

Identification of Alzheimer's disease from central lobe EEG signals utilizing machine learning and residual neural network



Islam A. Fouad^{a,*}, Fatma El-Zahraa M. Labib^b

^a Biomedical Engineering Dept., Misr University for Science and Technology, 6th October, Giza, Egypt

^b Department of Biomedical Engineering, New Ksar Ainy Teaching Hospital, Giza, Egypt

ARTICLE INFO

Keywords:

Alzheimer's disease
EEG signals
Pre-processing
Feature extraction
Classification
Support Vector Machine (SVM)
Random Forest (RF)
Logistic Regression (LR)
Resnet-50 CNN

ABSTRACT

Cognitive and behavioral deficits are some of the symptoms of Alzheimer's disease, a neurological disease caused by brain deterioration. Early diagnosis of the disease minimizes the disease's progression. In this way, patients' living standards can be maintained. An experienced specialist will assess the results of grueling diagnostic tests and a costly diagnosis process.

This motivation led to the creation of a fully automated computer system, which uses the proposed methods to detect Alzheimer's disease from EEG signals acquired from a minimum number of electrodes: three central lobe electrodes. This work presents the advantages of implementing the proposed system. Two different datasets were presented to evaluate the system's performance. After preprocessing the raw EEG data, wavelet transforms were applied to calculate statistical properties. First, the paper discusses the performance of different traditional machine learning classifiers: Diagonal Discriminant Analysis, Support Vector Machine (LSVM, RBF), K-Nearest Neighbors, Random Forest, Logistic Regression, and Naïve Bayes. By comparing the classifiers' performance, it is found that Naïve Bayes and LSVM classifiers yield the highest performance of 96.55% and 95.69% when applied to the first dataset and 96.55% and 94.52% for the second dataset, respectively. "ResNet-50" Convolutional Neural Network classifier was implemented to demonstrate the capability of constructing an effective AD diagnosing system. It yields an accuracy of 97.8261% when applied to the first dataset.

This work shows how Deep Learning improves the performance of classification in early medical diagnosis, which will enhance different diseases' treatment, as "Resnet - 50" convolutional neural network classifier outperformed Naïve Bayes Classifier.

1. Introduction

In addition to being the most prevalent form of age-related dementia, Alzheimer's disease is a major neurological disorder throughout the world. Brain activities in Alzheimer's patients are slower than in healthy people, and the most important symptom is the deterioration of cognitive functions [1,2]. Globally, 50 million people were affected by this disease in 2018, and the number is expected to rise to 82 million in 2030 [3]. There are three types of Alzheimer's disease: mild, moderate, and severe. A patient with Alzheimer's disease progressively loses cognitive functions over time. Early diagnosis of Alzheimer's is important to prevent the disease from progressing rapidly. For this reason, there are many related studies in the detection of Alzheimer's disease.

Electroencephalography (EEG) signals are an important tool widely used in the study of brain functions. EEG signals are widely used in the

development of computer-assisted diagnostic systems aiming at diagnosing many diseases due to the lower cost and lower time requirement compared to other biomedical imaging methods (such as CT-Computed Tomography, MRI Magnetic Resonance Imaging, PET- Positron Emission Tomography, fMRI - Functional Magnetic Resonance Imaging).

Various algorithms and new processing techniques are presented for the analysis of EEG signals. There are limitations that affect the performance of Computer Aided Diagnosis (CAD) systems that detect Alzheimer's from EEG signals. The inability to determine appropriate feature selection methods in the relevant literature studies or the failure to consider the EEG channels that carry important information are important factors for these limitations.

Various analysis methods are used to perform feature extraction directly from the signal in EEG-based CAD applications. A wavelet transform reveals important information in a signal's time-frequency

* Corresponding author.

E-mail addresses: islam.fouad@must.edu.eg (I.A. Fouad), fatma_zahraa_mlabib@yahoo.com (F. El-Zahraa M. Labib).

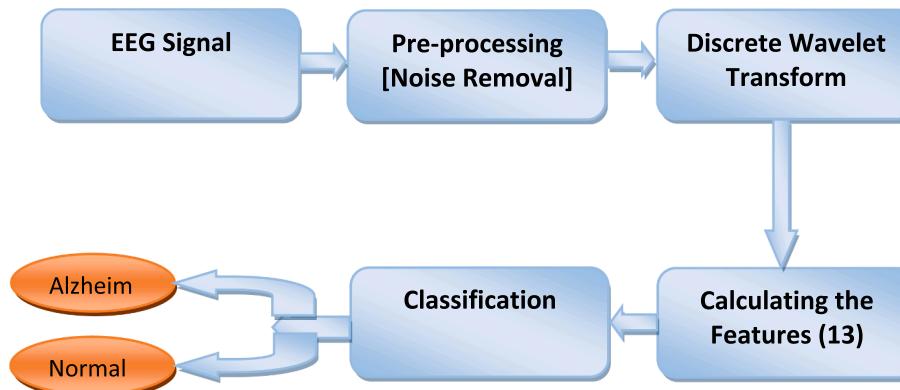


Fig. 1. Flow chart including the stages of the proposed method.

domain through analysis of frequency components [4]. However, the use of Hjorth parameters [5] has recently been preferred in many EEG-based diagnostic CAD systems and high classification performance has been achieved. In addition, the computational complexity of wavelet transform is much lower than other signal decomposition methods and other feature extraction methods of Hjorth parameters.

It has been proved that machine learning techniques can be used in a variety of fields and produce excellent results. In the field of environment, Buyrukolu et al. investigated the accuracy of the correlation value determined for the prediction of *Salmonella* in agricultural surface water [6]. In the field of sports, researchers proposed a stacked ensemble learning model to determine the positions of football players and achieved 83.9% classification success [7].

In another study carried out in the field of education, a literature review on the application of different Machine Learning algorithms on a large amount of data collected by academic institutes is presented [8].

To distinguish Alzheimer's disease, apart from the biomedical imaging methods mentioned above, there are many literature studies performed by analyzing EEG signals and classifying them with machine learning algorithms.

Kulkarni et al. [9] performed different feature extraction methods (Spectral features, wavelet transform features, complexity features) from EEG signals to diagnose Alzheimer's disease in their study. Classifying the obtained features with Support Vector Machines (SVM) classifier achieved a performance of 86% for spectral features, 88% for wavelet transform features, and 96% for complexity features, respectively. In a similar study, Bairagi [1] obtained spectral and wavelet transform-based features from EEG signals.

These features were classified with two different classifiers (SVM and KNN). Based on spectral-based features, they classified the features using the SVM algorithm with a classification performance of 94%. Ruiz-Gómez et al. [10] combined the spectral and nonlinear features obtained from EEG signals for Alzheimer's detection and classified the features they obtained with different machine learning algorithms. In their study, Alzheimer's and healthy controls achieved the highest rate of 78.43% with Multilayer Perceptron (MLP) and Quadratic Discriminant Analysis (QDA) algorithms. Amezquita-Sánchez et al. [11] applied in their studies multiple signal classification and empirical wavelet transform (MUSIC-EWT) methods to EEG signals to diagnose Alzheimer's disease. Then, Hurst exponent measures and fractal dimension properties were calculated. This study achieved 90.3% accuracy by classifying these features using advanced probabilistic neural networks (EPNNs). In a study by Kulkarni [12], they combined the spectral and complex features obtained from the EEG signal. With a maximum accuracy of 94%, these traits were then identified using the KNN classifier to identify Alzheimer's disease. Tzimourta et al. [13] investigated the statistical and spectral features with several classification techniques. With the Random Forest algorithm, 88.79% of the classifications were accurate. In their research, Buyrukolu proposed a method to classify Normal,

Mild Cognitive Impairment, and Alzheimer's conditions, it yields an accuracy of 91% [14]. In another study by Buyrukolu [15], they classified Alzheimer's conditions, Mild Cognitive Impairment, and Normal successfully with AdaBoost classifier. H. Yu et al. [16] proposed a network-based Takagi-Sugeno-Kang (N-TSK) method for AD identification which employs the complex network theory and TSK fuzzy system. The highest accuracy gives 97.3% for patients with closed eyes and 94.78% with open eyes. In 2023, Digambar V. Puri et al. [17] performed low-complexity orthogonal wavelet filter banks with vanishing moments (LCOWFBs-v) to decompose the AD and normal controlled (NC) EEG signals into sub-bands (SBs). The two features, Higuchi's fractal dimension (HFD) and Katz's fractal dimension (KFD) were extracted from EEG SBs. It yields an accuracy of 98.5% and 98.6% using the LCOWFBs-4 and LCOWFBs-6, respectively with a cubic-support vector machine classifier and 10-fold cross-validation technique. In a study by H. Yu et al. [18], they analyzed the EEG power change using Power Spectral Density during an acupuncture process. It was found that EEG power significantly increased in the delta and alpha bands under acupuncture and high-power level remained in alpha band after the acupuncture state. The result shows the functional networks in delta and alpha bands are small world networks (SWN) and acupuncture improves the SWN efficiency of functional network. H. Yu et al. [19] proposed a classification framework for different acupuncture manipulations, which employed the graph theory and machine learning method. The highest accuracy was achieved 92.14% with support vector machine. By further optimizing the network features utilized in machine learning classifiers, it was found that the combination of node betweenness and small-world network index is the most effective factor for acupuncture manipulation classification. In 2021, Pini L et al. [20] studied the clinical characterization of AD. Brain functional networks were used to provide a macro-scale scaffolding and explain the heterogeneity. TAR DNA binding protein 43 related pathologies (proteinopathies that mimic AD) were distinguished by network abnormalities.

Alzheimer's disease is difficult to diagnose. According to the report published in 2021, it is stated that 30% of people are misdiagnosed. Another difficulty in diagnosing Alzheimer's disease is that it is costly. Even if tests are carried out, especially in low-income countries, the lack of sufficient experts to examine them causes serious problems [21].

Due to these negativities, this study tends to investigate the availability of constructing an accurate and effective system that can detect Alzheimer's disease from central lobe EEG channels (only three electrodes) using the power of machine and deep learning. Moreover, the results of the proposed classification implemented on different datasets will demonstrate the capability of such approach for integration with fully automated disease diagnosis systems.

2. Materials & Methods

This section presents the definition of the EEG data set, the

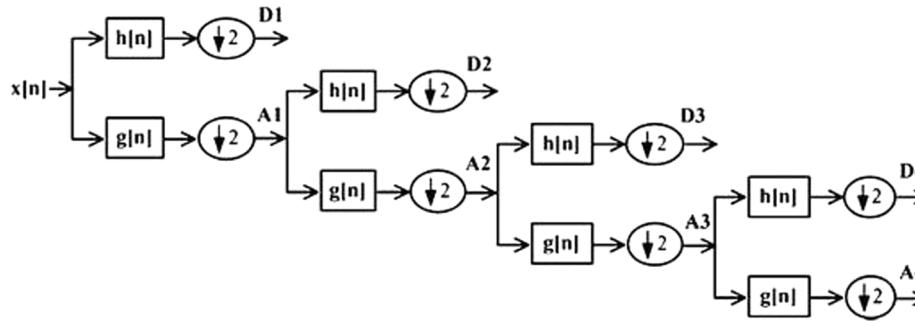


Fig. 2. Wavelet transform decomposition process [26].

preprocessing steps, and the steps of the presented analysis method. Fig. 1 presents the flow diagram of the proposed method for automatic AD detection from EEG signals. This study includes the following steps: (i) Pre-processing the EEG signal to remove the noise using Multi-Scale Principal Component Analysis (MSPCA). (ii) Application of Wavelet Transform method of the data received from each EEG channel. (iii) Calculation of statistical properties. (iv) Testing the accuracy of the proposed method by classifying it with the seven algorithms. (vi) Testing the classification performance in distinguishing Alzheimer's disease. All experimental stages of the study were carried out using MATLAB software.

2.1. Data sets

In this research, two different datasets were presented to evaluate the generalization ability. The first dataset was recorded by Florida State University researchers with a 19-electrode recorder using Biologic Systems Brain Atlas III Plus workstation according to the international 10–20 system. Recordings containing four groups (A, B, C, D) were performed under two resting conditions. The eyes of groups A and C were visually fixed and recorded with their eyes open. Groups B and D were recorded with their eyes closed. Groups A and B consisted of twenty-four healthy elderly subjects (age range of 61–83 years and mean age of 72 years) and subjects who did not have any neurological disorders. Groups C and D consisted of 24 probable AD patients (ages 53–85 years and mean age 69 years) diagnosed by the National Institute of Neurological and Communication Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Associated Disorders Association (ADRDA). EEG recordings are limited to the frequency range of 1–30 Hz and consist of 8-second recordings. The sampling frequency of the recordings was taken at 128 Hz. A technician was assigned during the recording to control the patient's alertness. This EEG dataset is available as open source [22].

The second dataset was posted by Matouš Cejnek and acquired with a 19-electrode recorder according to the international 10–20 system. Three groups of recordings (MCI, AD, NC) were performed. Group MCI was recorded from mild cognitive impairment subjects. Groups AD and NC were recorded from Alzheimer's and normal subjects, respectively. The dataset contains 7 MCI, 59 AD, and 102 NC subjects. For the purpose of logical comparison, only the first 8-second recordings from the first 24 AD and 24 NC subjects were selected for further implementation. The sampling frequency of the NC recordings was taken at 128 Hz. While the sampling frequency of the AD recordings was taken at 256 Hz. Those recordings were first down-sampled to 128 Hz using the decimation function found in the MATLAB signal processing toolbox. Decimation reduces the original sampling rate of a sequence to a lower rate. This EEG dataset is freely available at [23].

Principal Component Analysis (PCA) and Wavelet Transform are two analysis techniques that are combined in MSPCA. MSPCA is preferred because of its multi-scale structure for analyzing the changing signal or removing noise [24]. During EEG recording, noises may occur due to

interaction with muscles, eye movement, power lines, and other devices. For the purpose of eliminating noise in the EEG data obtained from each electrode, the MSPCA method was used in the proposed study. While applying the MSPCA method, the parameters were determined by taking the wavelet transform function 'sym4' and level 4.

2.2. Data epoching

For both datasets, there are 24 subjects, and each subject has (128 points [one second] X 19 channels) as one signal trial. A target is to identify whether each EEG segment belongs to a normal or a dementia state based on the given test set.

The epoching step is constructed by sequentially calculating the 19 channels of all eight-second segments. It has been previously discussed that the training set is 76 segments (40% of 192); that is, 76 segments. However, the subject's test set consists of 116 segments (the remaining 60%).

A Signal Processing Toolbox component associated with MATLAB version 9.13.0.2049777 (R2022b) was used in this study. Since it is a high-performance and robust tool, it can perform a wide range of signal-processing functions for data analysis. This implementation was conducted using a laptop with an i7-2.7 GHz processor Core (TM).

2.3. Wavelet transform

In biomedical signal analysis, wavelet transforms are used widely [4,25], especially for EEG signal analysis due to their non-stationary nature. Time-frequency analysis is possible with the wavelet transform by using broad time windows at low frequencies and small-time windows at high frequencies. Signals are decomposed using wavelet transforms by using sequential low-pass and high-pass filters and two down-samplers. A discrete main wavelet $g(n)$, whose mirror $h(n)$ is a low pass filter, can be represented by the high pass filter $g(n)$. As a result of the first high-pass and low-pass filters, A_1 and D_1 stand for approximation and verbose coefficients, respectively. As shown in Fig. 2, A_1 is fragmented further until the required number of dissociation levels is achieved [25,26].

A low-pass filter determines the broadening function $\varphi_{j,k}(n)$, and a high-pass filter determines the wavelet function $\psi_{j,k}(n)$.

$$\varphi_{j,k}(n) = 2^{\frac{j}{2}}h(2^j n - k) \quad (1)$$

$$\psi_{j,k}(n) = 2^{\frac{j}{2}}g(2^j n - k) \quad (2)$$

There are four parameters to consider: $j = 0, 1, 2, \dots, J-1$; $n = 0, 1, 2, \dots, M-1$; $k = 0, 1, 2, \dots, 2^j - 1$; $J = \log_2(M)$; and M is the length of the signal [27].

Using the fundamental frequency components in the signal, the maximum separation level can be calculated. In the Wavelet Transform, the coefficients are called the dot products of the original time series, while the coefficients A_i (Equation 3) and i are approximated by the determined basis functions. At the level, the details of the coefficients

I.A. Fouad and F. El-Zahraa M. Labib

Biomedical Signal Processing and Control 86 (2023) 105266

are described as D i (Equation 4):

$$Ai = \frac{1}{\sqrt{M}} \sum_n x(n) \times \varphi j, k(n) \quad (3)$$

$$Di = \frac{1}{\sqrt{M}} \sum_n x(n) \times \psi j, k(n) \quad (4)$$

Here k = 0, 1, 2, ..., 2j-1, and M is the length of the EEG time series at discrete points [4]. In the presented study, the wavelet transform process is performed by applying Daubechies' ('db4') main wavelet function to EEG signals.

2.4. Calculation of features

As a result of applying the Discrete Wavelet Transform (DWT) and decomposing the EEG signal to level 3 ('db4'), the following thirteen features were calculated:

Note: Treating the signal feature vector {X(1,j)} as a 1-D image, the following equations will be calculated.

- Contrast:

Contrast is the difference between the amplitude values of an instant signal and its neighbor over the entire time series. Calculating contrast is as follows:

$$Cont. = \sum_j [1 - j]^2 X(1, j) \quad (5)$$

where j is the column number in the signal row vector, and X is the EEG amplitude value.

- Homogeneity:

Homogeneity of the row vector refers to the uniformity of distribution elements of the co-occurrence matrix to its diagonals. Equation 6 can be used to calculate this:

$$Hom. = \sum_j \frac{X(1, j)}{1 + |1 - j|} \quad (6)$$

where j and X as mentioned above.

- Energy:

It is also known as the uniformity of the row vector, which is the sum of the squared elements of the co-occurrence matrix. A range of 0 to 1 is considered to be the range of energy. Here is how Equation 7 calculates energy:

$$Ener. = \sum_j X(1, j)^2 \quad (7)$$

- Correlation:

The correlation is the linear relationship between two-instantaneous amplitude values considering the entire row vector. This number ranges from -1 to 1. The following equation shows how correlation is calculated:

$$Corr. = \sum_j \frac{(i - \mu_1)(j - \mu_2)X(1, j)}{\sigma_1 \sigma_2} \quad (8)$$

where μ_1 , μ_2 , σ_1 , and σ_2 are the means and standard deviations of X_1 and X_2 which represent the partial probability density function.

- Root Mean Square:

Based on the row vector, the RMS gives the arithmetic mean of the squares of the mean values. As shown in Equation 9, RMS is calculated as follows:

$$RMS = \sqrt{\frac{\sum_{i=1}^M |x_i|^2}{M}} \quad (9)$$

- Standard Deviation:

SD refers to the mean square variation of instantaneous amplitude value $X(1, j)$ from its mean value. In normal mathematics, SD is equal to the square root of the variance. Equation 10 determines it as follows:

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^N (X(1, j) - \mu)^2} \quad (10)$$

- Entropy:

An entropy value can describe the textural features of a row vector. The formula for entropy can be found in equation 11.

$$Ent. = - \sum_{k=0}^{L-1} Pr_k (\log_2 Pr_k) \quad (11)$$

where Pr_k (probability of the kth amplitude value) = $Z_k / m * n$ (Z_k refers to the number of points with the kth amplitude level).

- Mean:

It is the central tendency of a probability distribution in a row vector of values. Calculation of mean value is shown in equation 12:

$$Mean = \frac{1}{N} \sum_{j=1}^N P(1, j) \quad (12)$$

- Skewness:

Skewness is defined as the degree of asymmetry in amplitude distributions around their means. Equation 13 can be used for calculating the skewness.

$$Skew. = \frac{1}{N} \sum_{j=1}^N \left(\frac{P(1, j) - \mu}{\sigma} \right)^3 \quad (13)$$

- Smoothness:

A relative smoothness measure is used to measure amplitude value contrast. As a result of Equation 14, the smoothness can be determined as follows:

$$Smth. = 1 - \frac{1}{1 + \sigma^2} \quad (14)$$

- Kurtosis:

Kurtosis (K) of a row vector can be calculated by comparing its peak or flatness to the normal distribution. In Equation 15, kurtosis is defined conventionally as:

$$Kurt. = \left\{ \frac{1}{N} \sum_{j=1}^N \left(\frac{P(1, j) - \mu}{\sigma} \right)^4 - 3 \right\} \quad (15)$$

- Variance:

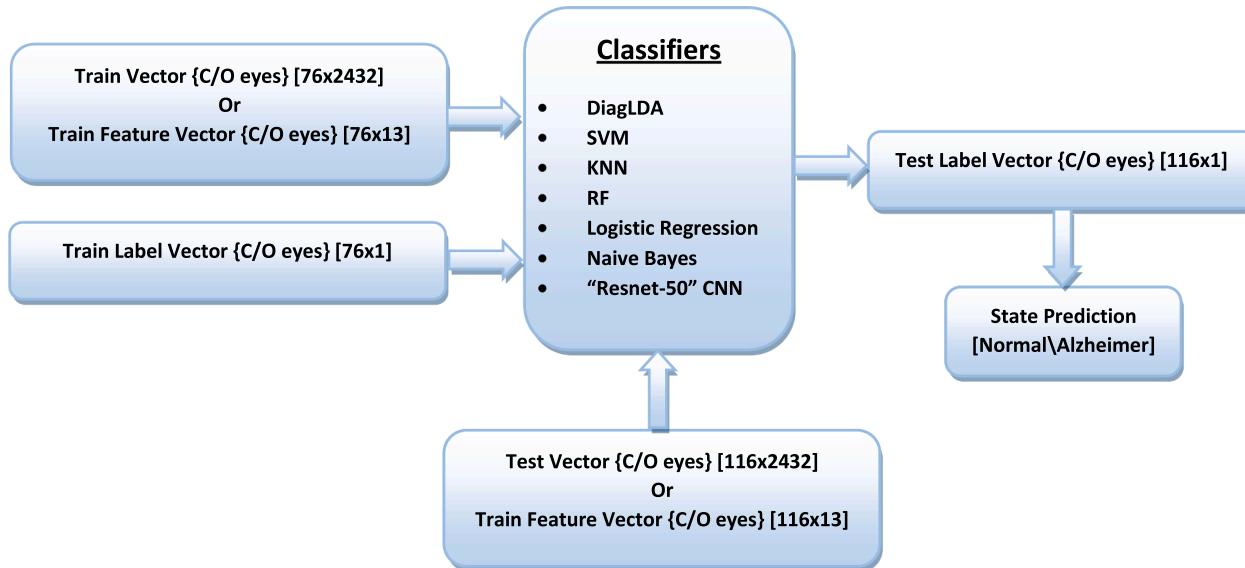


Fig. 3. General classification procedure for different classifiers.

Table 1
Hyper-parameters values for support vector machine.

Standardize	Kernel Function	Kernel Scale	SVM Procedure
True	Linear	23.4	Procedure I
True	RBF	3.3	Procedure II

Variance is computed using equation 16 as is square root of SD.

$$\text{Var} = \sqrt{SD} \quad (16)$$

• Inverse Difference Movement:

The texture of the row vector is represented by IDM. There is a range of 0 to 1, where 0.0 indicates a vector with a high level of texturing, and 1.0 indicates an untextured vector. The following equation describes IDM:

$$IDM = \sum_{j=1}^N \frac{P(1,j)}{1 + |1 - j|} \quad (17)$$

2.5. Classification

Fig. 3 depicts the general classification procedure for different classifiers. A detailed explanation of the proposed classifiers will be presented in the following sections.

1. Diagonal discriminant analysis

It is an algorithm for distinguishing groups of people using linear functions that is called a linear classifier. They are among the most popular Brain-Computer Interface (BCI) algorithms.

Data representing different classes can be separated by Linear discriminant Analysis (LDA) by utilizing hyperplanes [28,29]. According to the hyperplane, a function vector determines its class in a two-class problem.

Assume that the data have a normal distribution, and that the diagonal covariance matrix of each group is the same. This is known as diagonal linear discriminant analysis (DiagLDA). The distance between classes and the variance between classes must both be maximized to determine a separating hyperplane [29].

In BCI systems, this technique is well-suited due to its low computational requirements. Additionally, it produces high performance and is easy to use.

An individual (or sample) who was being tested (Normal) was given a non-target label of “-1” while a patient with Alzheimer’s was given a target label of “1” (DiagLDA method).

2. Support vector Machine (SVM)

In addition to dividing the data into two groups, this classifier uses a hyperplane to categorize a set of binary-labeled data [30,31]. Afterward, the classifier is trained on a specific dataset, and the hyperplane is optimized depending on its maximum gap with the dataset. By moving the data from the input area to the feature area, linear classification is achieved.

Classification can be performed using linear decision boundaries with a linear support vector machine. Many concurrent BCI problems have been solved using these classifiers [32,33,34,35,36,37].

There are several advantages to using SVMs. As a result of their strong generalization properties, SVMs are considered to have insensitivity to overtraining [38], and the curse of dimensionality is believed not to affect them [30,31].

Researchers have used SVMs to study the brain because they are an effective method for recognizing patterns in high-dimensional problems [35]. It is shown in Fig. 3 what the proposed algorithm looks like.

In this paper, two procedures related to SVM are discussed: “Radial Basis Function” and “Linear Function”.

Hyperparameters are parameters that can be adjusted in most machine learning and deep learning algorithms. Before training the models, it is important to set hyperparameters. A robust and accurate model relies heavily on hyperparameters. In this way, they prevent the model from overfitting or underfitting by finding a balance between bias and variance.

As shown in Table 1, a subset of the train data was used as a validation set for the choice of regularization parameters (hyperparameters).

3. 3. K-nearest neighbors (KNN)

In addition to dividing the data into two groups, this classifier uses a hyperplane to categorize a set of binary-labeled data [30,31]. Afterward, the classifier is trained on a specific dataset, and the hyperplane is optimized depending on its maximum gap with the dataset. By moving the data from the input area to the feature area, linear classification is achieved.

Based on their nearest neighbors’ distances (KNN), samples can be sorted using the K-nearest neighbor learning algorithm [39]. This

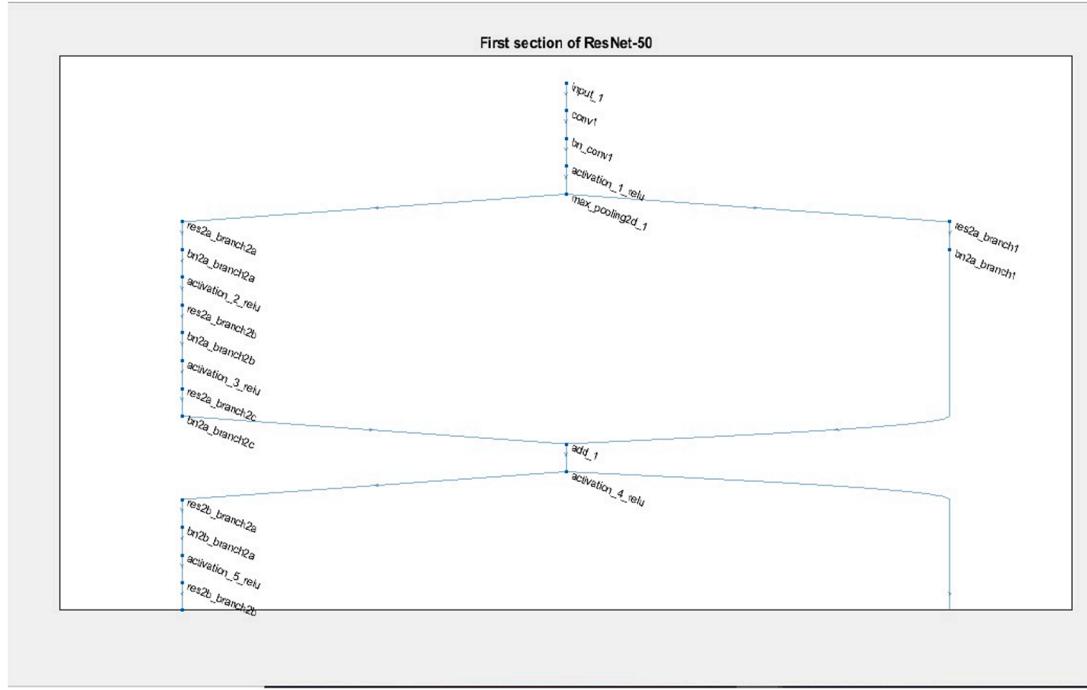


Fig. 4. “ResNet-50” model (construction of the first section only).

Table 2

Parameters' values of tried “Resnet-50” model.

Values	Model's Parameters
'crossentropyex'	LossFunction
Linear	Learning
32	MiniBatchSize
0.01	Learning Rate
fc1000	FeatureLayer

Table 3

First dataset “All participants-closed eyes” proposed classifiers’ accuracies.

Classifiers	All Participants Closed Eyes' EEG Data	
	Cut-off Frequency Range	Accuracy
DiagLDA		93.97%
KNN (N = 5)		88.79%
SVM (Linear)		96.55%
RF (T = 10)	1—30 Hz	93.97%
SVM (RBF)		85.34%
Logistic Regression		75.86%
Naive Bayes		95.69%

algorithm identifies new samples by using previously labeled training samples and their features. It is a memory-based algorithm that doesn't require fitting a model. A query point x_0 is used to find the k closest training points (Euclidean distances) to it. Assigning a new query to a cluster involves calculating the number of its neighbors. A random number will be used to break any voting bonds [40].

A KNN (with $N = 5$) method was used in the present study as a classifier for brain research, as shown in Fig. 3.

4. Random forest classifier

As a machine learning algorithm, random forests are derived from decision trees. Decision trees are diagrams in the form of trees that are used to make decisions. There are different branches on the tree representing different options, events, or reactions. [41].

Morgan and Sonquist [42] introduced the tree-based approach.

Table 4

First dataset “all participants-closed eyes” – processing time of the suggested classifiers.

Classifiers	All Participants	
	Training Time (sec.) / Trial	Testing Time (sec.) / Trial
DiagLDA	0.281545	0.017286
KNN	0.481505	0.040824
N = 5		
SVM Linear	0.167519	0.025668
RF	0.980015	0.081460
T = 10		
SVM RBF	0.320598	0.033832
Logistic Regression	0.781005	0.028388
Naive Bayes	0.889065	0.016013

Breiman et al. [43] developed classification and regression trees (CART) by extending the tree-based approach and strengthening the estimation mechanism. While Breiman [44] developed the decision tree approach used in classification and regression problems and proved that more than one tree can make more accurate decisions together with the method named bagging. The bagging method is based on constructing a set of trees in the classifiers randomly selected from the data set. Ho [45] developed the randomly selected sub-datasets with the random subspaces' method and each tree with the properties of the space in which it is included. The random subspace method, unlike the bagging method, uses a different number of classifier properties for each tree when creating trees.

The Random Forest algorithm described by Breiman [46], while determining the best split for the standard tree node, by adapting the rule of choosing the best estimator in the whole data set with the theory of randomness. The best split for each tree node is a random feature to be selected from the sub-dataset prepared for preloading. After the formation of many trees, the forest has become suitable for classification and each vote individually for the class that complies with its own rules. The Random Forest algorithm classifies the classification problems

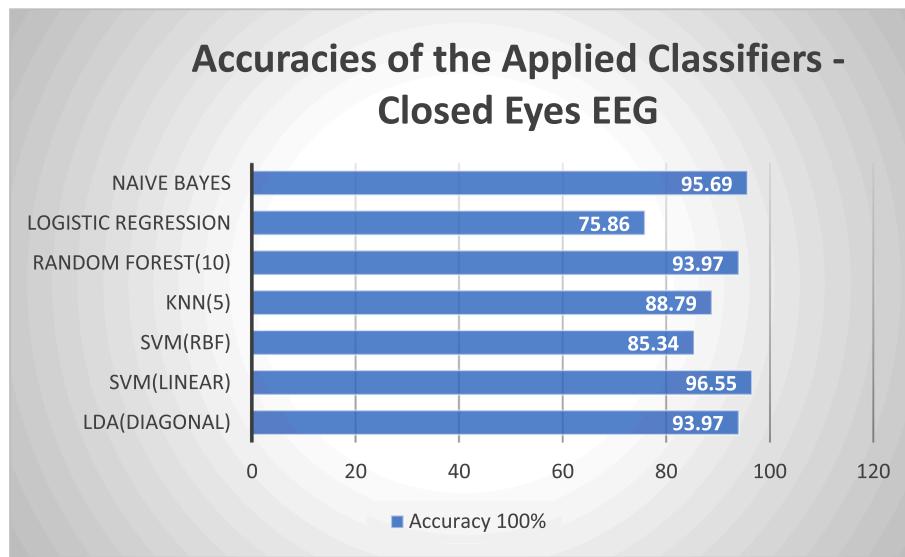


Fig. 5. First Dataset Proposed Classifiers' Accuracies on EEG of the Closed Eyes Participants.

Table 5

First dataset "all participants-opened eyes" proposed classifiers' accuracies.

Classifiers	All Participants	Accuracy
	Opened Eyes' EEG Data	
Cut-off Frequency Range		
DiagLDA	1—30 Hz	91.38%
KNN		76.72%
N = 5		
SVM Linear		87.93%
RF		93.1%
T = 10		
SVM		72.41%
RBF		
Logistic Regression		65.14%
Naive Bayes		85.73%

according to the majority right principle. That is, the class with the most votes indicate the prediction of the Random Forest.

In regression problems, the estimation of the Random Forest is found by taking the average of all the trees in the forest. According to the large

number theorem of probability theory, when the same experiment is repeated many times, the result obtained converges to the expected value (Dekking, 2005) [47]. Based on this theorem, it can be interpreted that large numbers are never overlearned for the Random Forest algorithm. Although one of the other basic features of the Random Forest algorithm is a collection of decision trees, no individual tree is pruned.

Table 6

Second dataset proposed classifiers' accuracies.

Classifiers	All Participants	Accuracy	Processing Time (Sec.)
	Closed Eyes' EEG Data		
Cut-off Frequency Range			
DiagLDA	1—30 Hz	93.97%	1.234509
KNN (N = 5)		93.10%	1.344245
SVM (Linear)		96.55%	1.298201
RF (T = 10)		93.10%	1.394024
SVM (RBF)		85.34%	1.31217
Logistic Regression		80.92%	0.741806
Naive Bayes		94.52%	0.724248

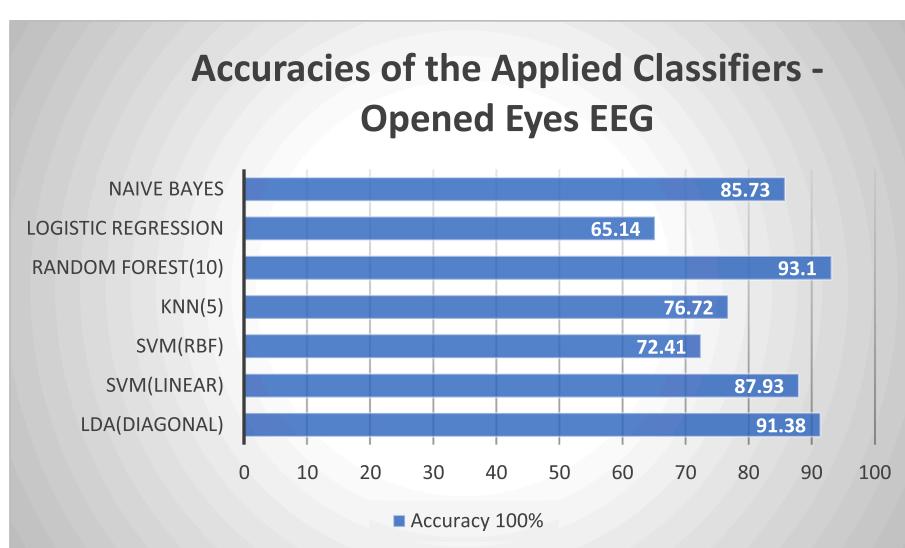


Fig. 6. First Dataset Proposed Classifiers' Accuracies on EEG of the Opened Eyes Participants.

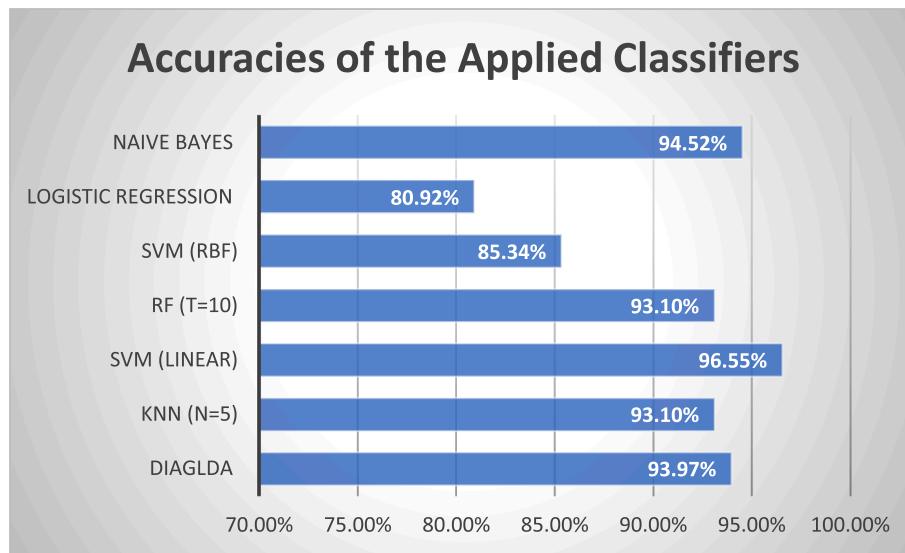


Fig. 7. Second Dataset Proposed Classifiers' Accuracies on EEG of All Participants.

Table 7

"Different EEG bands-closed eyes" Accuracies of the applied classifiers.

Classifier	LDA D	SVM L	SVM RBF	KNN	RF	LR	NB	LDA D	SVM L	SVM RBF	KNN	RF	LR	NB
EEG Band Accuracy	DELTA [1:3 Hz] 91.38%	90.52%	83.62%	83.62%	91.38%	82.76%	92.24%	THETA [4:7 Hz] 92.24%	94.83%	87.07%	81.03%	94.83%	81.03%	93.97%
EEG Band Accuracy	ALPHA [8:12 Hz] 94.83%	92.24%	78.45%	81.03%	92.24%	75.00%	95.69%	BETA [13:30 Hz] 97.41%	97.41%	84.48%	85.34%	94.83%	84.48%	98.28%

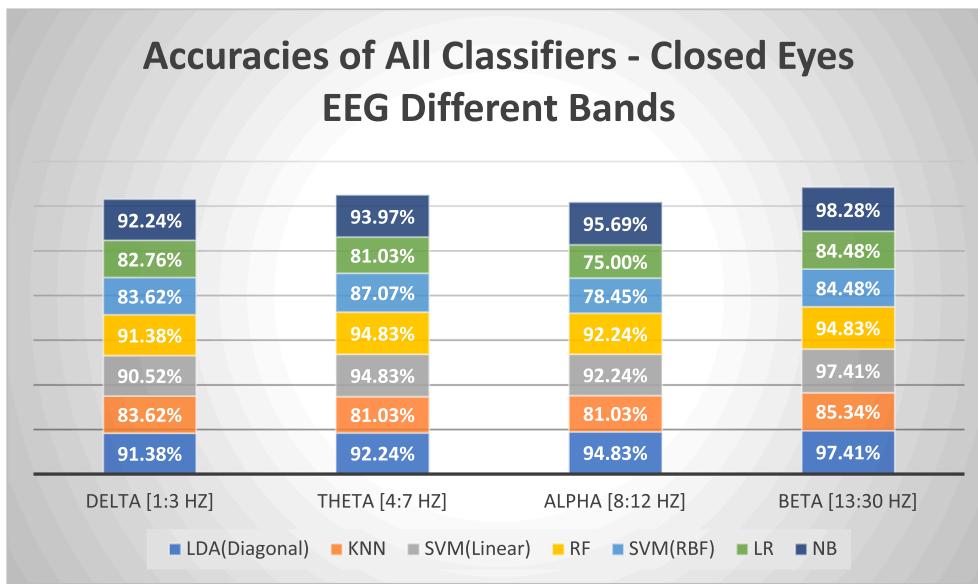


Fig. 8. Proposed Classifiers' Accuracies on Different EEG Bands of the Closed Eyes Participants.

This integrated tree system, which is built from randomly selected classifier features and sub-datasets, does not need feature extraction as it is needed in other algorithms. The points where hyperparameter optimization of this algorithm can be made can be considered as the number of decision trees, the minimum, and maximum of the branches of the trees, the separation criterion to determine the splits at the nodes, the minimum and maximum of the number of classifier features to be used,

and the purity value constants to be used in the formation of splits and branches if an entropy-based tree structure is to be constructed.

Using Random Forests has the following benefits:

1. It takes less time and less effort to practice using multiple trees because overfitting is avoided.

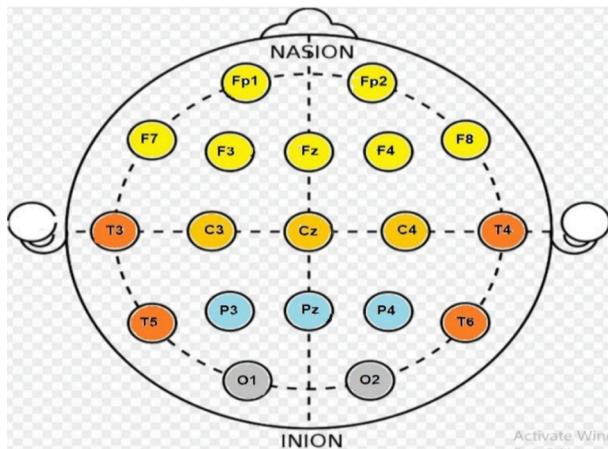


Fig. 9. Different Inspected EEG Electrodes According to 10–20 System. Every group is Colored differently.

2. For large datasets, it makes highly accurate predictions and performs well in large databases. The Random Forest is likely to excel in this area since big data is essential today.
3. If a large percentage of the data is missing, Random Forest decides which parameter is best for missing data.

An algorithm based on Random Forest with ten trees is implemented in the present study.

5. Logistic regression classifier

It is a popular classification model based on the logistic function [48]. A logistic function is an s-shaped curve that can take any input x value and map it to a value between 0 and 1.

Classification is made by assuming a linear relationship between dependent or independent variables. Therefore, logistic regression is considered a very inflexible classification algorithm that will oversimplify the real problem and result in a high bias [48].

Logistic Regression was introduced by Raymond Pearl and Lowell Reed in 1940 [49]. Logistic regression is used to classify independent variables to examine the probability of a categorical outcome [50,51]. Equation 5 shows the logistic regression equation.

$$W(t) = \Omega \frac{\exp(\alpha + \beta t)}{1 + \exp(\alpha + \beta t)} \quad (5)$$

In Equation 5, the expression Ω indicates the upper limit of the saturation level of W , the expression α the value of the curve on the x-axis, and the expression β the slope of the curve. The regression coefficients are calculated using this method by determining the relationship between the independent variables in terms of probability [52]. Logistic regression algorithm is frequently used in fields such as health [53], social sciences [54], and political sciences [55].

6. Naive Bayes classifier

When features in each class are distinct from one another, the Naive Bayes classifier is intended to be utilized. However, even when the independence assumption is incorrect, it appears to function effectively in practice. The data is classified by the Naive Bayes Classifier in two steps:

1. Train: Assuming that the characteristics are independent, the probability distribution's parameters are calculated using training examples.

2. Test: For any test sample, the method calculates the posterior probability of that sample belonging to each class.

Naive Bayes Classifier is a simple probabilistic classifier based on Bayes rule in Equations 6 and 7.

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)} \quad (6)$$

Table 8
"Different EEG electrodes {closed eyes-beta band}" Accuracies of the applied classifiers.

ELECTRODES CATEGORY	FRONTAL [Fp1, Fp2, Fz, F3, F4, F7, F8]				CENTRAL [C3, Cz, C4]				SVM RBF				NB
	LDA D	SVM L	SVM RBF	LR	NB	LDA D	SVM L	RF	LR	KNN	RBF		
ELECTRODES CATEGORY	FRONTAL [Fp1, Fp2, Fz, F3, F4, F7, F8]	81.90%	86.21%	64.66%	67.24%	82.76%	56.90%	87.07%	95.69%	89.66%	93.10%	87.93%	95.69%
ELECTRODES CATEGORY	TEMPORAL [T3, T4, T5, T6]	93.10%	88.79%	81.03%	87.93%	72.41%	92.24%	91.38%	91.38%	91.38%	91.38%	79.31%	89.66%
ELECTRODES CATEGORY	OCCIPITAL [O1, O2]	94.83%	95.69%	67.24%	89.66%	91.38%	87.07%	95.69%					
Accuracy													

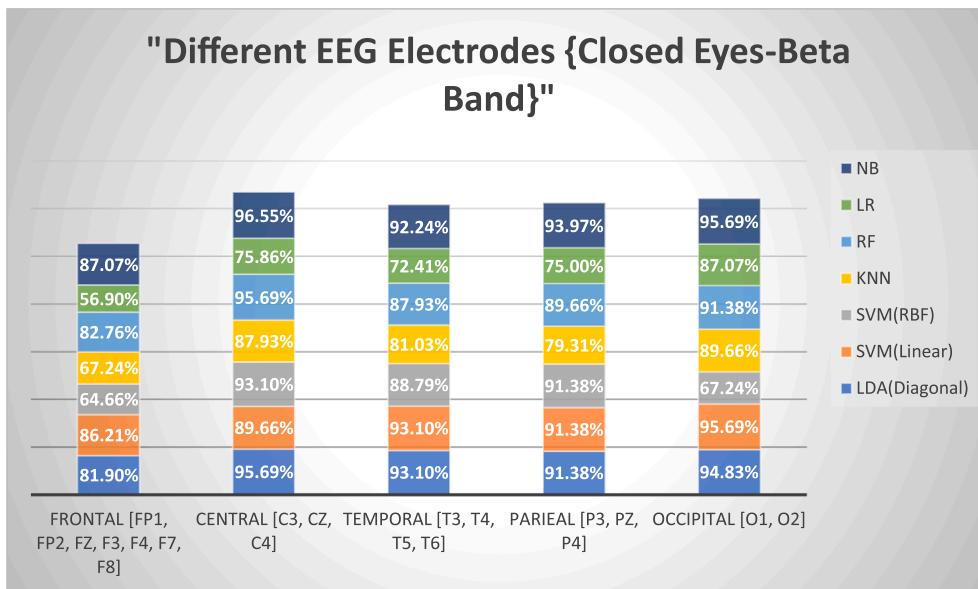


Fig. 10. Proposed Classifiers' Accuracies on Different EEG Bands of the Closed Eyes Participants.

$$P(\text{clx}) = P(x_1 \text{ lc}) \times P(x_2 \text{ lc}) \times P(x_3 \text{ lc}) \times \dots \times P(x_n \text{ lc}) \times P(c) \quad (7)$$

Using the Naive Bayes algorithm, the parameters required for classification are calculated using only the variables' standard deviation and mean [56]. According to equation 8, the Naive Bayes classifier selects the more likely classification according to its decision rule.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

7. ResNet-50 deep learning model

After generating the features from the EEG signals, the next step is to create the 2-D plots of the obtained features. Then, tools from the MATLAB deep learning toolbox were used. "Resnet-50" model was constructed. Training of the model was done on 1200 plots (3000*0.40%) and the remaining 1800 plots were used to test and validate the model.

Deep learning algorithms today are based on a computer simulation of the neurons in the human brain.

As a result of trying to accomplish it, a process has developed. Inspection of Donald Hebb's nerve cell structure led to the discovery of his neural networks. Afterward, design the model numerically in a computer environment.

Artificial nerves are created by creating artificial neural networks [57]. In recent years, deep learning has been used to describe the emerging and progressing nature of networks.

It has become increasingly important to use deep learning for classification, recognition, and detection in recent years [58]. Our daily lives are influenced by deep learning algorithms, which are commonly utilized in a wide range of technological structures.

To achieve high performance, this issue is more likely to arise. There is a wide range of classifications that have made recognition studies available.

The network is trained with several different learning strategies in deep learning architectures. Learning through teachers, learning without teachers, and learning through reinforcement are the three types of learning.

It is given what the input data at the input of the network will be as the output of the network in teacher learning. There is no difference in the weights of the network. To adjust the weights of the network, it is repeatedly inserted into the network as input data. As a result, weights are determined by training the network [59].

There is no knowledge of the input data at the output of teacher-less learning. Whenever data enters the network, it has a counterpart at the

exit that belongs to a nearby cluster. In this manner, each data item becomes part of a cluster at the output [60,61].

On the other hand, reinforcement learning does not provide a mechanism for determining what the input will be at the output data. There is a statement at the end that indicates whether or not it is true. When true, it sets a value of 1 and if false, it sets a value of 0. The network is trained in this manner [62].

In contrast to convolutional neural networks, deep learning represents an advanced structure. Data features are automatically extracted from input during the convolution process and added to the next layer.

Data is transferred between layers of a convolutional neural network using separate processes and various functions are carried out by each layer. The layers in a deep learning architecture are as follows, along with the operations they perform:

- Convolution layer: Conditional operations on image matrices of $3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11$

The size matrix is traversed in order to form it. It is the entire image that is specified in small-size matrices.

A new image matrix is generated when the tool waves over the matrix, highlighting its features [63].

- Oxford University's Visual Geometry Group developed the Vgg16 deep learning algorithm.

In this network, there are 16 convolutions and 3 fully connected layers, and it is created using a convolutional algorithm. Using the Maxpool algorithm, there are forty-seven layers with Full-connectedlayer, Softmaxlayer, Dropoutlayer, and Relulayer.

In contrast to traditional consecutive networks like AlexNet, VggNet has a different structure.

There is a structural difference between the Resnet micro module and other modules. Some layers are denied when swapping between them, and the lower layer may be preferred. It is important to allow this situation within the Resnet architecture so that the success rate will grow.

Fig. 4 shows the first section as the utilized network is large (50 sections).

Table 2 presents some values of the chosen parameters via training and testing of the data utilizing the previously referred model.

"ResNet-50" is one of the pretrained networks that can be

I.A. Fouad and F. El-Zahraa M. Labib

Biomedical Signal Processing and Control 86 (2023) 105266

Table 9

All applied classifiers' confusion matrices for the Closed Eyes' Data from All Central Electrodes.

Classifier	Observation Type	No. of Observations
DiagLDA	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 55
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 56
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 0
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 5
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 50
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 54
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 5
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 7
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
SVM(Linear)	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 7
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 54
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 7
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 7
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 7
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 0
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 5
SVM(RBF)	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 7
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 54
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 7
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 7
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 54
KNN	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 7
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 7
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 48
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 54
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 7
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 7
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 55
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 56
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 0
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 5
Random Forest	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 44
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 44
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 11
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 17
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 52
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 3
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 44
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 44
Logistic Regression	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 11
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 17
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 52
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 3
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 52
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 3
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
Naive Bayes	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 52
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 3
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 52
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60
	Positive Actual Value (Target) / Negative Predictive Value (Non-Target)	FN = 3
	Negative Actual Value (Non-Target) / Positive Predictive Value (Target)	FP = 1
	Positive Actual Value (Target) / Positive Predictive Value (Target)	TP = 52
	Negative Actual Value (Non-Target) / Negative Predictive Value (Non-Target)	TN = 60

Table 10

An analysis of the accuracy calculations for all classifiers used.

Classifiers	For "Central" Electrodes (C ₂ , C _z , and C ₃)			
	Accuracy	Precision	Sensitivity	Specificity
DiagLDA	95.6896%	91.6666%	100.000%	95.6521%
KNN (N = 5)	87.9310%	87.2727%	87.2727%	87.2727%
SVM (Linear)	89.6551%	87.7193%	90.9091%	89.2857%
Random Forest (T = 10)	95.6896%	91.6666%	100.000%	95.6522%
SVM (RBF)	93.1034%	97.9592%	87.2727%	92.3077%
Logistic Regression	75.8621%	72.1311%	80.0000%	75.8621%
Naive Bayes	96.5517%	98.1132%	94.5455%	96.2963%

implemented utilizing Neural Network Toolbox in MATLAB. It has been trained on the ImageNet dataset, which has 1000 object categories and 1.2 million training images [64]. Because this is a large network, Fig. 4 shows just the first section.

3.1. State prediction

Using the classifiers previously discussed, the state is almost specified. In the following section, the work will be analyzed from different perspectives.

4. Results & discussions

Two different datasets were presented. The accuracy of the results is calculated based on the percentage of trials in the test sets that were correctly estimated.

First: Dataset I

4.1. Inspecting the EEG signals of the closed eyes' participants

Each classifier's accuracy and results are displayed in Table 3 according to the evaluation standards. As shown in Table 4, predicting a trial state does not take more than three seconds. The calculated times are accurate and can be utilized for online applications.

According to Fig. 5, the accuracies of the discussed classifiers for closed-eye EEG data were obtained when all closed-eye participants were analyzed together.

Based on the results of the classifiers' accuracy comparisons, "Linear Support Vector Machines" and "Naive Bayes" give the greatest accuracy with 96.55% and 95.69% respectively. Logistic Regression was the least accurate method, achieving only 75.86% accuracy.

4.2. Inspecting the EEG signals of the opened eyes' participants

Open-eye participants had their EEG signals taken separately, and training and testing data are derived from the same criteria (40% for training and 60% for testing). Calculated accuracies are based on the correct estimation percentages in each set of test trials. Each classifier's accuracy is provided in Table 5, based on the evaluation standards.

Based on EEG data collected from subjects with opened eyes, Fig. 6 shows the accuracy of the proposed classifiers.

It is clear from the comparison of the accuracies of the suggested classifiers that the "Random Forest" and "Diagonal LDA" classifiers provided the highest accuracies with 93.1% and 91.38%, respectively. However, the "Logistic Regression" algorithm gave the lowest result as it achieved a maximum accuracy of 65.14% for the participants.

Generally, the "linear kernel functions" support vector machine classifier performs better than those with "RBF kernel functions". The reason for this is that EEG signals have a high dimensionality (a lot of features) [65], and it has been demonstrated in different EEG applications [66] and [67].

Second: Dataset II

Based on the percentage of the correctly predicted trials in the test, the classifiers' performance is evaluated. Table 6 and Fig. 7 show the second dataset classifiers' accuracies and processing time.

Based on Table 6 and Fig. 7, the results of the classifiers' accuracy comparisons show that "Linear Support Vector Machines" and "Naive Bayes" give the greatest accuracy with 96.55% and 94.52% respectively. It is also clear that the Logistic Regression was the least accurate method, achieving only 80.92% accuracy.

According to the previous analysis of the results, it is clear that:

- The proposed techniques are robust as applying them on two different datasets doesn't change the accuracies adversely, and the performance measurements show its reliability and effectiveness.

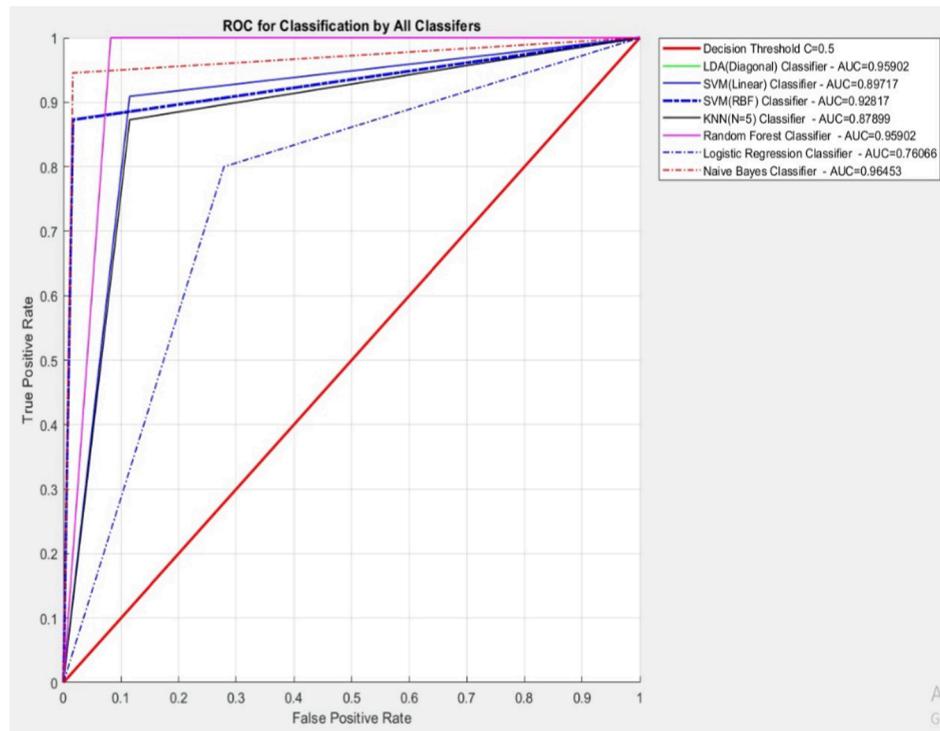


Fig. 11. "Receiver Operating Characteristic" Curves of the Proposed Classifiers.

Table 11

Confusion matrices for applied "Resnet-50" model regarding the data from all subjects.

		Actual Value	
		(Target) Positive	(Non-Target) Negative
Predicted Value	(Target) Positive	T.P. = 155	F.P. = 0
	(Non-Target) Negative	F.N. = 13	T.N. = 192

- Dealing with closed eyes EEG signals gives better results than dealing with open eyes EEG signals.

Consequently, the next steps of the suggested methodology will be

Table 12
Performances measurements for "Resnet-50" classifier.

Classifier	EEG from "Central" Electrodes (C2, CZ, and C3) for all subjects			
	Accuracy	Specificity	Sensitivity	Precision
"Resnet-50" Deep Learning Model	97.826 %	97.836 %	98.261 %	97.414 %

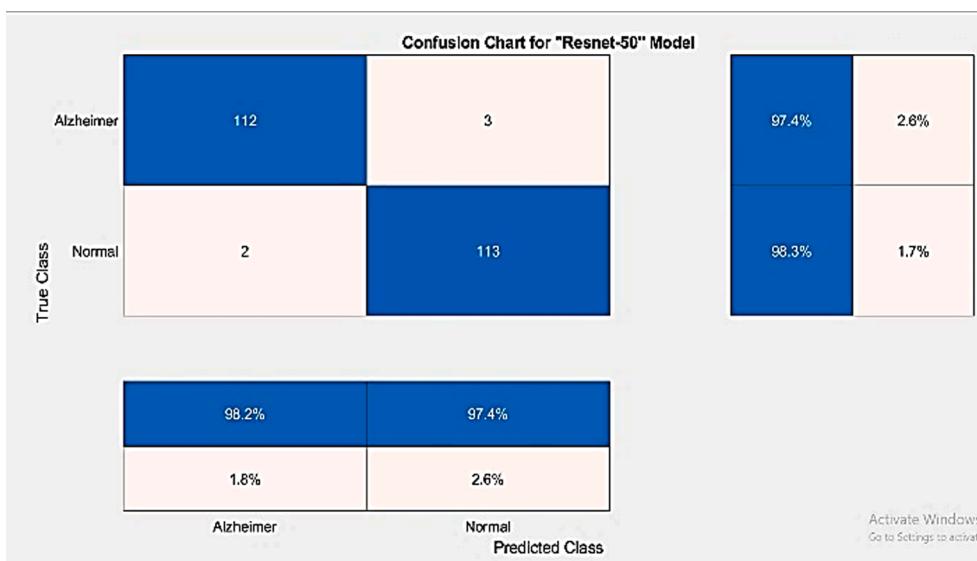


Fig. 12. Plot of confusion matrices for applied CNN model.

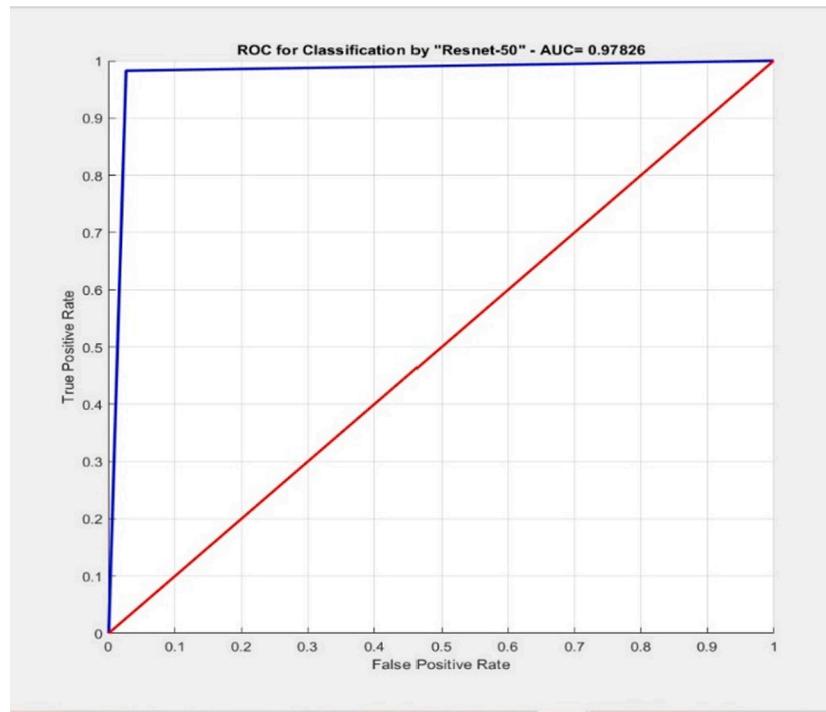


Fig. 13. "Receiver Operating Characteristic" Curve for applied CNN model.

Table 13
Accuracies from various researchers.

#	Research Group	Acc.	Classifier	EEG Dataset	Feature Inference
1	Kulkarni ve ark. [6]	96%	SVM	50 healthy + 50 Alzheimer's	Complexity properties
2	Ruiz-Gómez ve ark. [7]	78.43%	MLP	(37 healthy + 37 mild cognitive impairment) + 37 Alzheimer's	Spectral and nonlinear properties
3	Kulkarni [9]	94%	KNN	50 healthy + 50 Alzheimer's	Spectral and complexity properties
4	Amezquita-sánchez ve ark. [8]	90.3%	EPPN	37 mild cognitive impairment + 37 Alzheimer's	Hurst exponents and fractal dimension
5	Tzimourta ve ark. [10]	88.79	Random Forest	10 healthy + 8 mild Alzheimer's + 6 moderate Alzheimer's	Statistical and spectral properties
6	Bairagi [1]	94%	SVM	50 healthy + 50 Alzheimer's	Spectral properties
7	This work [Using EEG from only 3 electrodes]	97.83% 96.55% 95.69% 95.69%	ResNet-50 Naive Bayes DiagLDA Random Forest	Two different datasets Alzheimer's	Wavelet Transform and Statistical parameters

based on the first dataset EEG signals of participants with closed eyes.

Third: Inspecting the different EEG bands concerning the Closed Eyes' Participants

Taking EEG data separately from each band is now done for closed eyes' participants, and testing and training data are derived based on the same criteria (40% for Training, 60% for Testing). Accuracy is calculated based on the percentage of correctly estimated trials in test sets. As shown in Table 7 and Fig. 8, all subjects and each classifier show

accuracies based on the evaluation standards.

It is evident from Fig. 8 that the "EEG Beta Band" achieves the best performance among all the proposed classifiers, due to its highest average accuracy. Consequently, when trying to detect "Alzheimer" disease from EEG signals, the "Beta" band (which carry the high frequency range of EEG) is the most effective.

Fourth: Inspecting the different EEG electrodes

Again, constructing training and testing data using EEG of all closed eyes' participants simultaneously, 40% for Training and 60% for Testing. Now the investigation will concentrate on the different categories of EEG electrodes. As depicted in Fig. 9, there are 19 electrodes tested in this study according to 10-20 system. The electrodes are {Fp1, Fp2, Fz, F3, F4, F7, F8, Cz, C3, C4, T3, T4, Pz, P3, P4, T5, T6, O1, and O2}. They will be categorized into five different groups named [(F) Frontal, (C) Central, (P) Parietal, (O) Occipital, and (T) Temporal] as color codes presented in the figure.

The percentage of accurate estimated trials is used to calculate the number of accurate trials in the test sets. Table 8 and Fig. 10 illustrate the accuracies concerning all participants together and for each electrodes' group according to the evaluation standards.

Fig. 10 and Table 7 show that the central electrodes [C3, Cz, and C4] achieved the highest average accuracies for all seven proposed classifiers when evaluating the accuracies of the 19 electrodes using the mentioned classifiers. According to the following order, "detecting AD from EEG signals" is important and effective:

- Central Electrodes: The most effective electrodes.
- Occipital Electrodes.
- Temporal and Parietal Electrodes: They are almost the same.
- Frontal Electrodes: The least effective electrodes.

Because they are the most effective electrodes, the following part will focus on trying and analyzing the most effective types of electrodes previously mentioned.

In this study, the four possible observations are positive for a target (Alzheimer's Disease (AD)) and negative for a non-target (Normal):

- Positive observation, predicted positive observation: True positive

(T.P).

- Negative observation, predicted negative observation: True negative (T.N.).
- Negative observation, predicted positive: False positives (F.P.).
- Positive observation, predicted negative: False negatives (F.N.).

To evaluate the performance of each classifier, the accuracy, specificity, sensitivity, and precision will be computed for all classifiers.

A total of 116 observations were observed and predicted (192 observations [96 Normal and 96 AD] × 60% "The testing part").

Presented in [Table 9](#) are the confusion matrices for each of the seven classifiers.

Based on all participants' data, [Table 10](#) provides the accuracy calculations for seven classifiers using only three electrodes (C2, Cz, and C3).

Receiver Operating Characteristic (ROC) Curves are commonly used to represent the output of classification models. Here are the trade-offs between true positive and false positive rates for various likelihood thresholds in a predictive model. A comparison of ROC curves for EEG data obtained from electrodes C2, Cz, and C3 for all seven subjects is shown in [Fig. 11](#).

Taking all the accuracy calculations, [Table 9](#) and [Table 10](#) indicate that each of the four suggested classifiers gives high results (Naive Bayes, Random Forest, DiagLDA, and SVM(RBF)). AUCs (areas under the ROC curves) for the above-mentioned classifiers are high (more than 0.93) based on the analysis of ROC curves depicted in [Fig. 11](#). Consequently, the proposed models are appropriate.

Fifth: Inspecting the residual network model (ResNet-50)

The evaluation and analysis of the suggested deep neural network model will be presented in this section.

The confusion matrix of the presented model is depicted in [Table 11](#) and [Fig. 12](#).

[Table 12](#) shows the performance calculations of the suggested residual neural network model.

The most common visual representation of a predictive model's output is the "Receiver Operating Characteristic" (ROC) curve. A predictive model's true positive (TPR) and false positive rate (FPR) trade-off is summarized here for various likelihood thresholds. As shown in [Fig. 13](#), the ROC curve for the proposed model is plotted.

Comparative analysis

[Table 13](#) shows the relevant literature studies that used EEG signals to diagnose Alzheimer's. In all of the below-mentioned studies that perform machine learning-based Alzheimer's diagnosis, methods with high statistical features and high computational complexity have been proposed.

However, in the presented study, quite satisfactory classification performance was obtained by classifying only thirteen statistical features (Statistical Parameters) with more than one classifier.

5. Conclusion and future work

Positive developments in the health sector enable the world population to lead healthier and longer life. However, the increase in the number of unwanted diseases such as dementia causes undesirable results [[10,11,68,69,70,71,72](#)]. Although there is no definitive cure for Alzheimer's disease, early diagnosis of the disease can slow the process and improve quality of life [[73](#)]. Diagnosis of the disease requires many medical procedures such as neurological and psychiatric examinations, blood analysis, computed tomography, and the results obtained from these procedures should be evaluated comprehensively by the specialist [[71](#)].

As a result of all of these difficult diagnostic processes, it has been reported that 10–15% of cases fail to be diagnosed even by experienced specialists [[65,67](#)]. For all these reasons, it is important to develop a system using only biomedical signals or images. Biomedical imaging (such as CT, fMRI, PET) methods require high cost and time, as well as the lower cost and time requirement of EEG signals, which has been the

reason for preference in this study.

In this paper, two different datasets were presented. When applying several classification algorithms for both datasets, it was observed that machine learning and deep learning models achieved high accuracy reflecting the robustness and effectiveness of the suggested methods.

The most important limitation of the study, which can be compensated in future studies, is that the classification performance can be increased more by choosing different ensemble algorithms.

Wavelet transform is frequently utilized in the literature for time-frequency analysis of EEG signals due to its superior signal analysis capabilities. In this study, a simple system that uses wavelet transform and some statistical parameters together, which can distinguish EEG signals from Alzheimer's patients and healthy controls with high accuracy, is presented. With the data set used in the study, a machine learning-based study has not been presented yet. In this respect, the study cannot be compared with any other study. In addition to obtaining high classification performance with the suggested method, it also has a very high achievement in terms of computational complexity.

Declaration of Competing Interest

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

Data availability

I've shared the link of data.

Acknowledgment

The authors gratefully acknowledge the EEG Dataset providers:

- (Andriana S. L. O. Campanharo Fernando M. Ramos Aruane M. Pineda Luiz E. Betting 2020).
- Cejnek, Matouš (2017).

DECLARATIONS

Funding.

No funding was received for conducting this study.

References

- [1] V. Bairagi, EEG signal analysis for early diagnosis of Alzheimer's disease using spectral and wavelet-based features, *Int. J. Inf. Technol.* 10 (3) (2018) 403–412.
- [2] L.R. Trambaioli, N. Spolaòr, A.C. Lorena, R. Anghinah, J.R. Sato, Feature selection before EEG classification supports the diagnosis of Alzheimer's disease, *Clin. Neurophysiol.* 128 (10) (2017) 2058–2067.
- [3] R.H. Blank, *Alzheimer's Disease and Other Dementias: An Introduction*, in: *Social & Public Policy of Alzheimer's Disease in the United States*, Springer, 2019, pp. 1–26.
- [4] U. Orhan, M. Hekim, M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, *Expert Syst. Appl.* 38 (10) (2011) 13475–13481.
- [5] B. Hjorth, EEG analysis based on time domain properties, *Electroencephalogr. Clin. Neurophysiol.* 29 (3) (1970) 306–310.
- [6] S. Buyrukoglu, Y. Yilmaz, Z. Topalcengiz, Correlation value determined to increase *Salmonella* prediction success of deep neural network for agricultural waters, *Environ. Monit. Assessment* 194 (5) (2022) 1–12.
- [7] S. Buyrukoglu, S. Savaş, Stacked-based ensemble machine learning model for positioning footballer, *Arab. J. Sci. Eng.* 1–13 (2022).
- [8] R. Halde, Application of Machine Learning algorithms for betterment in education system, in: In 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), IEEE, 2016, pp. 1110–1114.
- [9] N.N. Kulkarni, V.K. Bairagi, Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier, *IEEE J. Res.* 63 (1) (2017) 11–22.
- [10] S.J. Ruiz-Gómez, et al., Automated multiclass classification of spontaneous EEG activity in Alzheimer's disease and mild cognitive impairment, *Entropy* 20 (1) (2018) 35.

- [11] J.P. Amezquita-Sanchez, N. Mammone, F.C. Morabito, S. Marino, H. Adeli, A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals, *J. Neurosci. Methods* 322 (2019) 88–95.
- [12] N. Kulkarni, Use of complexity-based features in diagnosis of mild Alzheimer's disease using EEG signals, *Int. J. Inf. Technol.* 10 (1) (2018) 59–64.
- [13] K.D. Tzimourta, et al., EEG window length evaluation for the detection of Alzheimer's disease over different brain regions, *Brain Sci.* 9 (4) (2019) 81.
- [14] S. Buyrukoglu, Early detection of alzheimer's disease using data mining: comparison of ensemble feature selection approaches, *Konya J. Eng. Sci.* 9 (1) (2021) 50–61.
- [15] S. Buyrukoglu, Improvement of Machine Learning Models' Performances based on Ensemble Learning for the detection of Alzheimer's Disease, in: In 2021 6th International Conference on Computer Science and Engineering (UBMK), IEEE, 2021, pp. 102–106.
- [16] H. Yu, X. Lei, Z. Song, C. Liu, J. Wang, Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification, *IEEE Trans. Fuzzy Syst.* 28 (1) (Jan. 2020) 60–71, <https://doi.org/10.1109/TFUZZ.2019.2903753>.
- [17] Digambar V. Puri, Sanjay L. Nalbalwar, Anil B. Nandgaonkar, Jayanand P. Gawande, Abhay Wagh, Automatic detection of Alzheimer's disease from EEG signals using low-complexity orthogonal wavelet filter banks, *Biomedical Signal Processing and Control*, Volume 81, 2023, 104439, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2022.104439>.
- [18] H. Yu, X. Wu, L. Cai, B. Deng, J. Wang, Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation, *IEEE Trans. Neural Syst. Rehabil. Eng.* 26 (5) (May 2018) 977–986, <https://doi.org/10.1109/TNSRE.2018.2828143>.
- [19] H. Yu, X. Li, X. Lei, J. Wang, Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (10) (2019) 1973–1984, <https://doi.org/10.1109/TNSRE.2019.2939655>.
- [20] L. Pini, A.M. Wennberg, A. Salvalaggio, A. Vallesi, M. Pievani, M. Corbetta, Breakdown of specific functional brain networks in clinical variants of Alzheimer's disease, *Ageing Res Rev.* 72 (2021), 101482, <https://doi.org/10.1016/j.arr.2021.101482>. Epub 2021 Oct 2 PMID: 34606986.
- [21] Alzheimer's Disease International & McGill University. World Alzheimer Report 2021, 2021.
- [22] A.M. Pineda, F.M. Ramos, L.E. Betting, A.S.L.O. Campanharo, Quantile graphs for EEG-based diagnosis of Alzheimer's disease, *PLoS One* 15 (6) (2020) e0231169.
- [23] <https://figshare.com/articles/dataset/dataset.zip/5450293/1>.
- [24] B.R. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, *AIChE J.* 44 (7) (1998) 1596–1610.
- [25] P. Jahankhani, V. Kodogiannis, K. Revett, EEG signal classification using wavelet feature extraction and neural networks, in: IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06), 2006, pp. 120–124.
- [26] H.U. Amin, et al., Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques, *Australas. Phys. Eng. Sci. Med.* 38 (1) (2015) 139–149.
- [27] R.C. Gonzalez, *Digital image processing*, Pearson Education India, 2009.
- [28] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Recognition*: second edition WILEY-INTERSCIENCE, 2001.
- [29] K. Fukunaga, *Statistical Pattern Recognition*, second edition, ACADEMICPRESS INC., 1990.
- [30] C.J.C. Burges, A tutorial on support vector machines for pattern recognition: *Knowledge Discovery and Data Mining* 2 (1998) 121.
- [31] K. P. Bennett and C. Campbell (2000) Support vector machines: Type Explorations Newsletter, 2:1.
- [32] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlfeld, K.-r. Müller, V. V. Nikulin, Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing: in *Advances in neural information processing systems* (2007) 113–120.
- [33] D. Garrett, D.A. Peterson, C.W. Anderson, M.H. Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, *IEEE Trans. Neural Syst. Rehabilitation Eng.* 11 (2003) 141.
- [34] A. Rakotomamonjy, V. Guigue, G. Mallet, V. Alvarado, Ensemble of SVMs for improving brain computer interface P300 speller performances. *International Conference on Artificial Neural Networks*, 2005.
- [35] A. Rakotomamonjy, V. Guigue, BCI competition III: Dataset II - ensemble of SVMs for BCI P300 speller, *IEEE Trans Biomed. Eng.* 55 (2008) 1147.
- [36] I.A. Fouad, F.E.M. Labib, Role of deep learning in improving the performance of driver fatigue alert system, *Traitement du Signal* 39 (2) (2022) 577–588. <https://doi.org/10.18280/ts.390219>.
- [37] I.A. Fouad, A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms, *Ain Shams Eng. J.* 1 (2022), 101895, <https://doi.org/10.1016/j.asnj.2022.101895>.
- [38] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000) 4.
- [39] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, s.l. : Springer, 2008.
- [40] C.D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University, s.l., 2008.
- [41] Olson M. Essays on random forest ensembles. Ph.D. Thesis. 3420 Walnut St., Philadelphia, PA 19104-6206; 2018.
- [42] J.N. Morgan, J.A. ve Sonquist, Problems in the analysis of survey data, and a proposal, *J. Amer. Statist. Ass.* 58 (1963) 415–434.
- [43] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone), *Classification and Regression Trees*. Wadsworth International, Belmont, CA, 1984.
- [44] L. Breiman, Bagging predictors, *Mach. Learn.* 26 (2) (1996) 123–140.
- [45] T.K. Ho, The random subspace method for constructing decision forests, *IEEE Trans. Pattern Anal Mach Intell* 20 (8) (1998) 832–844.
- [46] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [47] M. Dekking, *A Modern Introduction to Probability and Statistics*, Springer, 2005, pp. 181–190.
- [48] W.D. Hosmer, S. Lemeshow, *Applied Logistic Regression*, 2nd edition, Wiley, New York, 2000.
- [49] R. Pearl, L.J. Reed, J.F. K, The logistic curve and the consensus count of 1940, *Science*, 14 (1940) 895:901.
- [50] G. Yangın, Xgboost ve Karar Ağacı Tabanlı Algoritmaların Diyabet Veri Setleri Üzerine Uygulaması, Yüksek Lisans Tezi, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2019.
- [51] C.-Y.-J. Peng, R.L. Lee, G.M. Ingersoll, An introduction to logistic regression analysis and reporting, *J. Educ. Res.* 96 (1) (2002) 3–14, <https://doi.org/10.1080/00220670209598786>.
- [52] S. Baş, A. Uzun, *Tedarik Zincirinde Müşteri Siparişlerinin Lojistik Regresyon Analizi İle Değerlendirilmesi*, Ömer Halis Demir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 11(3) (2018) 67–81.
- [53] T.E. Hewett, K.E. Webster, W.J. Hurd, Systematic selection of key logistic regression variables for risk prediction analyses: a five-factor maximum model, *Clin. J. Sport Med. Off. J. Can. Acad. Sport Med.* 29 (1) (2019) 78–85, <https://doi.org/10.1097/JSM.00000000000000486>.
- [54] V.J. Felitti, R.F. Anda, D. Nordenberg, D.F. Williamson, A.M. Spitz, V. Edwards, J. S. Marks, Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *The Adverse Childhood Experiences (ACE) Study*, *Am. J. Prev. Med.* 14 (4) (1998) 245–258.
- [55] S.J. Cranmer, E.J. Menninga, P.J. Mucha, Kantian fractionalization predicts the conflict propensity of the international system, *Proc. Natl. Acad. Sci.* 112 (38) (2015) 11812–11816, <https://doi.org/10.1073/pnas.1509423112>.
- [56] S. Vijayarani, S. Dhayanand, M.P. Research Scholar, Data Mining Classification Algorithms for Kidney Disease Prediction, *Int. J. Cybern. Informatics* 4(4) (2015) 13–25, 2015.
- [57] D.O. Hebb, *The organization of behavior: A neuropsychological theory*, Psychology Press, 2005.
- [58] L. Deng, D. Yu, Deep learning: methods and applications. *Foundations and Trends®, Signal Process.* 7 (3–4) (2014) 197–387.
- [59] K. Makantasis, K. Karantzalos, A. Doulamis, N. Doulamis, deeply supervised learning for hyperspectral data classification through convolutional neural networks, in: *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, IEEE, 2015, July., pp. 4959–4962.
- [60] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015. arXiv preprint arXiv: 1511.06434.
- [61] I. Higgins, L. Matthey, X. Glorot, A. Pal, B. Uria, C. Blundell, et al., 2016. early visual concept learning with unsupervised deep learning. arXiv preprint arXiv: 1606.05579.
- [62] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, K. Talwar, semi-supervised Knowledge transfer for deep learning from private training data, 2016. arXiv preprint arXiv:1610.05755.
- [63] L. Liu, C. Shen, A. van den Hengel, The treasure beneath convolutional layers: cross-convolutional-layer pooling for image classification, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, p. 4749–4757.
- [64] Deng, Jia, et al., Imagenet: A large-scale hierarchical image database, *Comput. Vis. Patt. Recognit.*, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
- [65] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *A Practical Guide to Support Vector Classification*, National Taiwan University, May 2016.
- [66] I.A. Fouad, F.E.Z.M. Labib, M.S. Mabrouk, et al., Improving the performance of P300 BCI system using different methods, *Netw Model Anal Health Inform Bioinforma* 9 (64) (2020), <https://doi.org/10.1007/s13721-020-00268-1>.
- [67] F.-E.-Z.-M. Labib, I.A. Fouad, M.S. Mabrouk, et al., Multiple classification techniques toward a robust and reliable P300 BCI system, *Biomed Eng Appl Basis Commun.* 32 (02) (2020) 2050010.
- [68] W. H. Organization and others, *Dementia: a public health priority*. World Health Organization, 2012.
- [69] J. Jeong, EEG dynamics in patients with Alzheimer's disease, *Clin. Neurophysiol.* 115 (7) (2004) 1490–1505.
- [70] C. Patterson and others, "World alzheimer report 2018," 2018.
- [71] J. Dauwels, F. Vialatte, A. Cichocki, Diagnosis of Alzheimer's disease from EEG signals: where are we standing? *Curr. Alzheimer Res.* 7 (6) (2010) 487–505.
- [72] A. Alberdi, A. Aztiria, A. Basarab, On the early diagnosis of Alzheimer's disease from multimodal signals: a survey, *Artif. Intel. Med.* 71 (2016) 1–29.
- [73] R. Cassani, M. Estrella, R. San-Martin, F.J. Fraga, T.H. Falk, Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment, *Dis. Markers* 2018 (2018).

I.A. Fouad and F. El-Zahraa M. Labib

Islam A. Fouad is an Assistant Professor at the Department of Biomedical Engineering-Faculty of Engineering at MISR University for Science and Technology (MUST), 6th October, Egypt. He received his Ph.D. in System and Biomedical Engineering in 2014 from Cairo University, Egypt. His research interests are in the areas of digital image processing, microarray image analysis, brain-computer interface, and computer-aided diagnosis. He has several research papers published in peer-reviewed international journals. Dr. Islam serves also as a reviewer for different high reputation journals in related fields.

*Biomedical Signal Processing and Control 86 (2023) 105266*

Fatma El-Zahraa M. Labib is a senior biomedical engineer at the Department of Biomedical Engineering-New Ksar Ainy Teaching Hospital at Cairo University, Cairo, Egypt. She received her Ph.D. in System and Biomedical Engineering in 2020 from Cairo University, Egypt. Her research interests are in the areas of digital image processing, microarray image analysis, brain-computer interface, and computer-aided diagnosis. She has several research papers published in peer-reviewed international journals. Dr. Fatma serves also as a reviewer for different high reputation journals in related fields.