Cognitive Computation (2022) 14:1247–1268
https://doi.org/10.1007/s12559-022-10033-3

Efficient Deep Neural Networks for Classification of Alzheimer's Disease and Mild Cognitive Impairment from Scalp EEG Recordings

Saman Fouladi[1] · Ali A. Safaei[1] · Nadia Mammone[2] · Foad Ghaderi[3] · M. J. Ebadi[4]

Received: 25 October 2021 / Accepted: 26 May 2022 / Published online: 6 June 2022

Abstract

The early diagnosis of subjects with mild cognitive impairment (MCI) is an effective appliance of prognosis of Alzheimer's disease (AD). Electroencephalogram (EEG) has many advantages compared to other methods in the analysis of AD in an early stage. In this paper, two different deep learning (DL) architectures, including modified convolutional (CNN) and convolutional autoencoder (Conv-AE) neural networks (NNs), are proposed for classifying subjects into AD, mild cognitive impairment (MCI), and healthy control (HC) data based on scalp EEG recordings. The database includes 19-channel EEG recorded from 61 healthy control, 56 MCI, and 63 AD subjects. Time–frequency representation (TFR) is used to extract desirable features from EEG signals. Continuous wavelet transform (CWT) with Mexican hat function (MHf) as its mother wavelet is used for the selected TFR. The average accuracy obtained for the modified convolutional network and the convolutional auto-encoder network are 92% and 89%, respectively. The proposed networks in this study have superiority over those in similar studies not only by providing 10% increase in classification accuracy but also by improving the number of classes for similar data. In addition, the obtained accuracy of our networks was significantly higher than that of conventional machine learning methods. We believe the results illustrate DL architectures to be a good tool to handle EEG analysis, because of the ability to deal directly with inaccurate, inconsistent, and Para complete data, thereby providing a practical analysis.

Keywords Alzheimer · s disease · Convolution neural network · Auto-encoder neural network · Continuous wavelet transform

✉ M. J. Ebadi
ebadi@cmu.ac.ir

Saman Fouladi
smnfouladi@yahoo.com

Ali A. Safaei
aa.safaei@modares.ac.ir; ali.aa.safaei@gmail.com

Nadia Mammone
nadia.mammone@unirc.it

Foad Ghaderi
fghaderi@modares.ac.ir

1 Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

2 DICEAM Department, Mediterranea University of Reggio Calabria, Via Graziella, Feo di Vito, Reggio Calabria 89122, Italy

3 Human-Computer Interaction Lab, Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran

4 Department of Mathematics, Chabahar Maritime University, Chabahar, Iran

Introduction

Dementia is a comprehensive class of brain illnesses implying not only to a sporadic impairment but also to several syndromes distinguished by various emotional, cognitive, and behavioral disorders that make a long-time and usually slight decrease in the strength of thinking and remembering capabilities. This has a significant impact on the normal functioning of a person and can affect their life significantly [1].

Alzheimer's disease (AD), a neurodegenerative disease or a permanent neurodegenerative disorder, typically begins gradually, worsens over time, and causes a loss of connections between neurons. The AD prevalent is the same on the average of 1.4% of men and women aged 65–70, and on the average of 24% of men and women aged over 85. In recent years, pathology, biochemistry, imaging, and genetic review have provided ways for patients to recognize and provide treatment. The most common initial sign is hardness in recalling new events (losing short-time memory).

As the disorder progresses, the signs can involve problems in speech, losing direction (involving getting lost easily), losing motivation, mood swings, lack of self-care managing, and behavioral matters. As a condition of person decreases, he/she usually recede from society and family. Slowly, the functions of the body are losing, which finally leads to death. Although the progress speed can change, the life expectancy average after recognition is 3 to 9 years [2].

Mild cognitive impairment (MCI) is well known as isolated memory disorder and early dementia. It is a neurological impairment that arises in elders, which includes cognitive disorders with a minimal disorder in daily living necessary activities. MCI includes the beginning and progress of cognitive disorders beyond those expected relying on the education and age of the person, though which are not sufficiently important to meddle in their ordinary activities. It may happen as a transmutation phase between natural aging and dementia. The progression of MCI to Alzheimer's is usually progressive. An MCI is an interstitial process between the natural cognitive analysis of aging and acute cognitive impairment resulting from intellectual decline. This disorder involves more severe problems than the usual expectation of aging in memory, speech, thinking, and judgment. Mild cognitive impairment increases the risk of dementia due to Alzheimer's and other neurological disorders, but the disorders of some patients do not intensify and some will eventually recover. Symptoms of MCI may persist for many years, end in Alzheimer's or another type of mental decline, or may improve over time. Evidence suggests that MCI is often (not always) due to mild changes in Alzheimer's or other types of dementia. Some of these changes have been identified in the autopsy of corpses with MCI [3, 4].

Electroencephalogram (EEG) signals catch the brain's electrical activity and are one of the most important sources of information for studying brain activities and neurological disorders. For this reason, automated systems for detecting EEG changes have been investigated for consecutive years. Psychiatrists cannot make decisions quickly because of considerable overlaps of signs among various mental impairments. The EEG produces, in principle, a strong and almost inexpensive way to investigate dementia and AD in their early stages. Appropriate analysis of this electrical signal has a vital role in the field of brain-computer interfaces to form the communication channels between computers and the human brain. Previous researches that use EEG signals have focused on reducing the oscillatory rhythms of the brain, accompanied by reducing the complexity of the relative time-series and the raised compressibility of them. The analyses as mentioned above have usually been done on One-Channel EEG signals. The investigation of the level of screening on recorded EEGs of patients at peril could be used to underline the emergence of underscored progression of AD or leastwise protect more clinical attention. According to the above essential expressions about AD and MCI, we find that the accurate classification of AD and its early stage, MCI, plays a vital role in intercepting the progression of memory disorders and contributing to ameliorating the quality of life of AD patients.

Signal pre-processing methods must be used for extracting suitable and complete features from EEG signals. Empirically, CWT with the MHf used as mother wavelet has been suggested in the state of the art [5]. Machine learning (ML) models have been widely used to classify EEGs of AD and MCI subjects. ML methods are not appropriate to process high-dimensional data. DL is a developed ML technique able to extract the most suitable features directly from initial input data and widely used in previous studies. The proposed CNN and Conv-AE are compared with standard classifiers (logical regression (LR), multilayer perceptron (MLP), nearest neighbor (NN), random forest (RF), support vector machine (SVM), and stochastic gradient decrease (SGD)). Results demonstrated that the proposed CNN and the Conv-AE network outperformed all other approaches achieving an average accuracy rate up to 92% and 89%, respectively, in three classes (AD vs MCI vs HC) classification.

To the best of our knowledge, not only the accuracy of our methods but also the number of classes are excess to the existing classification methods. In Table 1, we prepare a brief survey of EEG classification algorithms using [5].

The following list shows the contributions of the paper:

- We firstly divided each signal into two-second epochs based on their frequency;
- We used TFR to show signals in both time and frequency;
- We successfully used CWT with the Mexican hat (Ricker) wavelet for signal pre-processing;
- We developed efficient DL approaches based on a CNN and Conv-AE to classify AD, HC, and MCI signals by only analyzing noninvasive recorded EEG signals;
- We developed six types of ML classifiers for classifying EEG signals into three classes of MCI, AD, and HC subjects;
- We evaluated our proposed methods using several evaluation criteria;

The organization of the paper is as follows. In "Related Work," some related works are reviewed. Material and methods including pre-processing signal, continuous wavelet transform, convolutional neural network, and Conv-AE neural network are discussed in "Material and Methods." "Experiment Results" includes the experimental results. The sum-up of the paper is given in "Discussion."

Table 1 EEG classification algorithms

Linear classifiers	Neural networks classifiers	Nonlinear Bayesian classifiers	Nearest neighbor classifiers	Combinations of classifiers	Other classifiers
Linear discriminant analysis (LDA) [6, 7, 9]	Multi-layer perceptron [17–25, 30]	Bayes quadratic [28, 40, 54]	K nearest neighbors [12, 15, 65, 66, 70]	Boosting [53, 54]	Logistic regression (LR) [58–61, 75, 76]
SVM [8, 11, 13, 14, 16, 71, 72]	Other neural network architectures [10, 16–27, 31–35, 37–39, 41]	Hidden Markov model [36, 57–59]	Mahalanobis distance [15, 29]	Voting [24, 55, 77, 93]	Dynamic wavelet network [62–67]
	DNNs [42–52]	Other Bayesian Network [60, 62–64, 73, 74]		Stacking [56, 57]	Decision tree [68, 69, 77]

Related Work

In order to distinguish HCs, ADs, and MCIs from a dataset of EEG signals, some state-of-art algorithms have been used as follows. To classify ADs and healthy persons, the authors in [74] used an artificial neural network (ANN) and obtained an accuracy of 90%. SVM classification was used in [75] to classify HC and AD subjects with an accuracy of 79.9% and various features on the base of coherence were appraised. The authors in [76] analyzed a set of EEG signals including ADs, MCIs, and HCs in the domain of frequency and found that the global frequency power (GFP) of theta and alpha was the best distinguishing indexes of classification of AD vs MCI and AD vs HC with the accuracies of 78% and 84%, respectively. A new system on the base of ANNs is proposed in [77], which is known as implicit function as squashing time (IFAST), to distinguish EEG signals involving AD and MCI subjects in relaxing position in which patients' eyes were either closed or open. In that paper, the weights of connections are used as inputs of a nonlinear ANN model that is auto-associative and implementing that model reached an accuracy of 92.33%. The IFAST method was examined in [78] to classify healthy subjects and MCI patients automatically, obtaining an accuracy of 93.46%. In [79], a model of SVM including regional spectral and complexity features is used to specify the types of 17 AD, 16 MCI, and 15 HC subjects of a dataset of EEG signals that were recorded according to a particular protocol involving various actions and cognitive duties. The investigated states were eye-closed relaxing, eye-closed counting, and eye-opened relaxing, and the classification accuracies were 79.2%, 85.4%, and 83.3%, respectively. The authors in [80] classified AD, MCI, and HC subjects collected from the ADNI database, a dataset of MRI scans, by the combination of 3D-CNN with the sparse auto-encoder method. They reported classification accuracies of 86.64%, 89.47%, 92.11%, and 95.39% for AD vs MCI, MCI vs AD vs HC, MCI vs HC, and AD vs HC, respectively. The classification accuracies for MCI vs HC, AD vs MCI, and AD vs HC from the ADNI dataset achieved 65.6%, 69.5%, and 91.4%, respectively in [81]. In [82], the authors obtained an accuracy of 96.85% for classification AD and HC subjects from fMRI data. Recently, some researchers, for example, in [42, 52], and [83], have used different DL methods to classify MCI vs AD vs HC subjects. In [42], the authors presented a stacked AE method to classify rapidly progressive dementia (RPD), Creutzfeldt-Jakob disease (CJD), HC, and AD subjects, and they obtained specificity, sensitivity, and average accuracy of 85%, 94%, and 88% in the classification of AD vs CJD subjects. Also, they achieved comparable results for HC vs CJD, and RPD vs CJD subjects. The average accuracies of 85%, 78%, 85%, and 82% were achieved for HC vs AD, MCI vs AD, HC vs MCI, and MCI vs AD vs HC subject classification, respectively. In the recently published paper [52], the average accuracies of 92.29%, 84.62%, 91.88%, and 83.33% were achieved to classify AD and HC, AD and MCI, HC and MCI, and MCI vs AD vs HC subjects, respectively. Despite providing high performance of classification in some of the above models, there is not an evident and standardized protocol of EEG recording for classification of MCI/AD. Also, they need highly complicated algorithms and costly foundations.

Material and Methods

The main steps of this research shown in Fig. 1 consists of the following: (1) recording EEG signals; (2) dividing signals into 2-s epochs based on frequency; the signals are divided into two-dimensional (2D) arrays with 19 rows representing 19 recorded signal channels and 512 columns; (3) vectorizing 2D arrays resulted from 2-s divisions to 1D arrays; (4) TFR of signals using CWT; and (5) converting 1D array of represented data in the time–frequency format to 4 dimensions inputs as multilevel convolution and multilevel Conv-AE for network training and signal classification into three classes of AD, MCI, and HC individuals. In this

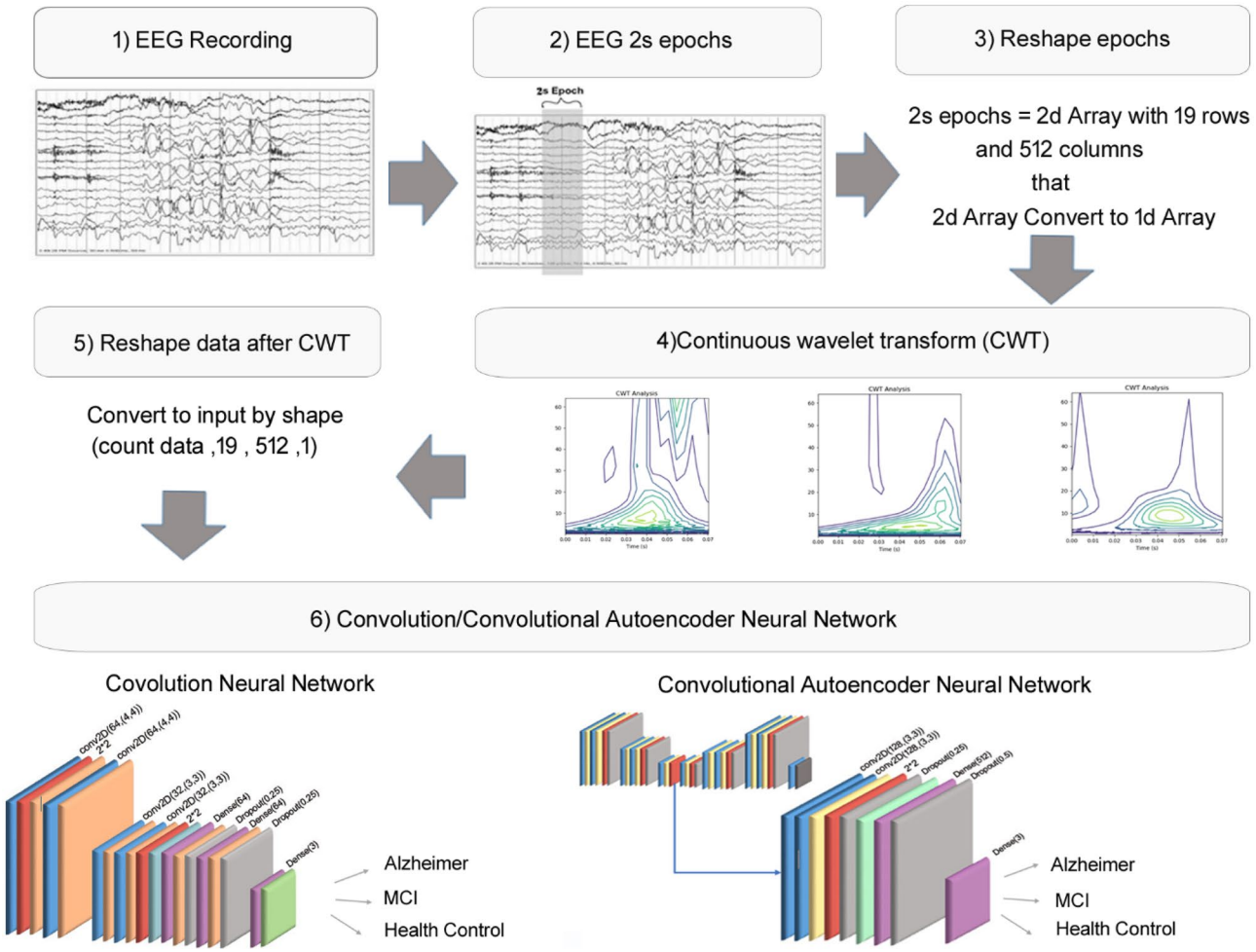

Fig. 1 EEG signals are recorded and stored on the computer. The signals are then converted to 2-s epochs by frequency. These epochs are 2D arrays with 19 rows and 512 columns, transformed into 1D arrays and then into wavelets that have signal-frequency data by CWT functions. All data is divided into training, and test sets, and their dimensions must change to be input to the networks. In this study, two deep convolution and auto-encoder convolutional networks were used to regulate the data into three classes AD, MCI, and HC

study, two networks of deep convolution and deep Conv-AE are designed to classify data. The convolution network consists of four convolution layers, two max-pooling layers, three dense layers, and two dropout layers, along with one softmax-layer designed for classification into three classes of AD, MCI, and HC. The architecture of the Conv-AE network includes two classes of training and classification. The training part consists of five convolution layers, three max-pooling layers, and two dropout layers in the encoder, as well as six convolution layers, three max-pooling layers, three dropout layers in the decoder, and a cropping layer to adjust the data size. Having been trained based on the network architecture, the output of the last encoder layer is applied as the input of a convolution network with two convolution layers, one max-pooling layer, two dropout layers, and two dense layers in order to classify the EEG data into three classes of AD, MCI, and HC.

In this study, after dividing the data into 2-s epochs, all the divisions are stored. Labels are arrays with 19 rows and 1 column, including the number considered for each group. The number 0 for the HC group, and the number 1 for the MCI group and the number 2 for the AD group are considered as data labels. Until this stage, the 2D arrays containing only signal time data are recalled one by one from the data folder. As the used function representing the time–frequency of signals only accepts 1D arrays as input, data are first converted to 1D arrays. Next, the TFR of the signals is performed using the Mexican hat wavelet transform function, and then, all the data are stored in a list and randomly divided into training and test data at specific percentages. The proposed deep networks are 2D, and the input data to these networks must be 3D. The first dimension corresponds to the number of initial data rows (19), the second dimension to the number of initial data columns (512), and the third dimension to 1. The

training and test data, together with their labels for network training and testing, must be entered into the fit function with four dimensions. The first dimension equals the total number of training or test data that are divided randomly in specific percentages, the second dimension refers to the number of initial data rows (19), the third dimension is equal to initial data columns (512), and the fourth dimension is set to 1. After training the network by the training data, the test data set is predicted based on what the network has learned.

EEG and Pre-processing Signal Recording

In the current work, we chosen 74 patients and 16 healthy persons randomly from the dataset investigated in [42] who were dementia patients, including 37 AD and 37 MCI individuals. The patients were at distinct clinical stages of the disease progression and were chosen at the IRCCS. Each patient has been appointed according to a continuous cooperation contract among IRCCS and DICEAM. The patients approved their cooperation by signing an informal approval form. The related protocol of the clinic has been authorized by the local Ethics Committee. An evaluating board performed all cognitive and clinic evaluations. The diagnostic process was in the accommodation of the standards of the Aging-Alzheimer's Association National Institute. Following verification of diagnosis, the patients have been classified by sex, age, marital status, education, the approximate age of the beginning of dementia, and Mini-Mental State Examination (MMSE) standard test. The quantity of current usage of any drugs, especially anti-depressants, anti-epileptic, memantine, and cholinesterase inhibitors (ChEis), has been considered with the fact that all patients had been using those for more than 3 months prior to assessment. The recording of all EEG signals was during a no-task resting state of the patients, which is well known as the gold standard. The international 10–20 standard system has placed the scalp electrodes. A2 is used as a contralateral referencing of all electrodes, which is placed on the behind of the outer right ear.

Prior to the recording of EEGs, all patients and the caregivers have been interviewed with the following questions:

a) How was the quality of your sleep last night?
b) How was the amount of the night prior to the recording?
c) How much time did your meal and what was its content?

The recording of EEGs was carried out in the morning. Within the EEG recording, the patients' eyes were closed, but they stayed awake: in fact, the medical technician hindered them from sleepiness by calling their names. None of the subjects slept within the EEG recording, as EEG visual reviews approved, which contained seeking the possibility of occurring sleeping patterns in the recording of EEGs. Despite being 5-min duration of any recording, several portions of the various recording length have been thrown away manually because of the appearance of small and big artifacts or unanticipated action. Though, this phase produces recording of the various period, it is essential the processing of sole epochs with artifact-free and keeping away from missing possibly beneficial information in any type of procedure of automated artifact rejection. Moreover, EEGs were band-passed filtered between 0.5 and 32 Hz, to set the bands examined related to AD assessment. Compared to plenty of prior research works, we do not work with a single channel in this paper. We attempt to clarify the distinctions and correlation between all 19 channels, at both scalp plane and per region. At the beginning stage of the study, two various patients did not have the certainly similar status of the pathology, as pointed out by the score of MMSE. In other words, the combination of patient and healthy samples of both AD and MCI, and HC is not homogeneous. This will reverberate in several published works variability.

2s EEG Signal Epochs

Each signal stored on the computer with a frequency of 256 Hz was divided into 2-s epochs with no overlap. As the frequency was 256 Hz, each epoch contained $N = 512$ (2×256) samples. The recorded epochs for each individual (EEG € (€ = 1, 2, …, M)) was stored on the computer with a size $N \times n$.

$$\mathrm{n} = \frac{\mathrm{EEG}}{256 \times 2} \tag{1}$$

The signals were recorded in 5-m durations, but some lasted more than 5 min, especially in the control group. The data from 37 AD, 37 MCI, and 16 HC individuals were used to have a balanced database. From 2-s epoch divisions of 90 randomly selected data, a total number of 16,449 data samples including 5480 AD, 5444 MCI, and 5525 HC was obtained.

Remark 1. It should be noted that the data were initially converted to 5-s epochs, each containing $N = 12080(256 \times 5)$ samples, and the accuracy of the network classification had been tested with that data size. Afterward, the classification accuracy significantly improved by changing the data division into 2-s epochs employed in this study.

Continuous Wavelet Transform

Most of the vital signals are non-stationary and their attributes change over time**.** To focus on the interest events of these signals, time–frequency analysis is usually used. Short-time Fourier transform (STFT) is a common way of time–frequency analysis. It has some limitations on

resolutions; e.g., an improvement in frequency resolution leads to a weakness in time resolution and vice-versa. To cope with STFT limitations, wavelet transform is proposed.

The weakness of the raw EEG signals results from low spatial resolution, low signal-to-noise ratio, or artifacts. Wavelet transform is used for reducing signal noise and feature extraction. Wavelet transform is a powerful signal processing method developed to fix the shortcomings of the Fourier transform. This method is like a mathematical microscope that can analyze EEG signals at different scales. In addition, the wavelet transformation is a common and popular method to remove signal noise. It can provide information related to the signals in various frequency bands [84]. Wavelet transform uses the feature of high-frequency bursts happening on short intervals and low-frequency signals that are widespread over time. There are two ways to use wavelet analysis, CWT and discrete wavelet transform (DWT). CWT utilizes a wavelet function and generates a scalogram similar to a spectrogram for time–frequency analysis. DWT breaks down a signal into two signals: (1) high-pass signal or difference using wavelet function and (2) low pass signal or medium using scaling function. Therefore, the use of CWT at various wavelet functions can provide a more detailed analysis [85]. In addition, CWT can remove signal noises and produce higher peaks with sharper amplitudes [86]. Wavelet transform techniques have advantages over Fourier transform techniques for analyzing EEG signals. They significantly decrease the computational time required for analysis without any reduction in the accuracy of feature extraction [87]. The EEG signal is continuous and has continuous parameters such as time and scale, for which CWT is the best method to extract useful features. The equations of CWT are:

$$CWT(a,b) = \int_{-inf}^{inf} x(t)\Psi^*(t)dt \tag{2}$$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi(\frac{t-b}{a}) \tag{3}$$

$$\Psi(t) = \text{MotherWavelet} \tag{4}$$

The "a" and "b" are continuous and refer to signal scaling, and time-shifting or time localization. These parameters change continuously over the time. There are two types of DWT: (1) based on frames with discrete parameters and selects positive or negative integer of powers of one. This parameter does not support the multi-resolution system, so it is not suitable for the analysis of EEG signals. And (2) based on wavelets that support multi-resolution analysis. The following shows the equation of DWT where 2^j refers to signal scaling and $2^j k$ refers to signal shifting [88]:

$$DWT(j,k) = \frac{1}{\sqrt{|2^j|}}\int_{-inf}^{inf} x(t)\Psi(\frac{t-2^j k}{2^j})dt \tag{5}$$

EEG signals are unstable and dynamic data. Therefore, time or frequency description alone is insufficient to provide comprehensive information on them, and TFR is more appropriate for such signals. TFR is a signal display (as a function of time) in both time and frequency, which serves as a function of representing and analyzing signals containing varying frequencies depending on time. For this representation, some functions known as wavelets are used. A wavelet is a set of mathematical functions used to decompose the continuous signal into its frequency-related components. A wavelet transform is a decomposition of a function based on wavelet functions. Wavelets (also known as girl wavelets) are scaled and transformed samples of a function (mother wavelet) with finite, oscillating, and extremely damp wavelengths. There are various methods and functions for wavelet

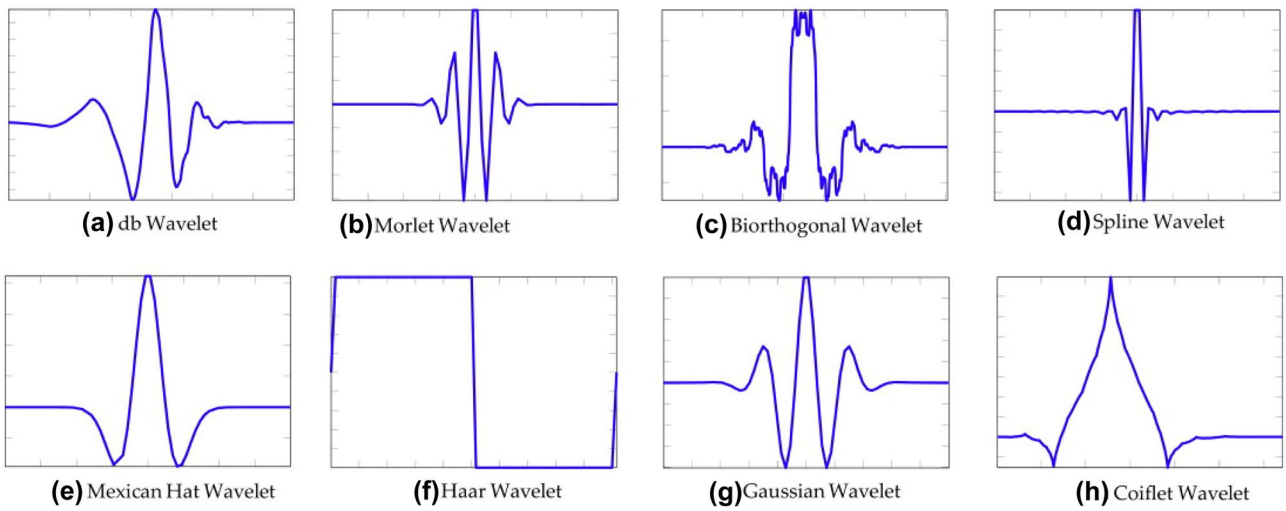

Fig. 2 All types of continuous wavelet functions [85]

transform that can be used based on the type of data and the intended use. Figure 2 demonstrates all types of wavelet functions with real and imaginary parts. They are smooth and similar to functions being derived from sinusoidal and exponential signals. However, all six types have various parameters to control the frequency band and oscillation.

Since EEG signals are continuous data and the Mexican hat wavelet function is suitable for continuous and seismic signals, the CWT method is used in this study with the Mexican hat (Riker) wavelet as the mother wavelet function for feature extraction and noise reduction.

In mathematics and numerical analysis, the Mexican hat (Riker) wavelet is:

$$\Psi(t) = \frac{2}{\sqrt{3\sigma}\prod^{1/4}}\left(1-(\frac{t}{\sigma})\right)^2 e^{-t^2/2\sigma^2} \quad (6)$$

It is the negative normalized second derivative of a Gaussian function and a particular type of continuous wavelet family (wavelets used in a CWT). The Mexican hat is often used to model seismic data. It is usually known as Mexican hat wavelet in the Americas for its sombrero shape when used as a kernel for 2D image processing [89].

After transforming 2D data arrays into a single 1D array, the wavelet transform is done for this array using the CWT function from the *scipy.signal* library in Python. Additional to the wavelet transform function, the scale parameter is another critical parameter in wavelet transform. The scale parameter for the wavelet transform acts like the scale for magnifying geographic maps. The upper-scale represents the general information, and the lower scale represents the partial information of a hidden pattern in a signal. In this study, scale (1, 19) was considered. There was no vital signal information extracted from the smaller scale, and no change in the results showed by the larger scale. The changes from the wavelet transform on a signal sample of MCI, HC, and AD can be seen in Figs. 3, 4, and 5, respectively.

Convolutional Neural Network

Deep learning–based models have recently achieved promising results in medical and healthcare trials. These models have outperformed CNN on signal processing and medical data analysis in terms of manual-crafted features. CNN is a type of the DL network in which various layers extract valuable features of input data, and these features are enormously learned by layers. These models are prevalent and practical in classification, medical image processing, and medical signal processing [90]. A CNN comprises three essential layers: convolution layer, pooling layer, and fully connected layer. The function of various layer combinations will result in network training. CNN layers can be embedded in other network architectures and make new networks for

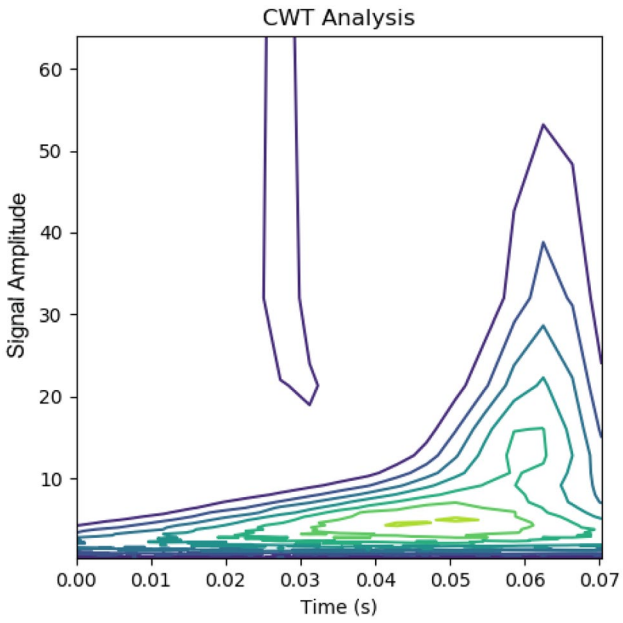

Fig. 3 TRF of MCI group

classification. Extracted features of EEG signals are used to classify data into AD, MCI, and HC classes. The subsequent main steps are applied to classify EEG signals with the aid of convolution-based models:

I. Feature extraction

In this stage, a convolution and a pooling layer are used to monitor and assess valuable features. Figure 6 demonstrates the function of stride by the kernel/filter

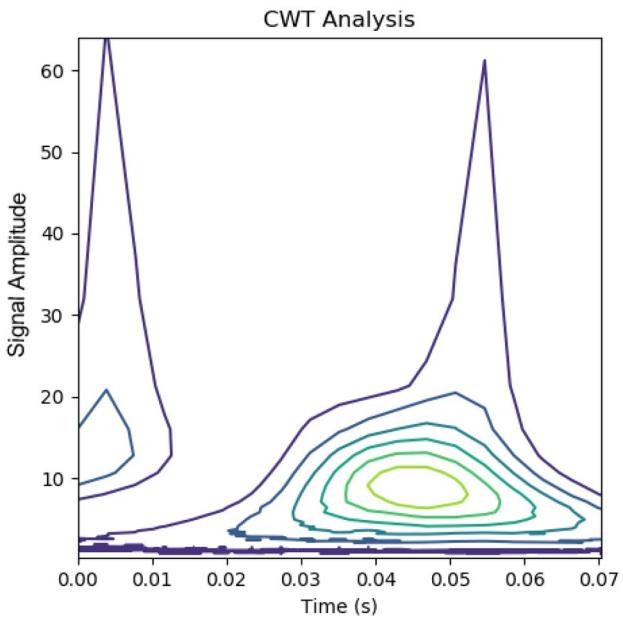

Fig. 4 TFR of HC group

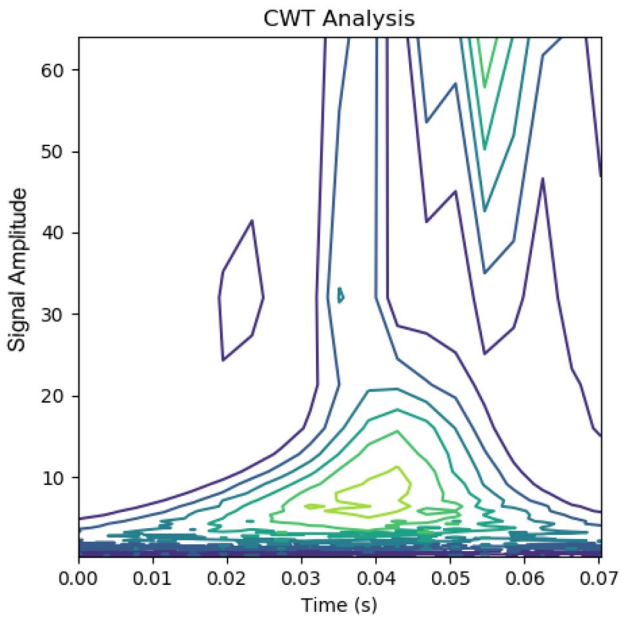

Fig. 5 TFR of AD group

for potential feature extraction. Then, the max-pooling layer is applied for spatial size reduction of the convolved features. It can cope with overfitting issues during network training. It evaluates the maximum value of the feature map generated by the convolution process. Based on Fig. 7, a max-pooling layer contains a stride of two and a kernel/filter size of two. Then, a piecewise linear function such as Rectified Linear activation function (ReLU) is used. This activation function generates output based on the input. If the input is positive, the output will be positive, and if the input is negative, the output will be zero. The ReLU is the standard activation function for most deep neural networks (DNNs) since it is an easy-to-use way of training networks and achieving better results.

II. Classification

In this step, extracted features are classified by fully connected layers and evaluated for the possibility of diseases in the input data. Figure 8 shows a fully connected layer for the classification of features into three classes [91].

In the design of a CNN architecture, standard hyper parameters such as convolution kernel size, channel number, padding type, learning rate, dropout, and stride affect both the efficiency and performance of the CNN. Hyper parameters specify the structure of layers like a feature map. To achieve desirable results from CNN, a suitable hyper parameter must be used based on the purpose of the project and the type of data. Nonetheless, there is neither obvious formula nor rules to optimize the hyper parameters; they are still specified based on the experience and insight of the developer [92]. In other words, as the CNN structure is sophisticated, it is challenging to design and implement an efficient CNN network. For instance, what is the optimal size for the convolution window and kernel, which activation function performs better, and how accurate results can be obtained by setting the learning rate of CNN. The answers to all these questions are determined by the values of hyper parameters. At present, there is a fixed architecture of CNN networks while hyper parameters are determined based on the experiences of the network developer.

The convolution kernel concept is represented in Fig. 6. The number and size of the convolution kernel matrix demonstrate the number and size of convolution kernels. Based on Fig. 6, the convolution kernel has a size of 3×3. Other odd sizes such as 5×5, 7×7, and 9×9 are also useable. The other hyper parameter is the number of convolution kernels which is an integer in the range of [1, 300]. The

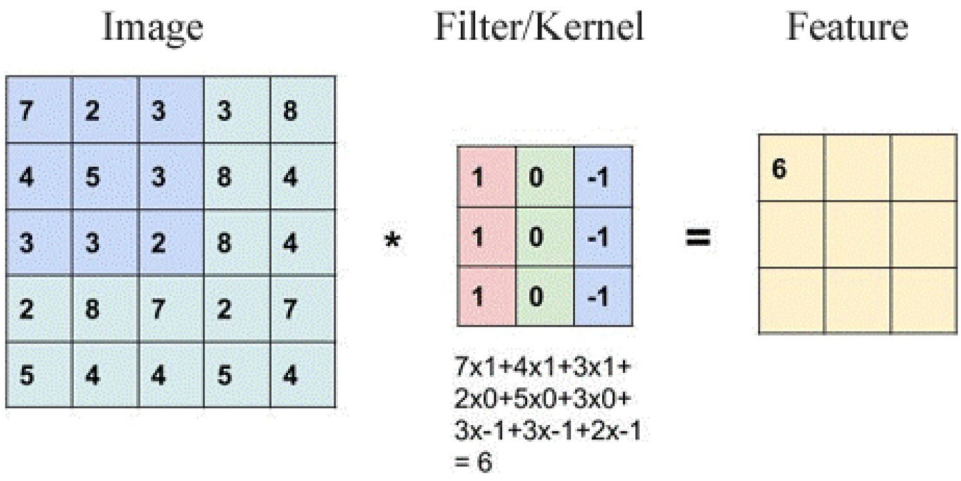

Fig. 6 CNN function (kernel size equals 3 with stride equals 1) [99]

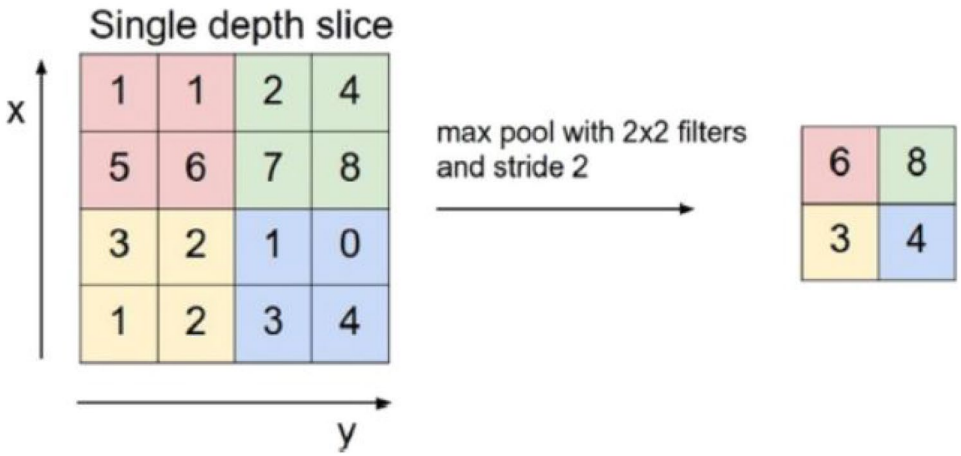

Fig. 7 Max pooling function with a pooled feature [99]

pooling layer reduces the dimension and sampling, usually known as the lower sampling layer. Generally, the pooling layer comes after the convolution layer. The extracted feature map by convolution layer is used as the input of the pooling layer. There are two types of pooling layers: (1) max-pooling and (2) average-pooling. In max-pooling, the final feature map is made of the maximum value in each window and average-pooling, the final feature map is made of the average value in each window. The convolution layers in combination with pooling layers can be used in the architecture of CNN several times. After these layers, several fully connected layers are used as the final layer of the network and the output of the final results. Dropout is a layer used to avoid overfitting during network training. Overfitting occurs in the neural network where the model is dependent on training data and cannot work with test data. Therefore, dropout is used to discard some neurons randomly. The learning rate is one of the main hyper parameters in deep neural network training. This hyper parameter controls the weight adjusting the speed of NNs based on the algorithm of loss gradient descent. The learning rate should have the right amount. If it is too small, the speed of gradient descent and thus the convergence speed will be low. Consequently, the error decline is slight, but the error of training is large after network training. On the other hand, a high rate of learning leads to a significant gradient with decreasing steps. This finally causes instabilities around the optimal weight and makes loss function divergent from the optimal solution [93].

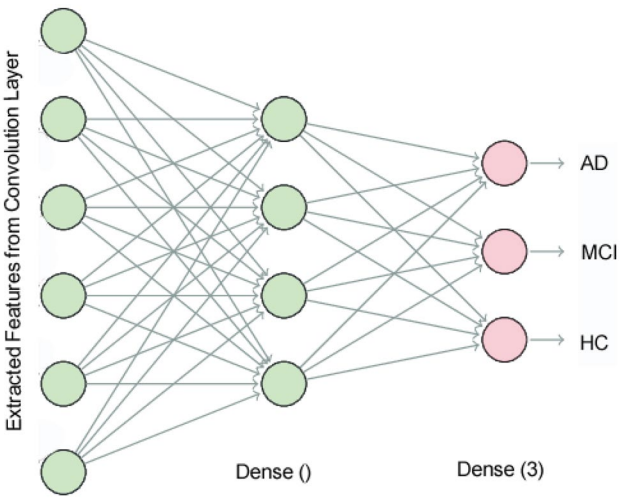

Fig. 8 Function of fully connected layer

As mentioned in [92] and [93], there are no fixed and pre-defined hyper parameters to achieve the best results and they are obtained based on training and testing networks of various hyper parameter values or based on developer experiences. In this study, common layers are used for designing networks and the results are improved by examining and changing some hyper parameters. The architecture details of proposed networks and the optimal hyper parameters are described as follows:

The suggested CNN architecture for this study is shown in Fig. 9. The 2D convolutional network requires 3D inputs designed by $19 \times 512 \times 1$ dimensions (of which 19 refers to EEG channels number, 512 is the number of signal array columns stored on the computer after 2-s epoch divisions, and 1 is the number of input data channels). From the total number of 16,449 data, 80% (13,159 numbers) were considered for network training and 20% (3290 numbers) for network testing. The network contains two convolution layers of 64 filter length and a 3×3 kernel function, followed by two convolution layers of 32 filter length and a 3×3 kernel function. Generally, a convolutional network is a hierarchical network that contains convolution layers,

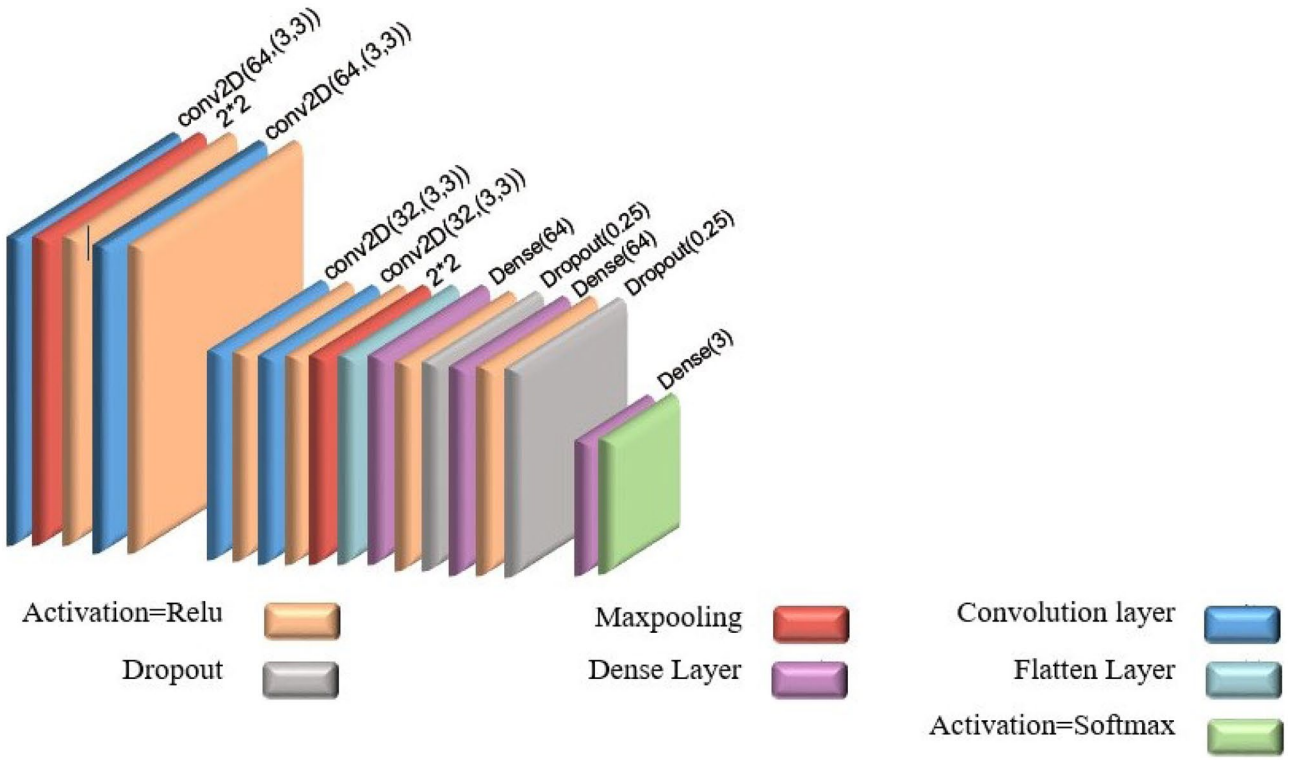

Fig. 9 Architecture of the proposed research convolution network

pooling layers, and then fully connected layers. However, these main layers are put together in different compositions. The network is composed of four convolutions with two pooling layers. Max-pooling with size 2×2 and step 2 is used in this architecture. A flatten layer converts the 2D output of the last pooling layer to a 1D array for sending to fully connected layers. A padding is used to control the output size of convolution layers. In this architecture, using the same padding was suggested to fill the boundaries of input data with similar values in bordering cells. To classify data, two fully connected layers with sizes of 64 and 3 were considered. The activation function for the last fully connected layer (with the size of 3) was softmax. To avoid overfitting, after two fully dense layers, a dropout layer with the rate of 0.25 was considered.

The input of the first dense layer with $U_{fc} = 64$ hidden layers is features extracted from the last convolution layer. The weights number is calculated based on last convolution layer output size ($y1 \times y2$), the filters number (k), and hidden layer number in dense layers. So, convolution layer weights is defined $W_{\text{conv}} = y1 \times y2 \times k \times U_{fc} = 4 \times 128 \times 32 \times 64 = 1{,}048{,}576$, where the number of current parameters to the first dense layer is $1{,}048{,}576 + 64$ (biases) $= 1{,}048{,}640$. Table 1 reports a summary of the training parameters of the CNN network. Based on Table 2, the summation of all network training parameters for classification input data into 3 classes can be computed from the sum of Param column data. The resulting sum is 1,098,243, where all parameters are trainable.

For this network, several optimization functions such as Adam, Adadelta, and SGD were examined. The obtained results from the Adam function were more precise than obtained results from other optimization functions, especially SGD. For achieving the least learning error, the learning rate with values of 0.01, 0.001, and 0.0001 was tested. The best value for the learning rate was 0.001. The CNN was trained during 100 epochs, and the data was sent to the network with a batch size of 20. Each epoch took 251 s and 19 ms, lasting 6 h to train. The network was implemented in a Jupyter notebook with a cross-library using GPU P2000 hardware of 24 cores and 128 GB memory.

Conv-AE Neural Network

As mentioned, the architecture of a Conv-AE network was also designed and implemented in this study. Instead of training the network and predicting the objective value Y for the input X, an auto-encoder is trained to reconstruct X input. EEG signal samples are large-scale data without independent dimensions. On the one hand, it can be difficult to extract features from auto-encoders. The sequence of signals can be understood by convolution to achieve local features. This sequence is based on the core convolution.

Table 2 Training parameters of CNN for three classes' classification

Layer (type)	Output shape	Param #
conv2d_21 (conv2D)	(None, 19, 512, 64)	1088
max_pooling2d_8 (maxPooling2D)	(None, 9, 256, 64)	0
activation_13 (Activation)	(None, 9, 256, 64)	0
conv2d_22 (conv2D)	(None, 9, 256, 64)	16,448
activation_14 (Activation)	(None, 9, 256, 64)	0
conv2d_23 (conv2D)	(None, 9, 256, 32)	18,646
activation_15 (Activation)	(None, 9, 256, 32)	0
conv2d_24 (conv2D)	(None, 9, 256, 32)	9248
activation_16 (Activation)	(None, 9, 256, 32)	0
max_pooling2d_9 (maxPooling2D)	(None, 4, 128, 32)	0
flatten_3 (Flatten)	(None, 16,384)	0
dense_7 (Dense)	(None, 64)	1,048,640
activation_17 (Activation)	(None, 64)	0
dropout_10 (Dropout)	(None, 64)	0
dense_8 (Dense)	(None, 64)	4160
activation_18 (Activation)	(None, 64)	0
dropout_11 (Dropout)	(None, 64)	0
dense_9 (Dense)	(None, 3)	195
Total params: 1,098,243 Trainable params: 1,098,243 Non-trainable params: 0		

Using multiple convolution cores to sample signals allows obtaining a variety of local features. At the same time, the features of the model can be reduced by lowering the sampling rate. Therefore, the final features of the sampled signal can be extracted with repeated convolution and small sampling [94]. In this study, the purpose of using the auto-encoder network is to train and classify data combining this network with the convolution network rather than reconstructing the input signals. Figure 10 shows the architecture of the proposed Conv-AE network of this research. The network consists of two sections; the first section includes a Conv-AE network to train, and the next section is a simple CNN to classify that use the output of the last layer of the encoder part of the Conv-AE network (first section). The first part architecture is composed of 2D multilayer convolutional networks for both the encoder and the decoder. The 3D input of this network is designed with $1 \times 512 \times 19$ dimensions (in which 19 refers to EEG channels number, 512 is the number of signal array columns stored on the computer after 2-s divisions and 1 is the number of input data channels). From the total number of 16,449 data, 67% of them (11,020 numbers) were considered as training set and 33% (5429 numbers) were considered as testing set. It is worth noting that the division of training and test data for this network had been done with 80% and 20% percentages, and then was changed to 67% and 33% in order to improve the division accuracy. The encoder part consists of two convolution layers with the filter size of 128 and 3×3 kernel function, two convolution layers with the filter size of 64 and 3×3 kernel function, and one convolution layer with filter size of 32 and 3×3 kernel function. In the encoder, after two

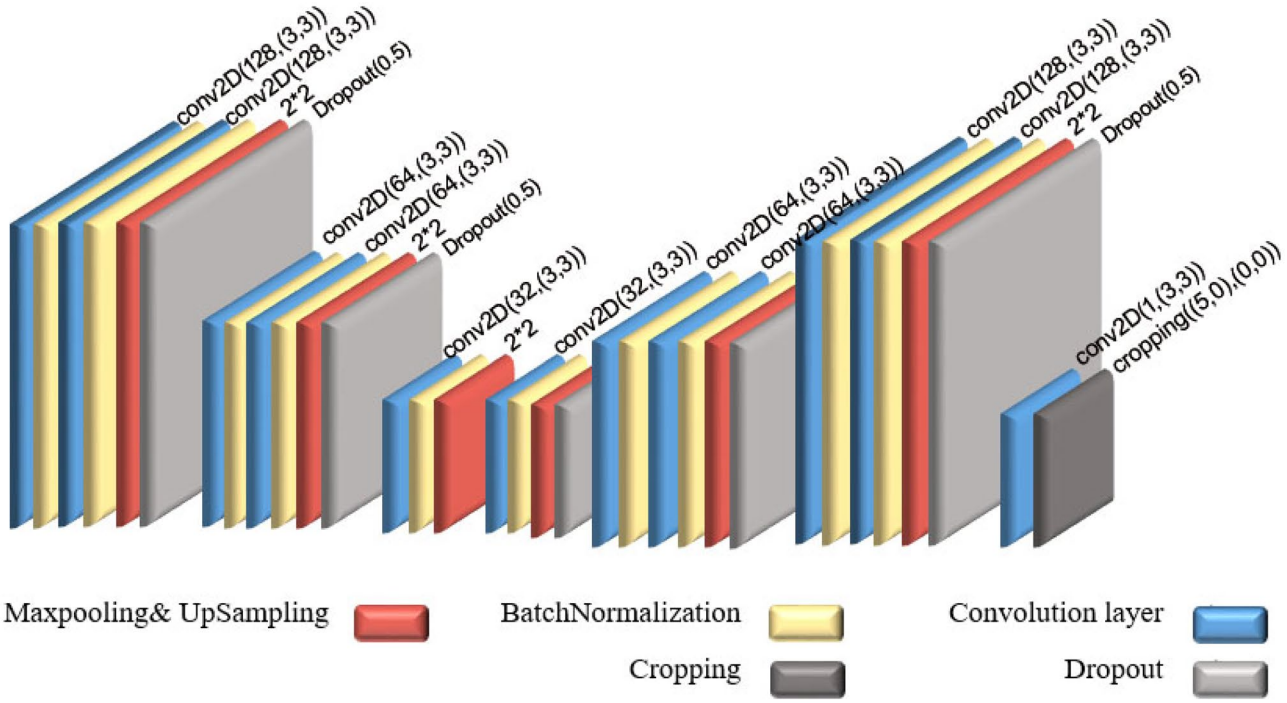

Fig. 10 The architecture of the proposed Conv-AE network

convolution layers, a max-pooling layer with size 2×2 and padding 2 is considered. The decoder section of the network contains a convolution layer with a filter size of 32 and 3×3 kernel function, two convolution layers with the filter size of 64 and 3×3 kernel function, and two convolution layers with a filter size of 128 and 3×3 kernel function, as well as a convolution layer of 1 filter length and 3×3 kernel function. In the decoder, a cropping layer adding five columns to the bottom of the data matrix is used to moderate the data size. There is not necessarily an upsampling layer after each convolution layer, and a 2×2 upsampling layer is used after two convolution layers. For this network, the same padding is used. In this network, to avoid overfitting, the batch normalization layer was used after the convolutional layer, and dropout (0.5) was used after each max-pooling layer except for the last layer in the encoder.

The last layer of the encoder section extracted valuable features of the EEG data and the output of this layer can be used as input for a simple CNN to classify data (Fig. 11). For the classification to be more accurate, the output is retrained by two convolution layers with a 128 filter size, 3×3 kernel function, and a 2×2 max-pooling layer with a step size of 2. To cope with overfitting, batch normalization and 0.25 dropout layers are used. To send max-pooling output to a 512 fully connected layer, a flatten layer is considered. Since there are three classes, the last layer is three fully connected, which is separated from the prior fully connected layer by a 0.5 dropout layer. The ReLU activation function was used for all layers.

For the Conv-AE network, the used parameters in the encoder part and classifier part are significant for training and classification. Extracted features from the encoder last layer are re-trained by a simple CNN, and the ultimate extracted features are used as the input of the first fully connected layer to the hidden layer of $U_{fc} = 512$. The number of weights W_{conv} depends on the output size of flatten layer and the number of hidden layers in the fully connected layer. The output of flatten layer equals $32 \times 128 \times 1 = 4{,}096$. Thus, the number of weights is $W_{\text{conv}} = \text{Outflatten} \times U_{fc} = 4096 \times 512 = 2{,}097{,}152$ and the number of existing parameters to the second fully connected layer is $2{,}097{,}152 + 512$ (biases) $= 2{,}097{,}664$. Table 3 illustrates the learning parameters of this network. Based on Table 3, the summation of all network training parameters for classification input data into three classes using Conv-AE can be computed from the sum of

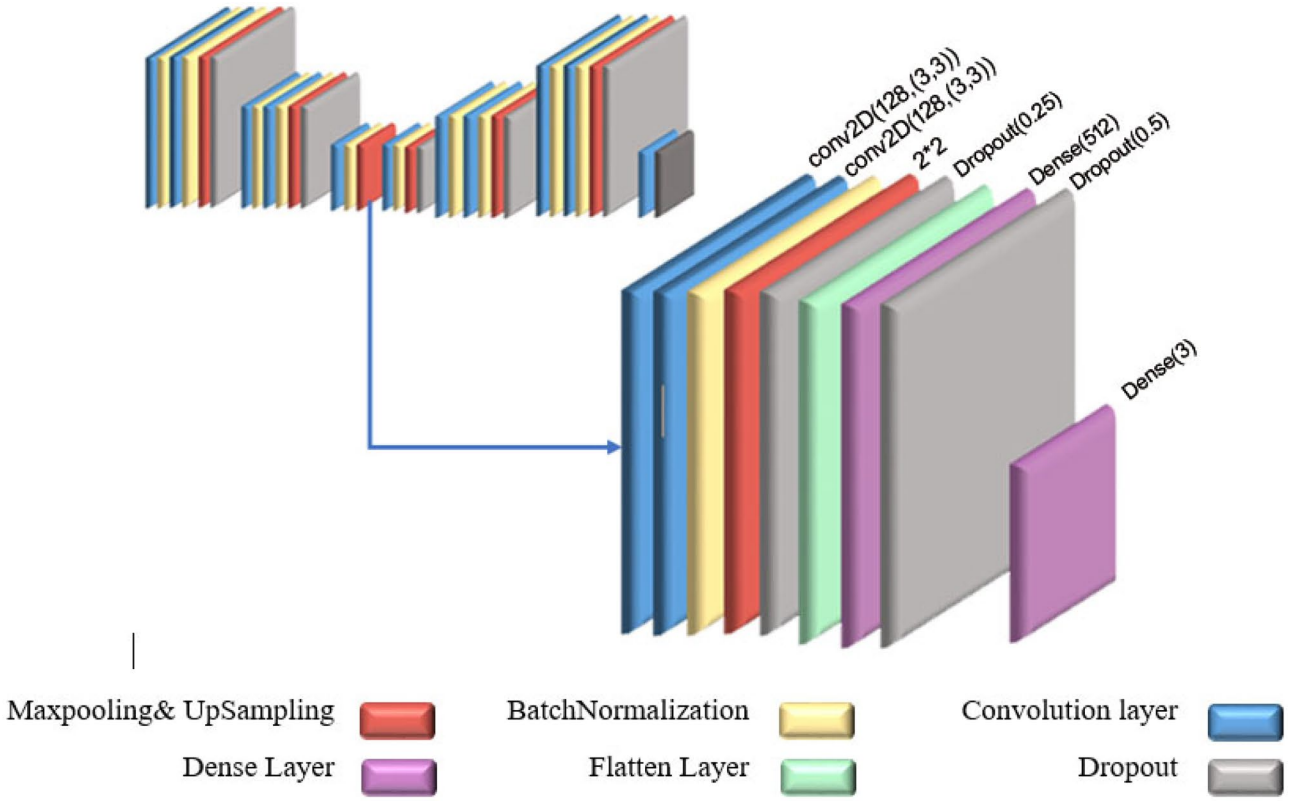

Fig. 11 The suggested architecture of the Conv-AE network classification section

Table 3 Training parameters of Conv-AE network for classification data into 3 classes

Layer (type)	Output shape	Param #
input_1 (Input Layer)	(None, 19, 512, 1)	0
conv2d_1 (conv2D)	(None, 19, 512, 128)	1280
batch_normalization_1 (Batch Normalization)	(None, 19, 512, 128)	512
conv2d_2 (conv2D)	(None, 19, 512, 128)	147,584
batch_normalization_2 (Batch Normalization)	(None, 19, 512, 128)	512
max_pooling2d_1 (MaxPooling2D)	(None, 10, 512, 128)	0
dropout_1 (Dropout)	(None, 10, 256, 128)	0
conv2d_3 (conv2D)	(None, 10, 256, 64)	73,792
batch_normalization_3 (Batch Normalization)	(None, 10, 256, 64)	256
conv2d_4 (conv2D)	(None, 10, 256, 64)	36,928
batch_normalization_4 (Batch Normalization)	(None, 10, 256, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None,5, 128, 64)	0
dropout_2 (Dropout)	(None,5, 128, 64)	0
conv2d_5 (conv2D)	(None,5, 128, 32)	18,464
batch_normalization_5 (Batch Normalization)	(None,5, 128, 32)	128
max_pooling2d_3 (MaxPooling2D)	(None,3, 64, 32)	0
conv8(conv2D)	(None,3, 64, 128)	36,992
conv2d_12 (conv2D)	(None,3, 64, 128)	147,584
batch_normalization_11 (Batch Normalization)	(None,3, 64, 128)	512
max7 (MaxPooling2D)	(None,1, 32, 128)	0
dropout_6 (Dropout)	(None,1, 32, 128)	0
flatten_1 (Flatten)	(None, 4096)	0
dense_1 (Dense)	(None, 512)	2,097,664
dropout_7 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 3)	1539
Total params: 2,564,003 Trainable params: 2,562,915 Non- trainable params: 1088		

Param column data. The sum of parameters is 2,564,003 which 2,562,915 parameters are used for learning and 1088 parameters are for non-learning parameters. The optimization function, learning rate, batch size, number of training cycles, development, and hardware space required for this network are similar to the convolution network discussed earlier. Each cycle duration was 118 s and 11 ms, generally lasting 3 and a half hours of training.

Experiment Results

The primary goal of current research is to classify EEG signals into AD, MCI, and HC classes. The database included 189 recorded EEG signals, of which 90 (37 AD, 37 MCI, and 16 HC data) were used. The signals were divided into 2-s time epochs. As EEG signals are continuous and dynamic data, the signal frequency at each time is important.

Applying continuous-wave transform and Mexican hat function, TFR of the signals was performed. The converted data were divided into training and test data with a specific ratio for convolutional and Conv-AE networks. Accuracy is one of the essential evaluation criteria for classification issues using NNs. The initial data set is usually divided into training data and testing data. The network obtains a training accuracy and a test accuracy in each epoch during network training. Final train accuracy and final test accuracy are the mean of all 100 training accuracy and testing accuracy, respectively. Table 4 shows the obtained results from two networks to classify EEG signals into three classes HC, MCI, and AD.

The classification was well performed by the proposed convolution network with 92% validation accuracy showing a 10% increase compared to that of previous studies.

Table 4 Results from the used networks for classification into 3 classes

	Training accuracy	**Validation accuracy**	**Training loss**	**Validation loss**
CNN	0.9759	0.9270	0.0607	0.2765
AE-CNN	0.8980	0.8926	0.2953	0.3940

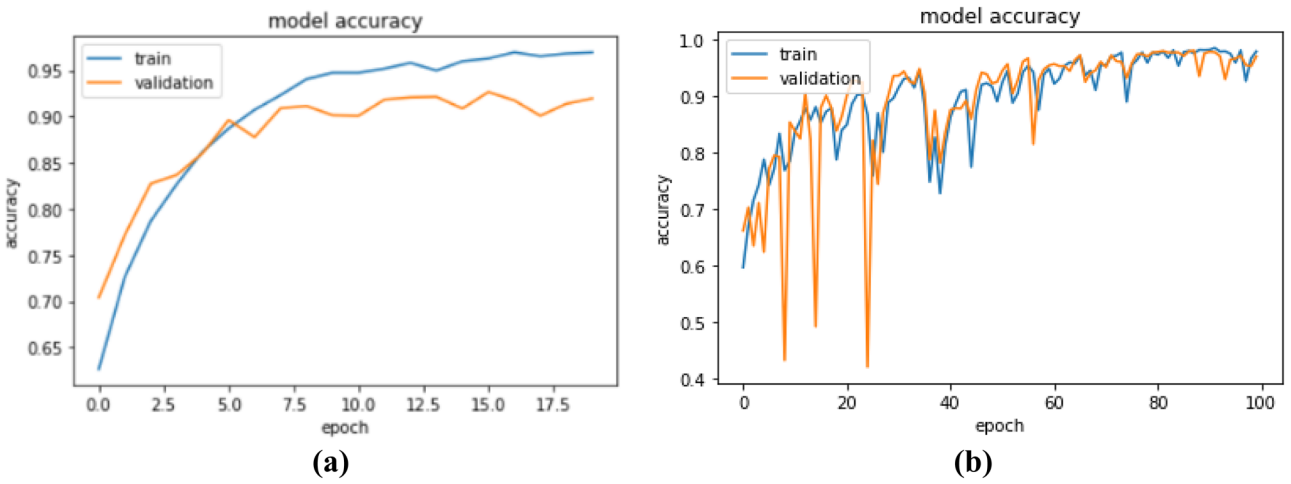

Fig. 12 Model accuracy diagram over 100 epoch: **a** convolutional network and **b** Conv-AE network

Also, compared to similar studies, the classification by the Conv-AE network showed better validation accuracy of 89%. Figure 12 shows the training and testing accuracies over 100 epochs for both convolution and Conv-AE networks. Although the running time for both networks was relatively long, especially the convolution network, significant results were achieved. For both networks, we used several convolution layers and created complex networks. It is justifiable owing to the high volume of data with the aim to increase classification accuracy. Moreover, we sent the data to the network in 20 batch size. It is confirmed concerning the time increased (the fewer the number of batch-size, the more time is needed for training). In comparison with other studies, despite using fewer layers or smaller datasets, networks trained the data in more than 100 training cycles resulting in poorer performance due to the simplicity of the network or the insufficient number of data for in-depth learning. Comparing the two networks used in this study, the convolution network generally operated 3% more accurately than the Conv-AE network. Although the convolution network is simpler than the auto-encoder network, according to what has already been mentioned, it has performed better in extracting features as well as learning and used all the parameters for learning.

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

In addition, further evaluation criteria such as precision, recall, and F1-score were calculated for both developed networks. In the ML field, precision criteria quantify the number of true-positive (TP) predicted classes. It is obtained by dividing the correctly prognosis positive items by the number of positive predicted examples. Precision and recall are often reported together. By combining precision and recall, F1-score has been calculated and used when a single criterion of a model's performance is required. In some studies, recall is called sensitivity and is used to calculate the ratio of retrieved instances among all relevant instances. An accurate classifier model obtained precision and recall, both equal to 1.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$F1 - score = \frac{2 \times (\mathrm{Recall} \times \mathrm{Precision})}{(\mathrm{Recall} + \mathrm{Precision})}$$

Table 5 summarizes the results of these criteria. The precision, recall, and F1-score obtained 70%, 91%, and 79% from CNN and were 70%, 88%, and 77% for AE-CNN, respectively.

The confusion matrix is a table consisting of two columns and two rows in two-class's classification research that reports the number of (TN: true negatives), (FN: false negatives), (TP: true positives), and (FP: false positives). The confusion matrix is not limited to two-class classification. It can be used in multi-class classification as well [95]. Precision, recall, and F1-score were calculated based on this matrix. Figure 13 demonstrates the confusion matrixes of the two proposed networks.

Table 5 Precision, recall, and F1-score obtained from proposed networks for classification into three classes

	Precision	Recall	F1-score
CNN	0.7048	0.9126	0.7957
AE-CNN	0.7098	0.8892	0.7784

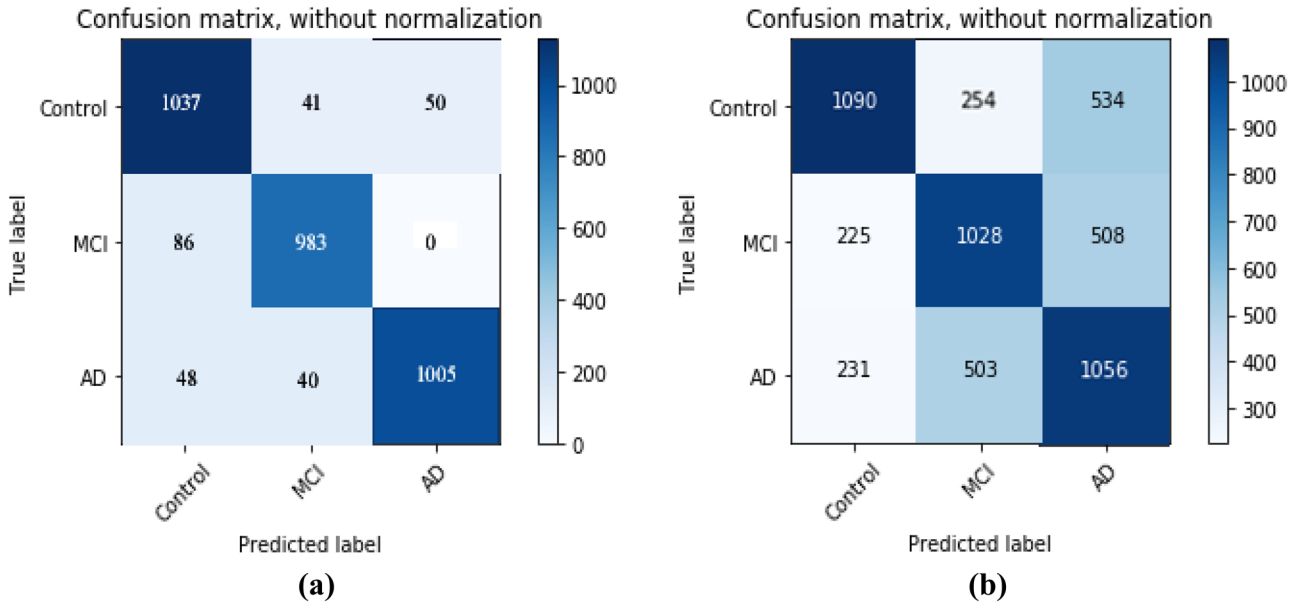

Fig. 13 The Confusion matrix of **a** convolutional network and **b** Conv-AE network

Sensitivity and specificity are two essential criteria that measure the validity of a diagnostic test in comparison with the standard golden test. Receiver operating characteristic (ROC) curve depicts the trade-off between the sensitivity and (1-specificity) in several cut-off points where the diagnostic test is continuous. This is a valuable method for evaluation the performance of a medical test [96]. The area under the ROC plot summarizes the information included in the curve, and if the area was equal to 1, the highest performance of the classifier model has been obtained. Figure 14 demonstrates the Roc plots after training and evaluating networks for (HC: class 0), (MCI: class 1) and (AD: class 2). Cross-validation is a method to evaluate ML and DL classifiers on a limited data set. In this technique, an initial data set is divided into K sections without overlapping. Therefore, the technique usually has been known as k-fold cross-validation. For instance, if the K value is, then the procedure is called fivefold cross-validation. This method is usually used to evaluate the performance of ML methods in dealing with unseen data.

It means a limited data sample is used to evaluate the model efficiency for predicting data not seen during network training. In this research, the value of 5 was chosen for the K parameter. Tables 6 and 7 summarized the result of fivefold cross-validation for CNN and AE-CNN, respectively [97]. The performance of classical ML models such

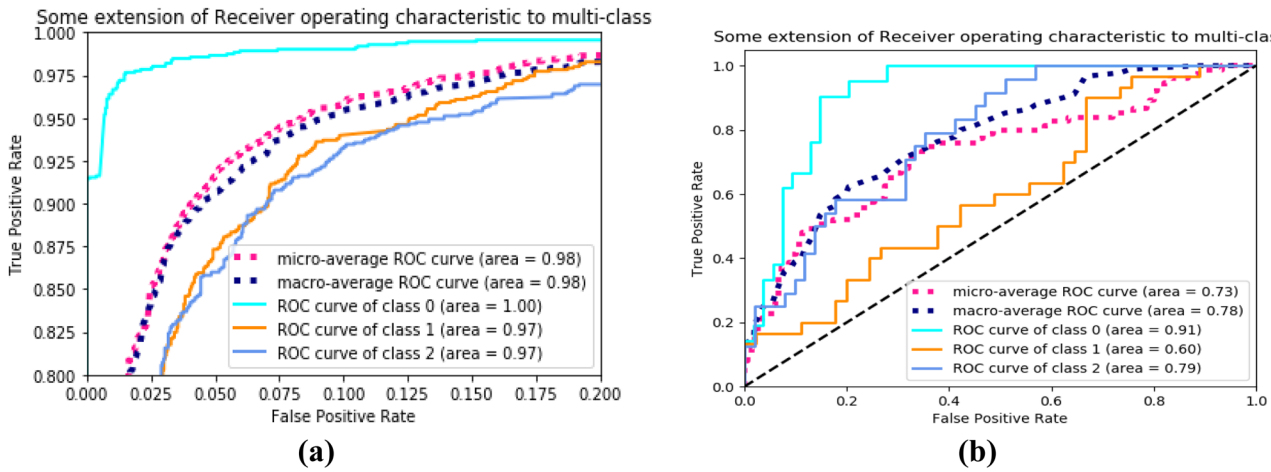

Fig. 14 ROC curve of **a** convolutional network and **b** Conv-AE network

Table 6 fivefold cross-validation of proposed CNN.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Training accuracy	0.9226	0.8989	0.9243	0.9121	0.9019
Validation accuracy	0.8955	0.8787	0.8953	0.8828	0.8823
Training error	0.1861	0.2238	0.1806	0.2035	0.2180
Validation error	0.2694	0.3114	0.2804	0.2990	0.3076

as MLP, SVM, RF, NN, LR, and SGD was compared for classification into three classes. The obtained accuracy was 64% for the random forest, 61% for support vector machine, 54% multilayer for perceptron, 40% for the nearest neighbor, and 36% for random gradient decrease and logical regression. The maximum accuracy obtained from these methods was 64%, which was 28% less than that of the convolution network and 25% less than that of the Conv-AE network.

Figure 15 demonstrates the difference between obtained results. Table 8 summarized the precision, recall, and F1-score for each set of HC, MCI, and AD data. For HC data, random forest achieved maximum accuracy of 84%. Maximum recall 95% and F1-score 85% were obtained by the nearest neighbor and random forest. For MCI data, the highest accuracy, the highest recall, and F1-score were calculated 56% for the SVM and 67% and 60% for the RF, respectively. Also, for AD data, the highest accuracy, recall, and F1-score were estimated for the SVM at 52%, 48%, and 50%, respectively.

In other studies, our data have been investigated using other neural network architectures. Their results are reviewed here. In one study, power spectral density for 19-channel EEG tracing and illustrating the corresponding spectral shape into 2D grayscale images (PSD-images) was used. In the first phase, 117 EEG signals (39 AD signal, 39 MCI signal, and 39 HC) were used to classify AD, MCI, and HC subjects. An especial one dimensional CNN is developed to achieve the related two and three classification studies. It achieves an average accuracy of 89.8% in binary classification and 83.3% in the three-way classification. Another layer was added to the network, but the accuracy was reduced to 78.49% in three-way classification. The maximum accuracy obtained from this study was 83.3%, which was 9% less than that of the convolution network and 6% less than that of the Conv-AE network [52].

Table 7 fivefold cross-validation of proposed AE-CNN.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Training accuracy	0.8922	0.8688	0.8949	0.8895	0.8708
Validation accuracy	0.8616	0.8378	0.8696	0.8573	0.8613
Training error	0.2061	0.2428	0.2002	0.2385	0.2379
Validation error	0.2874	0.3450	0.2912	0.3162	0.3366

In another study, 142 EEG signals contained 63 signals of AD subject, 56 signals of MCI subject, and 23 signals of HC subject were used to classify AD, MCI, and HC. In this study, the preprocessed EEG signals are used as input data to train CNN. At first, each channel of EEG signal is divided into non-overlapping parts of 2 s. Then TFR is applied for each epoch, and the average TFR on all of the epochs is calculated. In this research, CWT with the MHf was chosen as the mother wavelet. The features vector obtained from TFR is convolved with appropriate masks to achieve a high-level representation of features. After that, an auto-encoding network makes an additional packed hidden representation of 10 features. At the end step, the network is trained to classify and distinguish data into AD, MCI, and HC classes. It achieves an average accuracy of 82% in three-way classification. The accuracy obtained from this study was 10% less than that of the convolution network and 7% less than that of the Conv-AE network [42]. Although in this study, the same method was used for pre-processing the signals, our convolutional neural network architecture performed more accurately. Figure 16 shows a comparison of these results.

Discussion

AD is a neurological disorder that causes the disruption of the normal operation of brain functions. This disorder puts a heavy burden on the health system of the patient and the shoulder of the family. The detection of AD disorders at the early stages is necessary for slowing down the progress of the disease. To early detection, several approaches have been investigated in previous studies such as neuroimaging techniques and DL-based methods [98]. In addition, research in neurological disorders generally involves a wide range of data modalities (such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and EEG signals). Also, it depends on other considerations, e.g., the purpose of cognitive assessment, the type of non-invasive, economic, and time-saving approaches [99, 100].

AD patients, due to their condition, cannot stay motionless for a long time during examinations like imaging. EEG is a non-invasive test that can easily be tolerated by them [101]. Therefore, in this paper, an extensive set of EEGs have been used to classify data into three classes for MCI detection and AD prevention. These types of enormous data sets can be analyzed for further disease detection.

Extracting technical knowledge from and gaining practical insights into high-dimensional and sophisticated medical data is the primary challenge in health care and medicine. There is a wide variety of data types generated in modern medicine research, including imaging, signals, and text

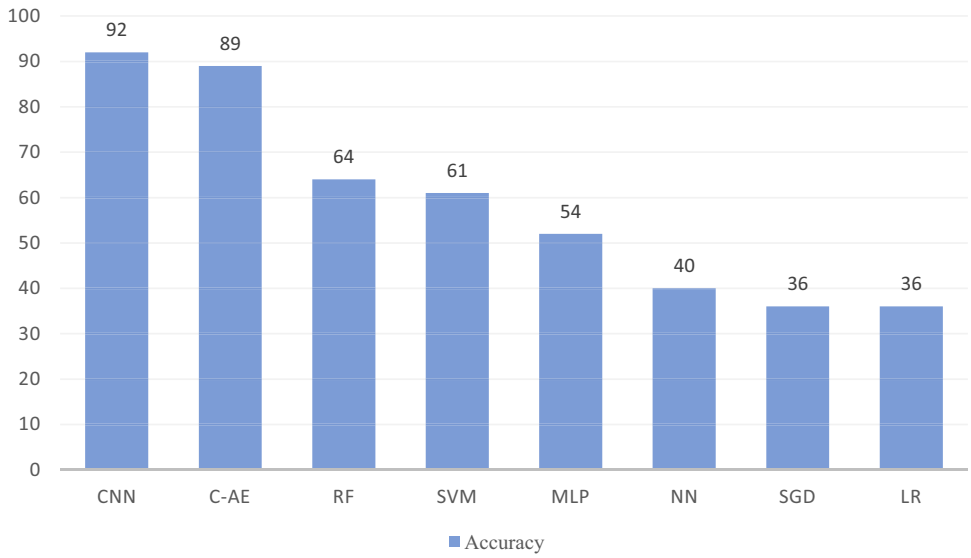

Fig. 15 Comparison of the classification accuracy obtained from DL networks and ML classifiers

which are heterogeneous, complex, and usually unstructured. Conventional data mining approaches are commonly needed to perform feature engineering on these data types to gain robust and useful features and then build prediction models of them. The recent advancements in DNNs provide new practical paradigms to obtain general learning models from the complicated data in this domain [102].

Rapid advances in hardware-based technologies and data acquisition tools have provided scientists and researchers new possibilities for collecting multimodal data such as signals, images, and sequences. These types of data are complex and huge. Their analysis and pattern recognition are a big deal and need sophisticated ML techniques. ANN-based systems and their deep architectures called DNNs have obtained promising results in pattern recognition problems [103]. According to recent studies, DL approaches can be a practical tool for using medical and biomedical data to improve human health. However, there are some limitations in the application of these methods in terms of their understandability for domain experts like physicians, health providers, and scientists. It is suggested to develop explainable and interpretable architectures to bridge human interpretation and deep neural networks. Still, DNNs have not been assessed for a wide range of characteristics that are useful and helpful in medicine and health care, such as their outstanding performance and ability to deal with multi-modality and complex data as well as end-to-end learning. To speed up these experiments, multiple challenges related to the aspects of health care data can be coped in studies within the field of DL. It is essential to improve tools and methods to join DL and health care information [102]. In most studies, DNNs are used for classification, but due to their non-linear nature, these networks are not interpretable and especially difficult to understand. Therefore, they are known as "black box" models. Numerous studies have been conducted on training deep networks with descriptions for their prediction and what they compute. This issue is especially important in healthcare and in the case of care where the issue of human life is raised [104].

The studies show that DL and ML techniques can be used in some medical settings to improve treatment outcomes, such as medical imaging, cancer therapy, and drug delivery. Most recent research concentrates on the prediction, diagnosis, and detection of diseases using AI [105]. DL models serve as valuable and practical tools in the hands of health

Table 8 Precision, recall, and F1-score of ML classifiers for three classes of AD, MCI, and HC

	AD			MCI			HC		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
SVM	0.52	0.48	0.50	0.56	0.50	0.53	0.74	0.86	0.80
RF	0.51	0.38	0.44	0.55	0.67	0.60	0.84	0.86	0.85
MLP	0.48	0.45	0.46	0.52	0.52	0.52	0.74	0.78	0.76
KNN	0.32	0.06	0.10	0.53	0.19	0.28	0.39	0.40	0.39
LR	0.34	0.33	0.34	0.35	0.34	0.35	0.39	0.40	0.39
SGD	0.35	0.37	0.36	0.36	0.35	0.36	0.40	0.38	0.39

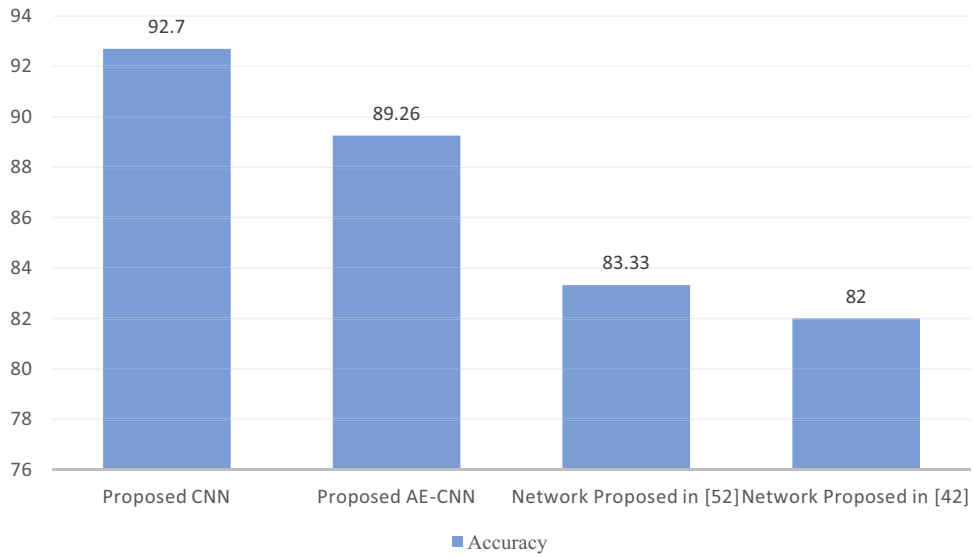

Fig. 16 Comparison of the classification accuracies obtained from proposed DNN and other studies that have used the same data set

providers and medical consultants to diagnose the disease at an early stage. AD as a sever neuron disease that is not curable entails the early detection to avoid its progression. An automated system should be developed to diagnose and classify the disease into multi classes. DL and ML methods have been applied to many medical issues like AD. Several AI techniques have been developed and designed for early detection of AD [106]. As mentioned, the primary purpose of this research is to develop DL models to classify EEG signals into three classes of AD, MCI, and HC and the secondary aim is detecting MCI patients to prevent AD. The development of the most accurate AI-based models for the classification of EEG signals or other medical data is a suggested tool to support physicians in decision making, and to help patients. To this aim, different studies, have used various DL architectures to obtain precise models. In some studies, EEG signals have been classified into three classes [52, 83, 94], and in some other studies, they have been classified into two classes such as AD and MCI, MCI and HC, or AD and HC [42, 78–82, 84–86].

DL techniques have a hierarchical learning process in which a model is capable of learning complex representations from raw data. Several DL architectures are used in research for various applications. Most popular and common models have been determined such as CNN, deep autoencoder (DA), generative adversarial network (GAN), and recurrent neural network (RNN). Based on the investigations, CNN, DA, and GAN are most frequently used for signal processing [107]. Based on the findings of recent studies, CNN and Conv-AE networks were implemented in this study to analyze and classify EEG signals.

Accuracy is one of the main criteria to evaluate DL models, and an accurate model is more reliable to predict the suitable input data class. Compared to previous studies, our study obtained promising results on the same dataset and networks. In [6] and [7], an ANN was suggested to classify EEG signals into two classes of and HC. The obtained accuracy of 90% is 2% less than our proposed CNN network and 1% more than that of our Conv-AE network. Overall, our study achieved better results as our proposed networks classified EEG signals into three classes. In [10], a nonlinear ANN model was developed to classify EEG signals into two classes of AD and MCI subjects in relaxing positions. The accuracy of 92.33% was obtained which is equal to the accuracy of our CNN and 3% more than that of our suggested Conv-AE. Notably, we classify data into three classes, but data in [10] were classified into two classes. In [13], a 3D-CNN with sparse auto-encoder networks was implemented to classify MRI scans from the ADNI database. The accuracy was reported 86.64% for AD vs MCI, 92.11% for MCI vs HC, 95.39% for AD vs HC, and 89.47% for MCI vs AD vs HC. In this study, MRI images were used and the classification was done in three classes resulting in an accuracy of 3% less than that of our suggested CNN model and equal to the accuracy of our Conv-AE model. In [14], classification accuracies of 65.6%, 69.5%, and 91.4%, respectively, were achieved for MCI vs HC, AD vs MCI, and AD vs HC using the ADNI dataset. In [42], a stacked AE method was developed to classify rapidly progressive dementia (RPD), Creutzfeldt-Jakob disease (CJD), HC, and AD subjects. The respective accuracies of 85%, 78%, 85%, and 82% were achieved for the classification of HC vs AD, MCI vs AD, HC vs MCI, and MCI vs AD vs HC groups. The two-class classification of this study was less accurate than our results for both CNN and Conv-AE. In [52], a CNN network was proposed to classify HC vs AD, MCI vs AD, HC vs MCI, and MCI vs AD vs HC subjects with achieved accuracies

of 92.29%, 84.62%, 91.88%, and 83.33%. In comparison, our proposed networks obtained promising results on three-class classification.

ML classifiers were also used widely to classify data and generate predicting models. In [8], an SVM was used to classify AD and HC, resulting in an accuracy of 79.9%. In [12], an SVM model was used to specify types of an EEG dataset, including AD, MCI, and HC. The subjects were in various states of eye-closed relaxing, eye-closed counting, and eye-opened relaxing with respective classification accuracies of 79.2%, 85.4%, and 83.3%. It is evident that DL models outperform ML methods in classifying EEG signals or MRI scans into a binary or triple class classification of AD, MCI, and HC.

Conclusions

In this work, two DL networks, including CNN and Conv-AE NN, were proposed for differentiating between HC, MCI, and AD subjects using EEG data. The classification was performed on 19-channel EEG signals recorded from 90 subjects, including 37 AD patients, 37 MCI patients, and 16 HC people. Each group signals were recorded in 5-m durations, though some lasted more, especially in the control group. At first, the signals were divided into 2-s epochs based on a sampling frequency of 256 Hz. After division, a total number of 16,449 data samples were resulted in the same size. Each data sample was an array with 19 rows and 512 columns. TFR was used to extract desirable features from EEG signals in both time and frequency dimensions. For this representation, MHf was proposed. In this research, the proposed CNN generally performed more accurately than the Conv-AE neural network for classifying EEG signals. Our proposed CNN improved accuracy by about 10% more than other studies that used our dataset. Also, compared to other studies, the performance of Conv-AE was promising. For future work, firstly, the use of other NN and other datasets is suggested. Secondly, the models can be used as a computer-aided diagnosis (CAD) system to help physicians in the detection of AD or MCI or HC person. Lastly, there are specific patterns in the TFR of signals of these three groups (Figs. 3–5), which can be considered as a tool for visual recognition. For their non-linear nature, DL networks are difficult to explain and understand and are known as "black box" models. For future AI-based research in medicine and healthcare, it is suggested to develop a novel deep neural network architecture that automatically explains its reason for each output prediction. In this case, physicians and health providers are convinced to use them and trust them for decision-making.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12559-022-10033-3.

Acknowledgements The authors thank IRCCS Centro Neurolesi of Messina (ITALY) for having provided the dataset.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare no competing interests.

References

1. Dementia. World Heal Organization 2021. https://www.who.int/en/news-room/fact-sheets/detail/dementia . (Accessed 2 Sept 2021).
2. Santos CY, Snyder PJ, Wu W-C, Zhang M, Echeverria A, Alber J. Pathophysiologic relationship between Alzheimer's disease, cerebrovascular disease, and cardiovascular risk: A review and synthesis. Alzheimer's Dement Diagnosis, Assess Dis Monit. 2017;7:69–87. https://doi.org/10.1016/j.dadm.2017.01.005.
3. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med. 2014;275:214–28. https://doi.org/10.1111/joim.12190.
4. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: Mild cognitive impairment. Neurol. 2018;90:126 LP – 135. https://doi.org/10.1212/WNL.0000000000004826.
5. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. J Neural Eng. 2018;15: 031005. https://doi.org/10.1088/1741-2552/aab2f2.
6. Bostanov V. BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Trans Biomed Eng. 2004;51:1057–61. https://doi.org/10.1109/TBME.2004.826702.
7. Garrett D, Peterson DA, Anderson CW, Thaut MH. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans Neural Syst Rehabil Eng. 2003;11:141–4. https://doi.org/10.1109/TNSRE.2003.814441.
8. Garcia GN, Ebrahimi T, Vesin J. Support vector EEG classification in the Fourier and time-frequency correlation domains. First Int IEEE EMBS Conf Neural Eng 2003 Conf Proceedings. 2003;591–4. https://doi.org/10.1109/CNE.2003.1196897.
9. Puri D, Nalbalwar S, Nandgaonkar A, Wagh A. EEG-based diagnosis of Alzheimer's disease using Kolmogorov complexity BT — applied information processing systems. In: Iyer B, Ghosh D, Balas VE, editors., Singapore: Springer, Singapore. 2022;157–65.
10. Müller KR, Krauledat M, Dornhege G, Curio G, Blankertz B. Machine learning techniques for brain-computer interfaces. Biomed Tech. 2004;49. https://doi.org/10.13109/9783666351419.11.
11. Rakotomamonjy A, Guigue V, Mallet G, Alvarado V. Ensemble of SVMs for improving brain computer interface P300 speller performances BT — artificial neural networks: biological inspirations — ICANN 2005. In: Duch W, Kacprzyk J, Oja E, Zadrożny S, editors., Berlin, Heidelberg: Springer Berlin Heidelberg. 2005;45–50.
12. Kaper M, Meinicke P, Grossekathoefer U, Lingner T, Ritter H. BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. IEEE Trans Biomed Eng. 2004;51:1073–6. https://doi.org/10.1109/TBME.2004.826698.

13. Palaniappan R. Brain computer interface design using band powers extracted during mental tasks. Conf Proceedings 2nd Int IEEE EMBS Conf Neural Eng. 2005;321–4. https://doi.org/10.1109/CNE.2005.1419622.
14. Anderson CW, Sijercic Z. Classification of EEG signals from four subjects during five mental tasks. InSolving engineering problems with neural networks: proceedings of the conference on engineering applications in neural networks (EANN'96) 1996 Jun (pp. 407-414). Turkey.
15. Haselsteiner E, Pfurtscheller G. Using time-dependent neural networks for EEG classification. IEEE Trans Rehabil Eng. 2000;8:457–63. https://doi.org/10.1109/86.895948.
16. Chiappa S, Bengio S. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems in Proc Eur Symp Artif Neural Netw (ESANN), 2004 (pp. 199–204).
17. Balakrishnan D, Puthusserypady S. Multilayer perceptrons for the classification of brain computer interface data. InProceedings of the IEEE 31st Annual Northeast Bioengineering Conference, 2005. 2005 Apr 2 (pp. 118-119). IEEE. https://doi.org/10.1109/NEBC.2005.1431953.
18. Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell. 2000;22:4–37. https://doi.org/10.1109/34.824819.
19. Congedo M, Lotte F, Lécuyer A. Classification of movement intention by spatially filtered electromagnetic inverse solutions. Phys Med Biol. 2006;51:1971–89. https://doi.org/10.1088/0031-9155/51/8/002.
20. Millán JD, Mourino J, Babiloni F, Cincotti F, Varsta M, Heikkonen J. Local neural classifier for EEG-based recognition of mental tasks. InProceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium 2000 Jul 27 (Vol. 3, pp. 632-636). IEEE. https://doi.org/10.1109/IJCNN.2000.861393.
21. Solhjoo S, Moradi MH. Mental task recognition: a comparison between some of classification methods. Proceedings of BIOSIGNAL 2004 International EURASIP Conference. 2004.
22. Cincotti F, Scipione A, Timperi A, Mattia D, Marciani AG, Millan J, et al. Comparison of different feature classifiers for brain computer interfaces. First Int IEEE EMBS Conf Neural Eng 2003 Conf Proceedings. 2003;645–7. https://doi.org/10.1109/CNE.2003.1196911.
23. Kohonen T. Essentials of the self-organizing map. Neural Netw. 2013;37:52–65. https://doi.org/10.1016/j.neunet.2012.09.018.
24. Pfurtscheller G, Flotzinger D, Kalcher J. Brain-computer interface—a new communication device for handicapped persons. J Microcomput Appl. 1993;16:293–9. https://doi.org/10.1006/jmca.1993.1030.
25. Lo Giudice M, Varone G, Ieracitano C, Mammone N, Tripodi GG, Ferlazzo E, et al. Permutation entropy-based interpretability of convolutional neural network models for interictal EEG discrimination of subjects with epileptic seizures vs. psychogenic non-epileptic seizures. Entropy. 2022;24. https://doi.org/10.3390/e24010102.
26. Bourbakis NG, Michalopoulos K, Antonakakis M, Zervakis M. A new multi-resolution approach to EEG brain modeling using local-global graphs and stochastic petri-nets. Int J Neural Syst. 2022;2250006. https://doi.org/10.1142/S012906572250006X.
27. Hoya T, Hori G, Bakardjian H, Nishimura T, Suzuki T, Miyawaki Y, et al. Classification of single trial EEG signals by a combined principal + independent component analysis and probabilistic neural network approach 2003, In Proc. ICA2003 (Vol. 197).
28. Kostov A, Polak M. Parallel man-machine training in development of EEG-based cursor control. IEEE Trans Rehabil Eng. 2000;8:203–5. https://doi.org/10.1109/86.847816.
29. Koutroumbas STK. Pattern recognition. 4th ed. Athens: Academic Press; 2008.
30. Nigam VP, Graupe D. A neural-network-based detection of epilepsy. Neurol Res. 2004;26:55–60. https://doi.org/10.1179/016164104773026534.
31. Morabito FC, Campolo M, Ieracitano C, Ebadi JM, Bonanno L, Bramanti A, et al. 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). IEEE. 2016.
32. Li F, Tran L, Thung K, Ji S, Shen D, Li J. A robust deep model for improved classification of AD/MCI patients. IEEE J Biomed Heal Informatics. 2015;19:1610–6. https://doi.org/10.1109/JBHI.2015.2429556.
33. Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification BT — medical image computing and computer-assisted intervention – MICCAI 2013. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, editors., Berlin, Heidelberg: Springer Berlin Heidelberg. 2013;583–90.
34. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. IEEE Int Conf Image Process. 2016;2016:126–30. https://doi.org/10.1109/ICIP.2016.7532332.
35. Suk H-I, Lee S-W, Shen D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage. 2014;101:569–82. https://doi.org/10.1016/j.neuroimage.2014.06.077.
36. Suk H-I, Lee S-W, Shen D, Initiative TADN. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct Funct. 2015;220:841–59. https://doi.org/10.1007/s00429-013-0687-3.
37. Franciotti R, Moretti DV, Benussi A, Ferri L, Russo M, Carrarini C, et al. Cortical network modularity changes along the course of frontotemporal and Alzheimer's dementing diseases. Neurobiol Aging. 2022;110:37–46. https://doi.org/10.1016/j.neurobiolaging.2021.10.016.
38. Li R, Zhang W, Suk H-I, Wang L, Li J, Shen D, et al. Deep learning based imaging data completion for improved brain disease diagnosis BT — medical image computing and computer-assisted intervention — MICCAI 2014. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors., Cham: Springer International Publishing. 2014;305–12.
39. Liu F, Shen C. Learning Deep Convolutional features for MRI based Alzheimer's disease classification. 2014.
40. Cao J, Zhao Y, Shan X, Wei H, Guo Y, Chen L, et al. Brain functional and effective connectivity based on electroencephalography recordings: A review. Hum Brain Mapp. 2022;43:860–79. https://doi.org/10.1002/hbm.25683.
41. Ieracitano C, Mammone N, Bramanti A, Hussain A, Morabito FC. A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. Neurocomputing. 2019;323:96–107. https://doi.org/10.1016/j.neucom.2018.09.071.
42. Barreto GA, Frota RA, de Medeiros FNS, On the classification of mental tasks: a performance comparison of neural and statistical approaches. Proc. 14th IEEE Signal Process. Soc Work Mach Learn Signal Process. 2004;2004(2004):529–38. https://doi.org/10.1109/MLSP.2004.1423016.
43. Deepa N, Chokkalingam SP. Optimization of VGG16 utilizing the arithmetic optimization algorithm for early detection of Alzheimer's disease. Biomed Signal Process Control. 2022;74:103455. https://doi.org/10.1016/j.bspc.2021.103455.
44. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77:257–86. https://doi.org/10.1109/5.18626.
45. Obermaier B, Neuper C, Guger C, Pfurtscheller G. Information transfer rate in a five-classes brain-computer interface. IEEE

Trans Neural Syst Rehabil Eng. 2001;9:283–8. https://doi.org/10.1109/7333.948456.
46. Solhjoo S, Nasrabadi AM, Golpayegani MRH, Classification of chaotic signals using HMM classifiers:EEG-based mental task classification. 13th Eur. Signal Process Conf. 2005;2005:1–4.
47. Seixas FL, Zadrozny B, Laks J, Conci A, Muchaluat Saade DC. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Comput Biol Med. 2014;51:140–58. https://doi.org/10.1016/j.compbiomed.2014.04.010.
48. Mahendran N, PM DRV. A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease. Comput Biol Med. 2022;141:105056. https://doi.org/10.1016/j.compbiomed.2021.105056.
49. Liang C, Lao H, Wei T, Zhang X. Alzheimer's disease classification from hippocampal atrophy based on PCANet-BLS. Multimed Tools Appl. 2022. https://doi.org/10.1007/s11042-022-12228-0.
50. Dhanusha C, Senthil Kumar AV, Jagadamba G, Musirin IB. Evolving Chaotic Shuffled Frog Leaping Memetic Metaheuristic Model-Based Feature Subset Selection for Alzheimer's Disease Detection. InSustainable Communication Networks and Application 2022 (pp. 679-692). Springer, Singapore.
51. Schoonhoven DN, Briels CT, Hillebrand A, Scheltens P, Stam CJ, Gouw AA. Sensitive and reproducible MEG resting-state metrics of functional connectivity in Alzheimer's disease. Alzheimers Res Ther. 2022;14:38. https://doi.org/10.1186/s13195-022-00970-4.
52. Keogh E, Mueen A. Curse of dimensionality BT — encyclopedia of machine learning and data mining. In: Sammut C, Webb GI, editors., Boston, MA: Springer US. 2017;314–5. https://doi.org/10.1007/978-1-4899-7687-1_192.
53. Hoffmann U, Garcia G, Vesin J-., Diserens K, Ebrahimi T. A boosting approach to P300 detection with application to brain-computer interfaces. Conf Proceedings 2nd Int IEEE EMBS Conf Neural Eng. 2005;97–100. https://doi.org/10.1109/CNE.2005.1419562.
54. ISSPA 2003. Seventh International Symposium on Signal Processing and its Applications. Proceedings (Cat. No.03EX714). Seventh Int. Symp. Signal Process. Its Appl. 2003. Proceedings. 2003;1. https://doi.org/10.1109/ISSPA.2003.1224625.
55. Subasi A, Erçelebi E. Classification of EEG signals using neural network and logistic regression. Comput Methods Programs Biomed. 2005;78:87–99. https://doi.org/10.1016/j.cmpb.2004.10.009.
56. Alkan A, Koklukaya E, Subasi A. Automatic seizure detection in EEG using logistic regression and artificial neural network. J Neurosci Methods. 2005;148:167–76. https://doi.org/10.1016/j.jneumeth.2005.04.009.
57. Rodrigues P, Teixeira J, Homero R, Poza J, Carreres A. Classification of Alzheimer's electroencephalograms using artificial neural networks and logistic regression. In Japan-Portugal Nano-Biomedical Engineering Symposium. 2011;33–34.
58. Subasi A. Epileptic seizure detection using dynamic wavelet network. Expert Syst Appl. 2005;29:343–55. https://doi.org/10.1016/j.eswa.2005.04.007.
59. Yu H, Wu S, Dauwels J. Efficient variational Bayes learning of graphical models with smooth structural changes. IEEE Trans Pattern Anal Mach Intell. 2022;1. https://doi.org/10.1109/TPAMI.2022.3140886.
60. Ocak H. Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. Signal Process. 2008;88:1858–67. https://doi.org/10.1016/j.sigpro.2008.01.026.
61. Guo L, Rivero D, Seoane JA, Pazos A. Classification of EEG signals using relative wavelet energy and artificial neural networks. Proc. First ACM/SIGEVO Summit Genet Evol Comput. New York, NY, USA: Association for Computing Machinery. 2009;177–184. https://doi.org/10.1145/1543834.1543860.
62. Subasi A. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl. 2007;32:1084–93. https://doi.org/10.1016/j.eswa.2006.02.005.
63. Polat K, Güneş S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Appl Math Comput. 2007;187:1017–26. https://doi.org/10.1016/j.amc.2006.09.022.
64. Zhang Y-D, Wang S, Dong Z. Classification of Alzheimer disease based on structural magnetic resonance imaging by Kernel support vector machine decision tree. Prog Electromagn Res. 2014;144:185–91. https://doi.org/10.2528/PIER13121310.
65. Chudhey AS, Jindal H, Vats A, Varma S. An autonomous dementia prediction method using various machine learning models BT — advances in data and information sciences. In: Tiwari S, Trivedi MC, Kolhe ML, Mishra KK, Singh BK, editors., Singapore: Springer Singapore. 2022;283–96.
66. Javaid H, Kumarnsit E, Chatpun S. Age-related alterations in EEG network connectivity in healthy aging. Brain Sci. 2022;12:218. https://doi.org/10.3390/brainsci12020218.
67. Mirzaei G, Adeli H. Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia. Biomed Signal Process Control. 2022;72: 103293. https://doi.org/10.1016/j.bspc.2021.103293.
68. Ding Y, Chu Y, Liu M, Ling Z, Wang S, Li X, et al. Fully automated discrimination of Alzheimer's disease using resting-state electroencephalography signals. Quant Imaging Med Surg. 2022;12:1063–78. https://doi.org/10.21037/qims-21-430.
69. Caza-Szoka M, Massicotte D. Windowing compensation in Fourier based surrogate analysis and application to EEG signal classification. IEEE Trans Instrum Meas. 2022;1. https://doi.org/10.1109/TIM.2022.3149325.
70. Prado P, Birba A, Cruzat J, Santamaría-García H, Parra M, Moguilner S, et al. Dementia ConnEEGtome: towards multicentric harmonization of EEG connectivity in neurodegeneration. Int J Psychophysiol. 2022;172:24–38. https://doi.org/10.1016/j.ijpsycho.2021.12.008.
71. Saengmolee W, Chaisaen R, Autthasan P, Rungsilp C, Sa-ih N, Cheaha D, et al. Consumer-grade brain measuring sensor in people with long-term Kratom consumption. IEEE Sens J. 2022;1. https://doi.org/10.1109/JSEN.2022.3147207.
72. Sivasangari A, Sonti K, Kanmani GP, Sindhu, Deepa D. Chapter 12 - EEG-based computer-aided diagnosis of autism spectrum disorder. In: Zhang Y-D, Sangaiah AKBT-CS and SP in IP, editors. Cogn Data Sci Sustain Comput Academic Press. 2022;277–92. https://doi.org/10.1016/B978-0-12-824410-4.00010-6.
73. Singh K, Malhotra J. Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features. Complex Intell Syst. 2022. https://doi.org/10.1007/s40747-021-00627-z.
74. Pritchard WS, Duke DW, Coburn KL, Moore NC, Tucker KA, Jann MW, et al. EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures. Electroencephalogr Clin Neurophysiol. 1994;91:118–30. https://doi.org/10.1016/0013-4694(94)90033-7.
75. Trambaiolli LR, Lorena AC, Fraga FJ, Kanda PAM, Anghinah R, Nitrini R. Improving Alzheimer's disease diagnosis with machine learning techniques. Clin EEG Neurosci. 2011;42:160–5. https://doi.org/10.1177/155005941104200304.
76. Huang C, Wahlund L-O, Dierks T, Julin P, Winblad B, Jelic V. Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. Clin Neurophysiol. 2000;111:1961–7. https://doi.org/10.1016/S1388-2457(00)00454-5.

77. Buscema M, Rossini P, Babiloni C, Grossi E. The IFAST model, a novel parallel nonlinear EEG analysis technique, distinguishes mild cognitive impairment and Alzheimer's disease patients with high degree of accuracy. Artif Intell Med. 2007;40:127–41. https://doi.org/10.1016/j.artmed.2007.02.006.
78. Rossini PM, Buscema M, Capriotti M, Grossi E, Rodriguez G, Del Percio C, et al. Is it possible to automatically distinguish resting EEG data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy?. Clin Neurophysiol. 2008;119:1534–45. https://doi.org/10.1016/j.clinph.2008.03.026.
79. McBride JC, Zhao X, Munro NB, Smith CD, Jicha GA, Hively L, et al. Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. Comput Methods Programs Biomed. 2014;114:153–63. https://doi.org/10.1016/j.cmpb.2014.01.019.
80. Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. 2015. arXiv preprint arXiv:1502.02506.
81. Aderghal K, Benois-Pineau J, Afdel K, Gwenaëlle C. FuseMe: Classification of SMRI Images by Fusion of Deep CNNs in 2D+ε Projections. Proc. 15th Int. Work. Content-Based Multimed. Index., New York, NY, USA: Association for Computing Machinery. 2017. https://doi.org/10.1145/3095713.3095749.
82. Sarraf S, Tofighi G. Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. 2016. arXiv preprint arXiv:1603.08631.
83. Morabito FC, Campolo M, Mammone N, Versaci M, Franceschetti S, Tagliavini F, et al. Deep learning representation from electroencephalography of early-stage Creutzfeldt-Jakob disease and features for differentiation from rapidly progressive dementia. Int J Neural Syst. 2016;27:1650039. https://doi.org/10.1142/S0129065716500398.
84. Lam Q, Nguyen L, Nguyen K. Build control command set based on EEG signals via clustering algorithm and multi-layer neural betwork. J Commun. 2018;406–11. https://doi.org/10.12720/jcm.13.7.406-411.
85. Bajaj N. Wavelets for EEG Analysis, in S. Mohammady (ed.), Wavelet theory, IntechOpen, London. 2020. https://doi.org/10.5772/intechopen.94398.
86. Dinç E, Baleanu D. Application of Haar and Mexican hat wavelets to double divisor-ratio spectra for the multicomponent determination of ascorbic acid, acetylsalicylic acid and paracetamol in effervescent tablets. J Braz Chem Soc. 2008;19:434–44. https://doi.org/10.1590/S0103-50532008000300010.
87. Schiff SJ, Aldroubi A, Unser M, Sato S. Fast wavelet transformation of EEG. Electroencephalogr Clin Neurophysiol. 1994;91:442–55. https://doi.org/10.1016/0013-4694(94)90165-1.
88. Taunk K, De S, Verma S, Swetapadma A. A brief review of nearest neighbor algorithm for learning and classification 2019. Int Conf Intell Comput Control Syst. 2019;1255–60. https://doi.org/10.1109/ICCS45141.2019.9065747.
89. Wang Y. The Ricker wavelet and the Lambert W function. Geophys J Int. 2015;200:111–5. https://doi.org/10.1093/gji/ggu384.
90. Saha P, Mukherjee D, Singh PK, Ahmadian A, Ferrara M, Sarkar R, et al. graph neural network based model for detecting COVID-19 from CT scans and X-rays of chest. Sci Rep. 2021;11:8304. https://doi.org/10.1038/s41598-021-87523-1.
91. Fouladi S, Ebadi MJ, Safaei AA, Bajuri MY, Ahmadian A. Efficient deep neural networks for classification of COVID-19 based on CT images: Virtualization via software defined radio. Comput Commun. 2021;176:234–48. https://doi.org/10.1016/j.comcom.2021.06.011.
92. Lee W-Y, Park S-M, Sim K-B. Optimal hyper parameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm. Optik (Stuttg). 2018;172:359–67. https://doi.org/10.1016/j.ijleo.2018.07.044.
93. Guo Y, Li J-Y, Zhan Z-H. Efficient hyper parameter optimization for convolution neural networks in deep learning: a distributed particle swarm optimization approach. Cybern Syst. 2021;52:36–57. https://doi.org/10.1080/01969722.2020.1827797.
94. Wen T, Zhang Z. Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. IEEE Access. 2018;6:25399–410. https://doi.org/10.1109/ACCESS.2018.2833746.
95. Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ. 1997;62:77–89. https://doi.org/10.1016/S0034-4257(97)00083-7.
96. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr. 2011;48:277–87. https://doi.org/10.1007/s13312-011-0055-4.
97. Brownlee J. A Gentle introduction to k-fold cross-validation. Mach Learn Mastery. 2018. https://machinelearningmastery.com/k-fold-cross-validation/.
98. Noor MBT, Zenia NZ, Kaiser MS, Mamun SAl, Mahmud M. Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. Brain Inform. 2020;7:11. https://doi.org/10.1186/s40708-020-00112-2.
99. Chen T, Shang C, Su P, Shen Y, Mahmud M, Moodley R, et al. Assessing significance of cognitive assessments for diagnosing Alzheimer's disease with fuzzy-rough feature selection. In UK Workshop on Computational Intelligence. Springer, Cham. 2021;450–462. https://doi.org/10.1007/978-3-030-87094-2_40.
100. Ruiz J, Mahmud M, Modasshir M, Kaiser MS, Initiative for 3D DenseNet ensemble in 4-way classification of Alzheimer's disease. In International Conference on Brain Informatics. Springer, Cham. 2020;85–96. https://doi.org/10.1007/978-3-030-59277-6_8.
101. Lo Giudice P, Mammone N, Morabito F, Pizzimenti R, Ursino D, Virgili L. Leveraging network analysis to support experts in their analyses of subjects with MCI and AD. Med Biol Eng Comput. 2019;57. https://doi.org/10.1007/s11517-019-02004-y.
102. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform. 2018;19:1236–46. https://doi.org/10.1093/bib/bbx044.
103. Mahmud M, Kaiser MS, McGinnity TM, Hussain A. Deep Learning in Mining Biological Data. Cognit Comput. 2021;13:1–33. https://doi.org/10.1007/s12559-020-09773-x.
104. Li O, Liu H, Chen C, Rudin C. Deep learning for case-based reasoning through prototypes: a neural network that explains its predictions. In Proceedings of the AAAI Conference on Artificial Intelligence. 2018;32(1).
105. Toh C, Brody JP. Applications of machine learning in healthcare, in T. Y. Kheng (ed.), Smart manufacturing — when artificial intelligence meets the Internet of Things, IntechOpen, London, 2021. https://doi.org/10.5772/intechopen.92297.
106. Ghazal TM, Abbas S, Munir S, Khan MA, Ahmad M, Issa GF, et al. Alzheimer disease detection empowered with transfer learning. Comput Mater Contin. 2022;70(3):5005–5019. https://doi.org/10.32604/cmc.2022.020866.
107. Mahmud M, Kaiser MS, Hussain A, Vassanelli S. Applications of deep learning and reinforcement learning to biological data. IEEE Trans Neural Networks Learn Syst. 2018;29(6):2063–79. https://doi.org/10.1109/TNNLS.2018.2790388.