on these results, dynamic classifiers such as the convolutional neural network had better performance than LDA and Elman for the diagnosis of mild Alzheimer's disease, especially by extracting appropriate linear and nonlinear features. The high accuracy of deep learning algorithms in solving complex decision problems attracted the attention of researchers in various sciences. Although the proper performance of these methods has expanded their application in medical diagnoses and classifications, the description of performance due to the non-linear nested structures prevents the possibility of explaining their decisions or acts to users. Therefore, it is valuable to evaluate and explain how these algorithms work using different methods such as LIME (Local Interpretable Model-Agnostic Explanations) [46] and Grad-CAM (Gradient-weighted Class Activation Mapping) [47]. This suggestion can be used as a future direction to explore and improve the performance of the model.

5. Conclusion

In conclusion, by analyzing the information extracted from the P3 component of the ERP signal, the present study investigated how the combination of proper linear and nonlinear features extracted from ERP and EEG signals can improve the chance of early mild Alzheimer's disease detection. In addition, the extracted frequency features indicate that the first changes in the brain signal are an increase in theta activity and a decrease in beta activity, which is accompanied by a decrease in alpha activity. Due to the importance of the alpha frequency band which is one of the best criterion measures of the subject's level of cognition and attention, this frequency band is divided into three parts alpha1, alpha2, and alpha3. The results have shown the effect of these sub-bands in increasing the accuracy of classifiers. The best sub-band for comparison is alpha2 because of significant differences in frequency characteristics. Comparing different classifiers has shown that in regard to the dynamic nature of the brain signal and the capability of convolutional neural network in learning unsupervised from data that is unstructured or unlabeled, this classifier contains higher and proper accuracy for the diagnosis of mild Alzheimer's disease.

CRediT authorship contribution statement

Elias Mazrooei Rad: Conceptualization, Data curation, Software, Formal analysis, Writing - original draft. **Mahdi Azarnoosh:** Supervision, Methodology. **Majid Ghoshuni:** Validation, Writing - review & editing. **Mohammad Mahdi Khalilzadeh:** Validation, Writing - review & editing.

Declaration of Competing Interest

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

References

[1] F. Roselli, B. Tartaglione, F. Federico, V. Lepore, G. Defazio, P. Livrea, Rate of MMSE score change in Alzheimer's disease: influence of education and vascular risk factors, Clin. Neurol. Neurosurg. 111 (4) (2009) 327–330.
[2] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, F.C. Morabito, A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings, Neurocomputing 323 (2019) 96–107.
[3] J.A. van Deursen, E.F.P.M. Vuurman, V.H.J.M. van Kranen-Mastenbroek, F.R. J. Verhey, W.J. Riedel, 40-Hz steady state response in Alzheimer's disease and mild cognitive impairment, Neurobiol. Aging 32 (1) (2011) 24–30.
[4] R.M. Chapman, G.H. Nowlis, J.W. McCrary, J.A. Chapman, T.C. Sandoval, M. D. Guillily, M.N. Gardner, L.A. Reilly, Brain event-related potentials: diagnosing early-stage Alzheimer's disease, Neurobiol. Aging 28 (2) (2007) 194–201.
[5] N. Malek, M.R. Baker, C. Mann, J. Greene, Electroencephalographic markers in dementia, Acta Neurol. Scand. 135 (4) (2017) 388–393.
[6] X. Bi, H. Wang, Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning, Neural Networks 114 (2019) 119–135.
[7] Morabito, Francesco Carlo, Maurizio Campolo, Cosimo Ieracitano, Javad Mohammad Ebadi, Lilla Bonanno, Alessia Bramanti, Simona Desalvo, Nadia Mammone, and Placido Bramanti. "Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings." In 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), pp. 1-6. IEEE, 2016.
[8] J. Omedes, I. Iturrate, L. Montesano, J. Minguez, Using frequency-domain features for the generalization of EEG error-related potentials among different tasks, in: In 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2013, pp. 5263–5266.
[9] J.-L. Song, R. Zhang, Automatic seizure detection using a novel EEG feature based on nonlinear complexity, in: In 2016 international joint conference on neural networks (IJCNN), 2016, pp. 1686–1695.
[10] R. Majid Mehmood, R. Du, H.J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors." Ieee, Access 5 (2017) 14797–14806.
[11] X. Li, C. Guan, H. Zhang, K.K. Ang, Discriminative ocular artifact correction for feature learning in EEG analysis, IEEE Trans. Biomed. Eng. 64 (8) (2017) 1906–1913.
[12] G. Vrbancic, V. Podgorelec, Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks, Elektronika ir Elektrotechnika 24 (4) (2018) 3–7.
[13] S. Kuanar, V. Athitsos, N. Pradhan, A. Mishra, K.R. Rao, Cognitive analysis of working memory load from EEG, by a deep recurrent neural network, in: In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 2576–2580.
[14] Z. Jiao, X. Gao, Y. Wang, J. Li, H. Xu, Deep convolutional neural networks for mental load classification based on EEG data, Pattern Recogn. 76 (2018) 582–595.
[15] J. Wagner, T. Solis-Escalante, P. Grieshofer, C. Neuper, G. Müller-Putz, R. Scherer, Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects, Neuroimage 63 (3) (2012) 1203–1211.
[16] Babiloni, Claudio, Claudio Del Percio, Roberta Lizio, Giuseppe Noce, Susanna Lopez, Andrea Soricelli, Raffaele Ferri et al. "Abnormalities of resting state cortical EEG rhythms in subjects with mild cognitive impairment due to Alzheimer's and Lewy body diseases." Journal of Alzheimer's Disease 62, no. 1 (2018): 247-268.
[17] Y.R. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals, J. Neural Eng. 14 (1) (2017) 016003, https://doi.org/10.1088/1741-2560/14/1/016003.
[18] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, H. Wang, Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition, Concurrency and Computation: Practice and Experience 30 (23) (2018), e4413.
[19] X. Wei, L. Zhou, Z. Chen, L. Zhang, Y.i. Zhou, Automatic seizure detection using three-dimensional CNN based on multi-channel EEG, BMC Med. Inf. Decis. Making 18 (5) (2018) 71–80.
[20] R. Hussein, H. Palangi, R.K. Ward, Z.J. Wang, Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals, Clin. Neurophysiol. 130 (1) (2019) 25–37.
[21] H. Xu, K.N. Plataniotis, Affective states classification using EEG and semi-supervised deep learning approaches, in: In 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), 2016, pp. 1–6.
[22] Bojarski, Mariusz, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).
[23] S.M. Mathews, L.F. Polania, K.E. Barner, Leveraging a discriminative dictionary learning algorithm for single-lead ECG classification, in: In 2015 41st Annual Northeast Biomedical Engineering Conference (NEBEC), 2015, pp. 1–2.

[24] S.M. Mathews, C. Kambhamettu, K.E. Barner, A novel application of deep learning for single-lead ECG classification, Comput. Biol. Med. 99 (2018) 53–62.

[25] S.M. Mathews, Dictionary and deep learning algorithms with applications to remote health monitoring systems, University of Delaware, 2017.

[26] J. Polich, J. Corey-Bloom, Alzheimer's disease and P300: review and evaluation of task and modality, Curr. Alzheimer Res. 2 (5) (2005) 515–525.

[27] C.E. Jackson, P.J. Snyder, Electroencephalography and event-related potentials as biomarkers of mild cognitive impairment and mild Alzheimer's disease, Alzheimer's & Dementia 4 (1S1) (2008) S137–S143.

[28] M.-S. Lee, S.-H. Lee, E.-O. Moon, Y.-J. Moon, S. Kim, S.-H. Kim, I.-K. Jung, Neuropsychological correlates of the P300 in patients with Alzheimer's disease, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 40 (2013) 62–69.

[29] V. Jelic, J. Kowalski, Evidence-based evaluation of diagnostic accuracy of resting EEG in dementia and mild cognitive impairment, Clinical EEG and Neuroscience 40 (2) (2009) 129–142.

[30] X. Jiang, G.-B. Bian, Z. Tian, Removal of artifacts from EEG signals: a review, Sensors 19 (5) (2019) 987.

[31] C. Micanovic, S. Pal, The diagnostic utility of EEG in early-onset dementia: a systematic review of the literature with narrative analysis, J. Neural Transm. 121 (1) (2014) 59–69.

[32] R.P. Brenner, C.F. Reynolds, R.F. Ulrich, Diagnostic efficacy of computerized spectral versus visual EEG analysis in elderly normal, demented and depressed subjects, Electroencephalogr. Clin. Neurophysiol. 69 (2) (1988) 110–117.

[33] H. Hampel, S. Lista, S.J. Teipel, F. Garaci, R. Nisticò, K. Blennow, H. Zetterberg, L. Bertram, C. Duyckaerts, H. Bakardjian, A. Drzezga, O. Colliot, S. Epelbaum, K. Broich, S. Lehéricy, A. Brice, Z.S. Khachaturian, P.S. Aisen, B. Dubois, Perspective on future role of biological markers in clinical therapy trials of Alzheimer's disease: a long-range point of view beyond 2020, Biochem. Pharmacol. 88 (4) (2014) 426–449.

[34] A.A. Gouw, C.J. Stam, Electroencephalography in the differential diagnosis of dementia, Epileptologie 33 (2016) 173–182.

[35] K. Strimbu, J.A. Tavel, What are biomarkers? Current Opinion in HIV and AIDS 5 (6) (2010) 463–466.

[36] I. Szirmai, A. Kamondi, EEG investigations in cognitive impairments, Ideggyogyaszati szemle 64 (1–2) (2011) 14–23.

[37] F.Z. Brill, D.E. Brown, W.N. Martin, Fast generic selection of features for neural network classifiers, IEEE Trans. Neural Networks 3 (2) (1992) 324–328.

[38] R.H. Chowdhury, M.BI. Reaz, M.A.B.M. Ali, A.AA. Bakar, K. Chellappan, T. G. Chang, Surface electromyography signal processing and classification techniques, Sensors 13 (9) (2013) 12431–12466.

[39] C.D. Papadaniil, V.E. Kosmidou, A. Tsolaki, M. Tsolaki, I.(. Kompatsiaris, L. J. Hadjileontiadis, Cognitive MMN and P300 in mild cognitive impairment and Alzheimer's disease: A high density EEG-3D vector field tomography approach, Brain Res. 1648 (2016) 425–433.

[40] K.J. Friston, K.E. Stephan, R. Montague, R.J. Dolan, "Computational psychiatry: the brain as a phantastic organ." The Lancet, Psychiatry 1 (2) (2014) 148–158.

[41] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural Eng. 16 (3) (2019) 031001, https://doi.org/10.1088/1741-2552/ab0ab5.

[42] Y.i. Zheng, Q.i. Liu, E. Chen, Y. Ge, J.L. Zhao, "Exploiting multi-channels deep convolutional neural networks for multivariate time series classification." Frontiers of Computer, Science 10 (1) (2016) 96–112.

[43] Y.i. Zheng, Q.i. Liu, E. Chen, Y. Ge, J. Leon Zhao, in: Time series classification using multi-channels deep convolutional neural networks, Springer, Cham, 2014, pp. 298–310.

[44] C.A.M. Lima, A.L.V. Coelho, M. Eisencraft, Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study, Comput. Biol. Med. 40 (8) (2010) 705–714.

[45] D. Hedges, R. Janis, S. Mickelson, C. Keith, D. Bennett, B.L. Brown, P300 amplitude in Alzheimer's disease: a meta-analysis and meta-regression, Clinical EEG and neuroscience 47 (1) (2016) 48–55.

[46] Mathews, Sherin Mary. "Explainable artificial intelligence applications in NLP, biomedical, and malware classification: a literature review." In Intelligent Computing-Proceedings of the Computing Conference, pp. 1269-1292. Springer, Cham, 2019.

[47] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: In Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.