Discrete Wavelet Packet Transform (Dwpt) Based Feature Selection And Classification Using Optimization Based Deep Convolutional Neural Network (Dcnn) And Long Short-Term Memory (Lstm) For Alzheimer's Disease Detection In Eeg

Rashmi R Nath[1], Dr. S. Prabhu[2]

[1]Research Scholar, PG and Research Department, Department of Computer Science, Park's College Tirupur, Tamilnadu , India
[2]Research Supervisor, Department of Computer Science, Park's, College, Tiruppur, Tamilnadu, India

Received: 15.07.2024 Revised: 10.08.2024 Accepted: 06.09.2024

ABSTRACT
Alzheimer's disease (AD) stands out as one of the main causes of dementia worldwide and it represents around 65% of all dementia cases, affecting mainly elderly people. AD is composed of three evolutionary stages: Mild Cognitive Impairment (MCI), Mild and Moderate AD (ADM) and Advanced AD (ADA). It is crucial to create a tool for assisting AD diagnosis in its early stages with the aim of halting the disease progression.The main purpose of this study is to develop a system with the ability of differentiate each disease stage by means of Electroencephalographic Signals (EEG). Thereby, an EEG nonlinear multi-band analysis by Wavelet Packet was performed enabling to extract several features from each study group which are not processed in existing works. Classic Machine Learning (ML) and Deep Learning (DL) methods have been used for data classification per EEG channel. In this research work, initially, disturbances and interference from the EEG dataset are eliminated using a Band-Pass Elliptic Digital Filter (BEF). Then Variational mode decomposition (VMD) was employed to decompose the filtered signals. VMD has excellent noise resistance, better decomposing performance, and stability. Also this paper proposes a new method called Discrete Wavelet Packet Transform (DWPT) processed on a nonlinear multiband analysis of the filtered EEG signals. Hence both the statistic features and the nonlinear features are considered to improve diagnostic performance. The improved EEG was categorized using a Combination of a deep convolutional neural network (DCNN) and long short-term memory (LSTM) called DCNN-LSTM. Here, the ideal DCNN hyperparameter is determined using the Adaptive clonal selection algorithm (ACSA), which might increase the classifier's accuracy. Finally, the sensitiveness, accuracy, quality of diagnosis, and the area beneath the receiver operating characteristic (ROC) curve were additionally computed in order to compare and evaluate the performance of the various recommended approaches.

Keywords: Alzheimer's disease, average energy, discrete wavelet transform, electroencephalogram, Discrete Wavelet Packet Transform (DWPT), Band-Pass Elliptic Digital Filter

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder in which electrical activities of the brain become slower than normal people [1,2]. It is the first cause of dementia and its main symptom is rapid impairment of cognitive functions [3,4]. It is most commonly seen in people over 65 years and its incidence rate increases exponentially with age [5]. The number of people with neurocognitive disorders is about 50 million and is projected to exceed 100 million by 2050 [6]. AD includes three stages of mild AD (AD1), moderate AD (AD2), and severe AD, respectively. During these stages, the patient loses his/her cognitive abilities such as memory and thinking, gradually. Using treatments for people who are in the early stages of AD can greatly prevent the progression of disease and delay the onset of dementia. Therefore, early diagnosis of AD is very important. In many studies, the term mild cognitive impairment (MCI) has been used to categorize people who have many AD symptoms, but these symptoms are not severe enough to be called AD [7].

The treatment of neurological damages caused by AD is very difficult. Early detection of AD allows early treatment and thus delays the onset of AD symptoms [8]. Therefore, early detection of AD is of particular importance and can reduce the incidence rate of this disease as well as delay in disease progression. In the last decades, different neurophysiological techniques have been used for diagnosing patients with

AD. Electroencephalography (EEG) is a non-invasive, low-cost, high-temporal-resolution, and portable technique that reflects the electrical activity the brain. The EEG signal can usefully describe the brain dynamics in neurological disorders [9].

Therefore, early diagnosis is crucial in medical intervention to maintain daily functioning longer, reduce brain damage, and give the patient time to make plans for the future [10]. Different neuroimaging techniques, namely positron emission tomography (PET), EEG, magnetic resonance imaging (MRI), and single photon emission computed tomography (SPECT) are used for the diagnosis of AD [11]. EEG records electrical waves produced by nerve cells in the brain. EEG offers superior millisecond-level temporal resolution in comparison to other neuroimaging techniques. Recording the EEG at rest is advantageous when examining AD patients as it requires little cooperation and is not stressful [12]. In addition, EEG is a relatively easy, inexpensive and widely available, portable and noninvasive method. Because of these advantages, EEG is more preferred [13].

Early diagnosis of diseases by EEG provides an opportunity for early treatment. It is therefore particularly promising. In the relevant literature, there are limited studies investigating the connection between EEG signals and Alzheimer's diagnosis. Amezquita-Sanchez et al. (2019) [14] suggested a model for AD and mild cognitive impairment classification tasks. They used enhanced probabilistic neural network (EPNN) to automatic classify and MUSIC-EWT algorithm to decompose the EEG signals. The accuracy of the model is 90.3%. The convolutional neural network (CNN), linear discriminant analysis (LDA), kNN, and support vector machine (SVM) approaches in diagnosis of AD using EEG signal. Using CNN approach, the accuracy of AD was 89.1% and the accuracy of the healthy population was 75%. Early diagnosis and treatment can slow down the progression of the disease. Therefore, there is a need to develop easy and applicable methods and it is of critical importance.

The contribution of the research is discussed as follows

Step 1: the disturbances and interference from the EEG dataset are eliminated using a band-pass elliptic digital filter (BEF).

Step 2: Variational mode decomposition (VMD) was employed to decompose the filtered signals.

Step 3: To get the properties of the EEG signal, processed on a nonlinear multiband analysis of the filtered EEG signals using the Discrete Wavelet Packet Transform (DWPT) technique. Then, by including a variety of signal properties into feature vectors, the DWPT technique was utilized to improve diagnostic performance.

Step 4: The improved EEG was categorized using a Combination of a deep convolutional neural network (DCNN) and long short-term memory (LSTM) called DCNN-LSTM. Here, the ideal DCNN hyperparameter is determined using the Adaptive clonal selection algorithm (ACSA), which might increase the classifier's accuracy.

In terms of structure, this paper is organized in five main sections: In Section 2, the existing works are described. Thereafter, the methodology concerning the signal processing and the classification procedure is explained in Section 3. The obtained results and the inherent discussion were covered in Section 4. Lastly, Section 5 makes remarks about conclusions.

RELATED WORKS

Recognizing EEG of AD, MCI and HC is a matter of classification. This section details the latest techniques for diagnosing AD through the combination of EEG and deep learning. These techniques can be divided into two-class classification and three-class classification, and the latter has been the focus of research in recent years.

Ferri et al. [15] collected the resting-state EEG signals from 89 AD subjects and 10 HC subjects, and preprocessed them by using Fast Fourier Transform (FFT) and eLORETA source estimation. They extracted spectral power density calculated by FFT and neural current density by LORETA as the input features of the neural network, gaining a two-class classification accuracy rate of 89.00% on the stacked ANN with autocoder. Although this study innovatively used neural current density as a classification feature, the classification accuracy rate is relatively low.

Duan et al. [16] collected two groups of resting-state EEG signals, i.e., 22 MCI subjects and 38 HC subjects in Group 1 and 17 AD subjects and 24 HC subjects in Group 2 to achieve two two-class classifications of MCI vs. HC and AD vs. HC. Both groups of EEG data were preprocessed with the same band-pass filter. The functional connectivity of EEG was extracted as the input to CNN by calculating the coherence between two electrodes. In their paper, the recognition accuracy rate of CNN model ResNet-18 is 93.42% for MCI and 98.54% for AD. This study is the first to consider the difference of spatial attribute of EEG in AD classification and achieve excellent classification performance, but it lacks direct comparative analysis of AD vs MCI and the challenge of accurately diagnosing MCI remains unresolved.

Two-class classification is less complex than three-class classification, but the latter is more significant in research. MCI is an intermediate stage between HC and AD. At present, it is uncertain whether MCI will definitely develop into AD [17], but the diagnosis of MCI can help to prevent and treat AD. Therefore, it has more clinical application value.

Ieracitano et al. [18] collected EEG signals of 63 AD subjects, 63 MCI subjects and 63 HC subjects, and used two-dimensional grayscale PSD images of EEG as classification features to construct a CNN-based three-class classification model. Although this model is superior to traditional machine learning techniques such as SVM and MLP, its accuracy rate is only 83.33%. Therefore, the study considered using the average PSD image of the entire EEG as a classification feature in the future to reduce the effect of epoch variability. It will also investigate mapping spectral feature representations that are more sensitive to AD and MCI EEG features to improve the accuracy of the classification model.

You et al. [19] proposed a cascade neural network which used the sequential characteristics of EEG and the human gait characteristics to classify AD, MCI and HC, with a classification accuracy rate of 91.70%. This attention-based spatiotemporal graph convolutional network can automatically extract the features of EEG and gait data, reducing human intervention and achieving good classification results. Also, the study used only gait data when differentiating between HC and patients, eliminating the need to use EEG data from HC and saving a lot of data collection time. Although this study provides a new way of thinking about which data to collect to diagnose AD, it is difficult to reproduce the EEG and gait fusion data used when subdividing patients into AD and MCI.

Rad et al. [20] extracted different bands of frequency characteristics, statistical properties, Lyapunov index of task-state EEG, and selected optimal features that had the greatest impact on AD, MCI and HC classifications with variance analysis. The multi-channel deep convolutional neural network constructed by the author reached a three-class classification accuracy rate of 97.50%. This study presents the first analysis method based on the four task states of closed-eye, open-eye, recall and stimulation EEG, opening up a new way of using task-state EEG to diagnose AD. However, the EEG was collected in three self-defined electrode positions (Pz, Cz and Fz), rather than 19 or 64 channels in the 10–20 international system. The three-channel recording of human brain electrical activities is not comprehensive, and the data used in this paper cannot represent the real situation of AD and MCI patients, so the experimental results are somewhat fortuitous.

A paper published by Amini et al. [21] used the time-dependent power spectrum descriptor to extract EEG features of each channel as CNN input. The accuracy rate is 82.30% in the data set consisting of the resting-state EEG of 64 AD subjects, 64 MCI subjects and 64 HC subjects. The TD-PSD method well extracts time-domain and frequency-domain features from the 180 s EEG signals at each epoch (one epoch every 300 s, from 60 to 240 s), minimizing EEG background artifacts. However, the paper did not improve or innovate in CNN. Instead, it only proved that CNN has obvious advantages compared with SVM and LDA in this data set. Moreover, the accuracy rate of the algorithm used in the paper is low, especially for HC (only 75%).

To sum up, the current use of EEG combined with deep learning technology for AD diagnosis is mostly to convert one-dimensional EEG into high-dimensional data for discrimination, but the one-dimensional deep learning model is rarely used directly. This not only makes data preprocessing more difficult, but also increases the human intervention in the research process, reducing the reproducibility of research results.

PROPOSED METHODOLOGY

In this work, the EEG dataset was substantially reduced by using the band-pass elliptic digital filter. Then Variational mode decomposition (VMD) was employed to decompose the filtered signals. Important characteristics might then be extracted from the EEG data by applying the DWPTapproach to process on a nonlinear multiband analysis of the filtered EEG signals. After that, a lot of signal characteristics were added to the DWPTmethodology to build feature vectors and enhance the efficacy of diagnosis. After the required dataset was created, the DCNN-LSTM optimized using ACSA was used to classify the improved EEG is shown in figure 1.

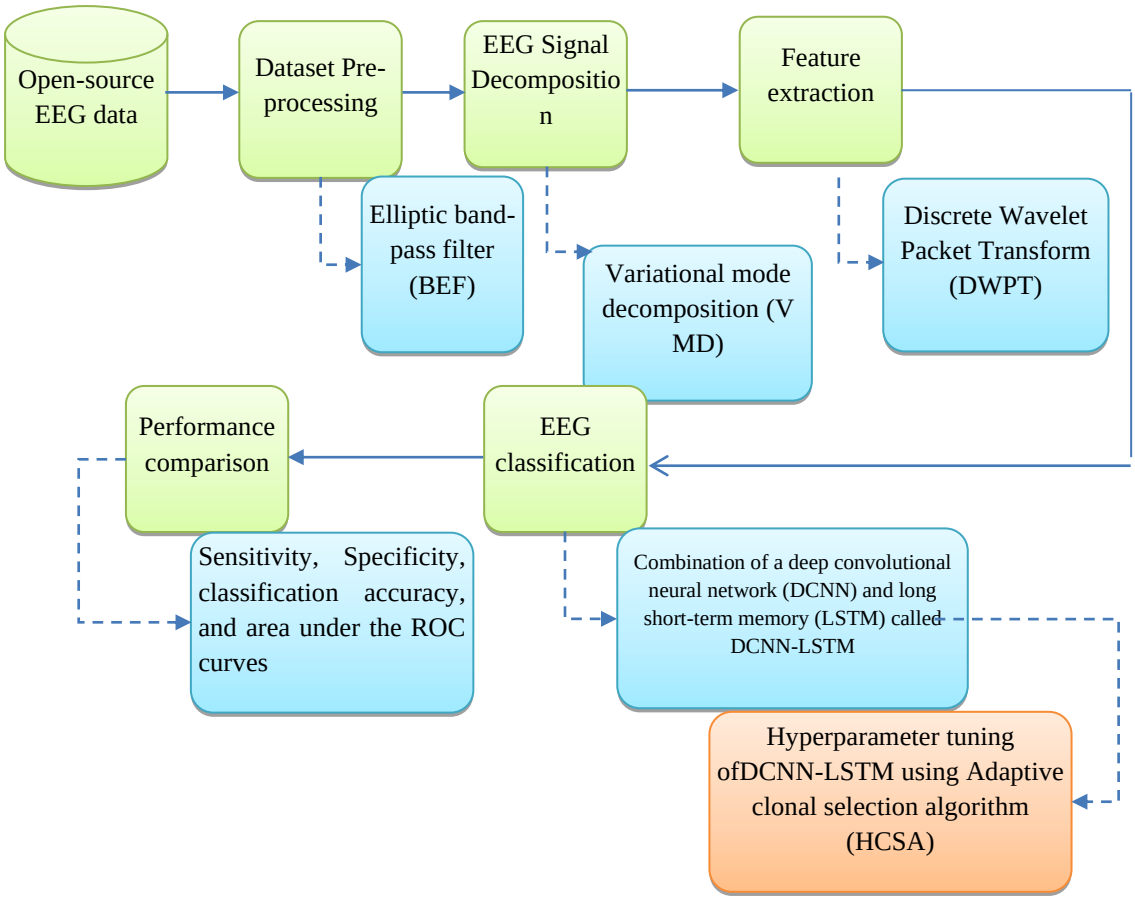

Fig.1. Blok Diagram of Proposed Methodology

Dataset Description

Datasets from the Behavioral and Cognitive Neurology Unit of the Department of Neurology including the Referral Center for Cognitive Disorders at the Hospital das Clinics in Sao Paulo, Brazil, were used in this study to gather information on AD patients and control volunteers. The Mini-Mental State Examination (MMSE) and the Brazilian version of the Clinical Dementia Rating - CDR scale was used by qualified neurologists to recognize all AD individuals and Control volunteers and gather the necessary datasets [22]. 86 participants were split into three groups and provided the multi-channel EEG datasets. With a median age of 66.89 years and 8.18 StD, the first batch of 35 Control subjects (CS) is made up of 18 men and 19 females. Based on individual interviews, the requirements for being included for those in the cognitive normal subgroup included an MMSE of at least 25, a CDR score of 0, a median MMSE of 28, and a mean deviation of 2.2. Moreover, there was no indication of functional cognitive decline prior to recording. The second patient group met the NINCDS-ADRDA [23] along with DSM-IVTR [24] parameters. It consisted of 31 mild-AD patients (a total of 12 male and 19 female), having the median age of 75.23 years as well as 5.55 StDs. The criteria for inclusion that followed significantly reduced the number of individuals with mild Alzheimer's disease: 0.5 CDR 1 as well as MMSE 24, which produced a median MMSE of 19.48 and a standard deviation of 3.16.

Twenty-two individuals with intermediate Alzheimer's disease—seven men and fifteen women—make up the third group. Their average age is 73.77 years, and they have 10.16 StDs.

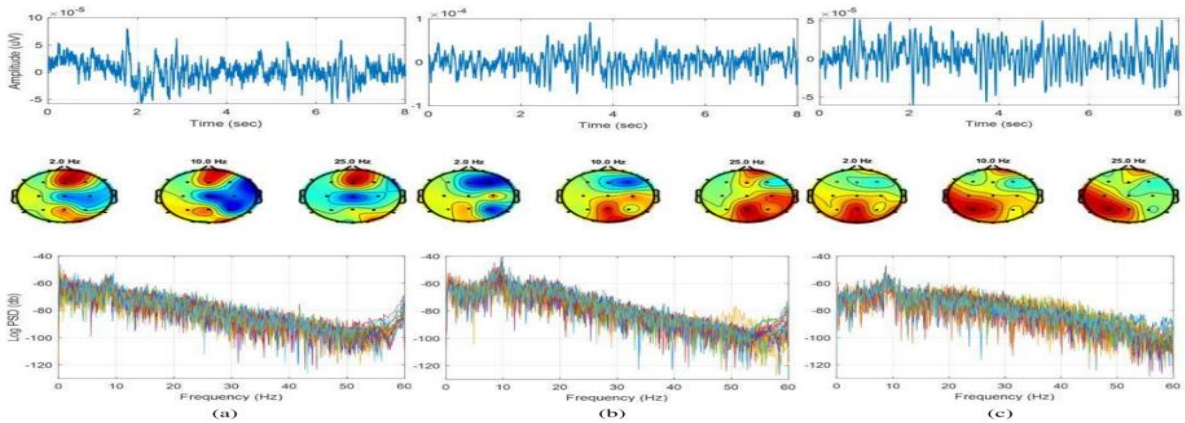

Fig 2. Samples of EEG data, electrode mappings, and power spectrum density pattern for three different EEGs: (a) control; (b) mild AD; and (c) moderate AD.

People with intermediate Alzheimer's disease needed to have an MMSE score of 20, with an average of 14.18 and a variance of 3.69, as well as a CDR score of 2. Inclusion in both AD cohorts (AD1 and AD2) was also necessary due to the existence of cognitive and functional decline in the year prior, as long-term interviews with competent informants indicated. Several other diseases that may possibly contribute to cognitive loss were examined in both AD groups, including B12 (vitamin) deficiency, diabetes, renal illness, thyroid disease, addiction to alcohol, liver or lung disease [25]. An example of electrode mappings, EEG power spectrum density, and EEG signals for three different datasets—control, moderate Alzheimer's disease, as well as moderate Alzheimer's disease—are shown in Figure 2 on a logarithmic scale. The patterns of waves from the Fp1 electrodes collected from three different individuals from three different datasets make up the EEG signal sample. Three different arbitrary frequencies—2, 10, and 25 Hz—for which electrode maps are given in order to show how the three datasets differ from one another.

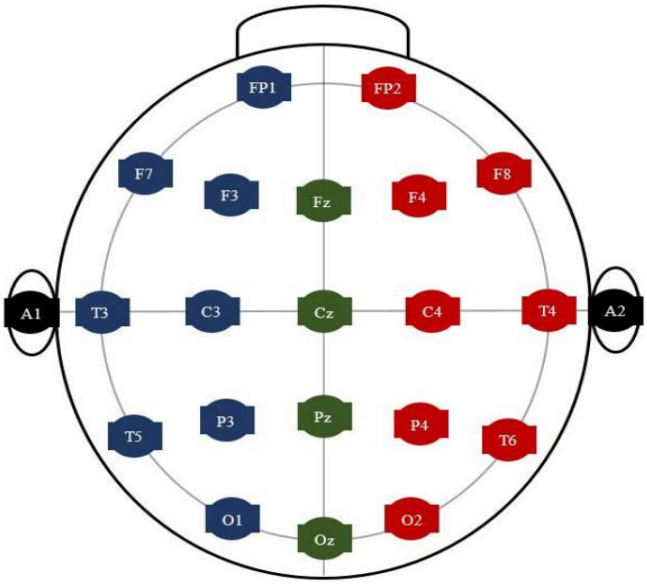

Fig 3. Distribution of the EEG acquisition system electrodes on the scalp

Using the Braintech 3.0 equipment collecting system (EMSA Medical Equipments Inc., Brazil), the EEG dataset was acquired at a sampling rate of 200 Hz and an accuracy of 12 bits. The International 10-20 protocol was followed while positioning electrodes with the EEG data collection instrument. The subjects' earlobes, A1 on the left and A2 on the right, as well as a pair of electrodes were used to collect EEG datasets for this study. The subjects' Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2, Fz, Cz, Pz, and Oz. Figure 3 depicts the dispersion of each electrode. Everyone was awake, calm, and wearing closed eyelids when the assessment was done. EEG artifacts such as muscle movements and blinking were meticulously eliminated from the data by two experienced neurophysiologists. Following that, a minimum of 28 epochs, that extending eight seconds, were selected by examining each person's eyes [26].

Data pre-processing

The research data is publicly available at https://github.com/tsyoshihara/Alzheimer-sClassification-EEG/tree/master/data. EEGs were obtained from 100 patients, including 14 HC subjects, 37 MCI patients, and 49 AD patients. Prior to removing noise and interference from brainwave patterns, the recorded signals from the EEG are analyzed and filtered using a pre-processing block. An elliptic band-pass filter is used to successfully confine the signals to a frequency range of 0.1 to 60 Hz. Beyond the EEG dataset, noises, interferences, and artifacts were recorded. The electrodes, magnetic fields from electronic devices, blood pressure, breathing, limb movement, eye blinking, and other people's actions were all responsible for these distortions, noises, and interferences [27]. During the preprocessing stage, the EEG signals were filtered using a band-pass filter to remove noise and disruptions caused by the EEG recording process. Numerous finite impulse response (FIR) and infinite impulse response (IIR) filters were in use. In this experiment, a band pass IIR elliptic digital filter with cutoff frequencies of 0.1 and 60 Hz was employed. Two skilled neurophysiologists carefully removed EEG artifacts from the EEG data, such as blink and contractions of the muscles [28].

Band pass elliptic digital filter: Band-pass filters allow signals inside the band-pass area to flow through while attenuating frequencies outside of the band-pass zone. The passband and stopband ripple can be adjusted individually. If the stopband and passband ripples approach zero independently, the filter can change into a Chebyshev type I or type II filter. As a result, in order to have a thin transition band, both the passband and stopband ripples must be permitted. Elliptical (also known as Cauer) filters provide a smaller transition band than other filters of the same order [29]. Discrete-time filters can be built in either IIR or FIR modes. The IIR filter is a feedback device with an infinite impulse response, as shown in Figure 2.Equations (1), (2), and (3) offer the mathematical difference equations that define the functioning of the IIR filter. For the IIR elliptic band-pass filter, we choose 100 kHz sampling frequency (f_{samp}), 2 kHz lower cut-off frequency (f_{lco}), 2.5 kHz upper cut-off frequency (f_{uco}), 1 dB pass band ripple, and 80dB stop-band attenuation. Because the filter is of fourth order, it contains four zeros and four poles in the complex plane, as illustrated in [29]. The distance between the zeros and the center is one. Using the suitable Bilinear transformation method, an analogue filter with a transfer function $H(sp)$ may be turned into a digital filter with a transfer function $H(zp)$. A projection between the s-plane and z-plane variables can be constructed if $sp = f(zp)$, This matching is utilized for the band-pass filter, and the equivalent frequency may be found by replacing $sp = j\omega$ and $zp = e^{j\omega}$, where ω is the angular frequency and $\omega = \frac{2\pi f}{f_{samp}}$. Equations (4) and (5) define the bilinear transformation employed in the band-pass filter.

$$y[n] = \sum_{k=0}^{M} rev_k x[n-k] - \sum_{k=1}^{N} frw_k y[n-k] \quad (1)$$

$$H(sp) = \frac{\sum_{k=1}^{M} frw_k z^{-k}}{\sum_{k=1}^{N} rev_k z^{-k}}, rev_0 = 1 \quad (2)$$

$$H(zp) = \frac{frw_0 \prod_{k=0}^{M}(1-rev_k z^{-1})}{\prod_{k=1}^{N}(1-frw_k z^{-1})} \quad (3)$$

$$s = \frac{1-2\cos\omega_0 z^{-1}+z^{-2}}{1-z^{-2}} \quad (4)$$

$$\Omega = \frac{\cos\omega_0 - \cos\omega}{\sin\omega} \quad (5)$$

Where rev_k and frw_k are the reverse and forward filter coefficients, respectively, T_s represents the sampling period and ω_s corresponds to the center frequency of the band-pass/stop filter.

Variational mode decomposition (VMD) for EEG Signal

For signal processing, especially with regards to EEG signals and other biological signals, the feature extraction stage is crucial for achieving the optimum results. To evaluate and break down the EEG signal into its component features, a variety of feature extraction techniques are used. A well-liked and widely utilized method, VMD, was employed in the current investigation. VMD is an adaptive signal decomposition method based on Wiener filter, Hilbert transform and heterodyne demodulation. Its purpose is to decompose a real-valued input signal $x(t)$ into a set of sub-modes $\{u_k\}$ with a particular sparsity. Each of the limited bandwidth sub-modes is tightly centered around a central frequency ω_k. In the process of VMD, each submode u_k is transformed by Hilbert transform to obtain the corresponding unilateral spectrum. The analytical signal of each component is mixed with a pre-estimated center frequency $e^{-j\omega kt}$ and the spectrum of each mode is modulated to the corresponding base-band. In order to obtain the bandwidth of the sub-modes, the constraint variational problem is introduced to calculate the square L^2norm of the gradient of the demodulated signals and estimate the bandwidth of each mode. The optimized variational model constructed is shown in Equation (6).

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_i \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \ \sum_{k=1}^{K} u_k(t) = x(t) \end{cases} \quad (6)$$

where, K is the number of mode components, $\{u_k\} = \{u_1, u_2, \cdots, u_K\}$ and $\{\omega_k\} = \{\omega_1, \omega_2, \cdots, \omega_K\}$ are components and corresponding central frequency respectively. $\sum k := \sum_{k=1}^{K}$ is equivalent to the sum of all band components.

By introducing the quadratic penalty factor α and the Lagrange multiplication operator $\lambda(t)$, the constrained variational problem is transformed into a non-constrained variational problem. The quadratic penalty factor can guarantee the signal reconstruction accuracy under the condition of Gaussian noise, and the Lagrange multiplier emphasizes the strictness of constraints.

$$L(\{u_k\},\{\omega_k\},\lambda) := \alpha \sum_{k=1}^{K} \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \quad (7)$$

Then, the $\{u_k^{n+1}\}, \{\omega_k^{n+1}\}$,and λ^{n+1} are updated alternately by using the alternate direction method of multipliers (ADMM). By seeking the 'saddle point' of the extended Lagrange expression, that is, meeting the stop condition of iteration, mutually independent frequency band components $\{u_k\}$ are finally concluded. The decomposition process of the variational model is summarized in Algorithm 1.

Algorithm: VMD

Initialize: $\{\hat{u}_k^1\}, \{\omega_k^1\}, \hat{\lambda}^1, n \leftarrow 0$

Repeat:

$n \leftarrow n + 1$

For k=1: k do

Update $\hat{u}_k$ for all $\omega \geq 0$;

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) - \sum_{i>k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2 \propto (\omega - \omega_k^n)^2}$$

Update ω_k:

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega}$$

End for

Dual ascent for all $\omega \geq 0$:

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right)$$

Until convergence: $\frac{\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon$

Once the iterations converge, the final decomposed modes and their central frequencies are used to reconstruct the original signal. The reconstructed signal $\hat{x}(t)$ is obtained by summing up all the modes:

$\hat{x}(t) = \sum_{k=1}^{K} u_k(t)$ (8)

Discrete Wavelet Packet Transform (DWPT)and Nonlinear Analysis for feature extraction

In this study, seven significant features (parameters) were combined with the VMD technique to generate feature vectors and improve AD system performance diagnosis. To construct the feature vectors, the significant features were computed for all extracted components, which included both the IMFs and the residual, from all signals of a certain trial. These feature vectors were then processed individually through the classifiers in order to reduce the proposed system complexity and simplify the training process. In order to handle EEG signals as efficiently as possible, the feature-extraction approach is essential. To deal with the relatively small number of values that indicate the properties of the EEG signal, the signals are recorded and split into long time-series. In order to handle EEG signals as efficiently as possible, the feature-extraction approach is essential. To deal with the relatively small number of values that indicate the properties of the EEG signal, the signals are recorded and split into long time-series. These numbers are often referred as being features since they are merged into a vector known as the vector with features. Consequently, feature-extraction methods may be characterized by methods for converting signals into feature vectors. To extract features, a variety of feature-extraction methods are employed. The most popular and often applied approach, DWPT, has been applied in this investigation. We propose

to use DWPT in this study to create vectors of attributes that include the parameters as follows: Logarithmic band power (LBP), Standard deviation (StD), Variance (VAR), Kurtosis (KUR), Average Energy (AE), Root mean square (RMS), and NO (Norm).

Using the STFT, non-stationary data analysis is not possible, such as EEG data. This can be attributed to STFT's consistent accuracy over the entire wavelength range. The technique of multi-resolution wavelet transform is employed to assess various frequencies at various resolutions. Additionally, the wavelet transform would be the ideal choice to avoid the related dimensionality problem because it may provide fewer attributes for the signal to be analyzed. Wavelet transformations, for example, may analyze signal properties in the temporal along with frequency domains by decomposing signals across multiple steps via a single function [30]. This function is provided by and is known as the mother function ξ.

$$\xi(t) = \frac{1}{\sqrt{2}}\xi\left(\frac{t-y}{x}\right) \ldots x, y \in WS, x > 0, \quad (9)$$

where the wavelet space (WS) is represented by the parameter and the scaling and shifting parameters, respectively, by x and y. Wavelet transform is represented by the following equation.

$$F(x,y) = \frac{1}{\sqrt{x}}\int \xi\left(\frac{t-y}{x}\right) dt \quad (10)$$

DWT was utilized in this work because it provides a very useful wavelet representation. Low- and high-pass filters are commonly employed in first-level decomposition to offer a depiction of the digital information as detail (D1) and estimate (A1) components. The following equation provides the DWT decomposition:

$$f(t) = \sum_{k=-\infty}^{k=+X\infty} DC_{n,k}\,(2^{-n}t - k) + \sum_{k=-\infty}^{k=+X\infty}\sum_{k=-\infty}^{k=+X\infty} 2^{\frac{-j}{2}} \times ac_{j,k} \times \psi(2^{-j}t - k) \quad (11)$$

where the detail coefficients are represented by $DC_{n,k}$ and the approximation coefficients by $ac_{j,k}$, respectively; n indicates the level and ∅ is the function of scale. The process is repeated once the first estimate is deconstructed. At the end of the process, n+1 deconstructed signals are present. The mother wavelet function utilized in this study is Daubechies 4 (db4); level 4 was selected since it provides the best qualities for correctly detected signal features. The following techniques were employed to produce the extra feature vectors: There are $n = 1, 2, \ldots, N$, discrete signal samples in S(n), where N is the number of signal samples.

1. Logarithmic band power (LBP)
   $$LBP = log\left(\frac{1}{N}\sum_{n=1}^{N}|S(n)|^2\right) \quad (12)$$
2. The signal's Standard Deviation (σ)
   $$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(S(n) - \mu\ )^2} \quad (13)$$
3. The signal's variance V

$$V = \frac{1}{N}\sum_{n=1}^{N}(S(n) - \mu)^2 \quad (14)$$

4. The kurtosis of the signal
   $$kurt = \frac{E(S(n)-\mu)^4}{\sigma^4} \quad (15)$$
   where E( ) is the expected value of the signal samples.
5. Average Energy (AE)
   $$AE = \sum_{n=1}^{N}|S(n)|^2 \quad (16)$$
6. Root mean square (RMS)
   $$RMS = \sqrt{\frac{\sum_{n=1}^{N}|S(n)|^2}{N}} \quad (17)$$
7. NO (Norm)
   $$NO = \sqrt{\sum_{n=1}^{N}|S(n)|^2} \quad (18)$$

More specifically, Discrete Wavelet Packet Transform (DWPT) provides a simultaneous decomposition of low and high frequencies. Unlike the Discrete Wavelet Transform (DWT), it also considers high frequencies after the 1st level of decomposition. This procedure is iterative, as the signal is successively filtered both by a low-pass and a high-pass filter (obtaining approximation and detail coefficients, respectively) [31].

The DWPT equation is defined as follows:

$$W_{s,m}(n) = \begin{cases} \sqrt{2^s}\sum_{k=0}^{N-1} \_o(2n-k)W_{(s-1),\left(\frac{m}{2}\right)}(k), & if\ m\ even \\ \sqrt{2^s}\sum_{k=0}^{N-1} \_1(2n-k)W_{(s-1),\left(\frac{m-1}{2}\right)}(k), & if\ m\ odd \end{cases} \quad (19)$$

where s is the scale of decomposition ($s = \{0, 1, 2, \ldots, \log_2(n)\} \in Z$), N corresponds to the signals length, m denotes the position in the tree at scale s ($m = \{0, 1, 2, \ldots, 2^j - 1\}$), n are the samples ($n =$

$\{0, 1, 2, \ldots, N - 1\}$), and k represents the time ($k \in R$). $W_{s,m}$ are the coefficients of the Wavelet Packet at the scale s and position m. $\square_0$ *and* $\square_1$ are the high-pass and low-pass Wavelet filters, respectively. These filters depend from the selected wavelet [24]. It must be pointed out that $W_{0,0}$ is the original normalized signal — y(n), the sub-band signal $W_{s,m}(n)$ has its frequency limited in $[m\pi/2^s , (m + 1)\pi/2^s)]$, where π is the normalized angular frequency. The filters applied vary according to the mother wavelet family and subfamily.

Non-Linear Analysis

According to the nonlinear dynamic theory, a complex system (such as the human brain) is characterized by nonlinear dynamic properties. Due to physiological events, the brain environment is constantly changing over time. Consequently, brain waves exhibit a nonlinear and chaotic behavior. Additionally the degree of complexity of the brain represents the time series randomness. Depending on the intensity of the brain activity, the EEG signal can be more or less complex, containing more or less information per signal fragment. Briefly, many methods of extracting nonlinear features have been increasingly explored, making them a powerful approach for EEG signal characterization. Indeed, through the detection of patterns in the time series, it is possible to infer about EEG signal behavior and predict the same kind of occurrences in the future [32].

The present method suggested the extraction of nonlinear and statistic features from each node and from each one of the channels of all participants' signal segments with the purpose of accomplish a nonlinear and multiband analysis. Generally, the non-linearity characteristics correspond to entropy, exponents and fractal dimensions parameters.

Energy

The sub-bands energies are the most used EEG features for detecting AD due to the power shift effect from high to low frequencies that progressively occur as AD progresses. The energy of a signal $x(n)$ is defined as

$$E = \sum_{n=1}^{N} |x(n)|^2 \quad (20)$$

Entropy

Entropy concept arises with the intention of describing the molecules distribution in a given system. This thermodynamic definition explains how molecules are organized, considering the size and atomic configuration of each one. Thus, the assessment of entropy allows the quantitative evaluation of the degree of randomness and uncertainty of a given sequence of data. Entropy is a measure that considers the amount of energy present in the complex system. In fact, entropy features are commonly used, as their analysis allows us to accurately evaluate the non-linear behaviour characteristic of the EEG signals [17,24,26]. Several measures of entropy were used (Shannon, logarithmic, norm, approximate, sample, and permutation entropy), leaving the feature selections process elect the one(s) more relevant for the EEG classification.

Shannon Entropy:

$$SE = -\sum_{n=1}^{N} |x(n)|^2 \log [|x(n)|^2] \quad (21)$$

Logarithmic Entropy:

$$LE = -\sum_{n=1}^{N} \log [|x(n)|^2] \quad (22)$$

Norm Entropy:

$$NE = \sum_{n=1}^{N} |x(n)|^l \quad (23)$$

where l represents the normalized power equal to 1.1 on present study [33].

Approximate Entropy:

$$AE(e, r, N) = H^e(r) - H^{e+1}(r) \quad (24)$$

where N is the data length (suggested to be 1000 of the signal standard deviation), r is the similar tolerance (between 0.1 and 0.25) and e represents the embedding dimension (between 2 and 3). H is the Heaviside function that results from intermediate calculations [33].

• Sample Entropy:

$$SE(e, r, N) = -\ln \frac{A^{e+1}(r)}{B^e(r)} \quad (25)$$

Where,

$$A^{e+1}(r) = d[x_{e+1}(i), x_{e+1}(j)] < r; \quad (26)$$

And

$$B^e(r) = d[x_e(i), x_e(j)] < r; \quad (27)$$

as the factors N, r, e are the same used in Approximate Entropy, the recommended values are similar, and d is the Chebyshev distance between two sets of simultaneous data points [33].

• Permutation Entropy:

$PE(e) = -\sum_{j=1}^{m!} p_j \ln p_j$ (28)

where e is the embedding dimension and $p_j$represents the probability of the j th permutation occurring [30].

Chaos Theory

Chaos theory is an approach closely related to dynamic systems. A dynamic system does not share the properties of a system in equilibrium, wherefore certain unpredictable disturbances may influence its behavior. In this way, those perturbations cause the system transition from one state to another. The concept of phase space represents the set of all possible states through which a dynamic system can pass over time. There are two main exponents which provide a comprehensive framework of chaos.

- Hurst Exponent

$HE = \frac{\epsilon - 1}{2}$ (29)

Where ϵ derives from power-law [26].

- Lyapunov Exponent

$LE(x_o) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} ln\left|f'(x_k - 1)\right|$ (30)

Where f' is the derivative of iterator function f.

Fractal Analysis

Fractal structures present self-similarity properties that enable detailing both irregular processes and structures. Regarding brain activity study, fractal geometry has been showing to be a useful approach to identify neurological pathologies. In fact, monitoring the selfsimilarity of the brain rhythms allows the characterization of the patient clinical condition. Fractal feature extraction such as the Higuchi fractal dimension is considered to be helpful in EEG signal classification [34].

• Higuchi Exponent:

$L(k) \propto k^{-D}$ (31)

where k indicates the time interval, L(k) is the length of the curve in the k time interval and D is the Higuchi Exponent.

For each 38 study participants, 10 non-linear features (Energy, Shannon Entropy, Logarithmic Entropy, Approximate Entropy, Sample Entropy, Permutation Entropy, Norm entropy, Hurst Exponent, Lyapunov Exponent, and Higuchi Exponent) are calculated from the 18 DWP nodes signals of all 5 s segments per channel. At the same time, 10 timeseries statistics used for a fast signal analysis, such as Maximum, Mean, Median, Mode, Kurtosis, Standard Deviation, Variance, Skewness, Root Mean Square and Asymmetry, have been also extracted from each 18 DWP nodes. Thus, a total of 20 features have been computed in each DWP node 5 s signal analysis per channel. Treating the 20 extracted features of all segments per DWT Node analysis as time series distributions, 3 statistics have been used to compress them along the channel, reducing in this way the dimensionality of the problem. Mean, variance (var) and standard deviation (sd) have been picked for this purpose. At the end of the process, each DWT node analysis per channel produced 60 features (20 time series features digitally compressed by 3 statistics, respectively).

DCNN-LSTM with ACS for classification of EEG data

The extracted features from the EEG signal are given as input to the classifier. The classifier specifies that the new observation belongs to which category. The present study proposed a Combination of a deep convolutional neural network (DCNN) and long short-term memory (LSTM) called DCNN-LSTM. The DCNN-LSTM has the hyperparameters which is tuned by proposed optimization algorithm called Adaptive clonal selection (ACS) algorithm**.**

Convolutional neural network

A particular type of multilayer perceptron is a CNN, but a simple neural network cannot learn complex features, unlike a deep learning architecture. CNNs have shown excellent performance in many applications [35], such as image classification, object detection, and medical image analysis. The main idea behind a CNN is that it can obtain local features from high layer inputs and transfer them to lower layers for more complex features. A CNN comprises convolutional, pooling, and fully connected (FC) layers. A typical CNN architecture along with these layers is depicted in Fig. 3. The convolutional layer encompasses a set of kernels for determining a tensor of feature maps. These kernels convolve an entire input using “stride(s)” so that the dimensions of an output volume become integers. The dimensions of an input volume decrease after the convolutional layer is used to execute the striding process. Therefore,

zero padding [36] is required to pad an input volume with zeros and maintain the dimensions of an input volume with low-level features. The operation of the convolutional layer is given as:

$$F(i,j) = (I * K)(i,j) = \sum\sum I(i+m, j+n)K(m,n) \quad (32)$$

where I refers to the input matrix, K denotes a 2D filter of size $m \times n$, and F represents the output of a 2D feature map. The operation of the convolutional layer is denoted by $I * K$. To increase nonlinearity in feature maps, the rectified linear unit (ReLU) layer is used. ReLU computes activation by keeping the threshold input at zero. It is mathematically expressed as follows:

$$f(x) = max(0, x) \quad (33)$$

The pooling layer performs a downsampling of a given input dimension to reduce the number of parameters. Max pooling is the most common method, which produces the maximum value in an input region. The FC layer is used as a classifier that makes a decision on the basis of features obtained from the convolutional and pooling layers.

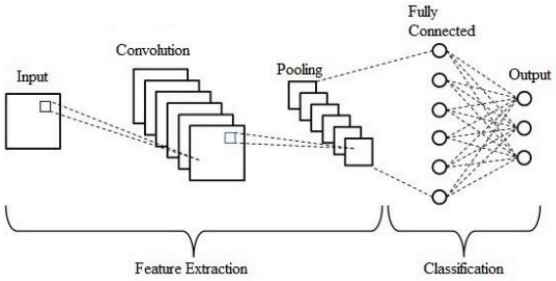

Fig. 3. A typical architecture of the convolutional neural network

Long short-term memory

Long short-term memory is an improvement of recurrent neural networks (RNNs). LSTM proposes memory blocks instead of conventional RNN units in solving the vanishing and exploding gradient problem [37]. It then adds a cell state to save long-term states, which is its main difference from RNNs. An LSTM network can remember and connect previous information to data obtained in the present. LSTM is combined with three gates, such as an input gate, a "forget" gate, and an output gate, where x_t refers to the current input; Ct and Ct− 1 denote the new and previous cell states, respectively; and ht and ht− 1 are the current and previous outputs, respectively. The internal structure of LSTM is shown in Fig. 4.

The principle of the input gate of LSTM is shown in the following forms.

$$\begin{aligned} i_t &= \sigma(W_i.[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_i.[h_{t-1}, x_t] + b_i) \\ C_t &= f_t C_{t-1} + i_t \tilde{C}_t \end{aligned} \quad (34)$$

where (3) is used to pass h_{t-1} and $x_t$through a sigmoid layer to decide on which portion of information should be added. Subsequently, (4) is employed to obtain new information after h_{t-1} and x_t are passed through the tanh layer. The current moment information, $\tilde{C}_t$, and long-term memory information C_{t-1} into $C_t$are combined in (5), where W_i refers to a sigmoid output, and $\tilde{C}_t$ refers to a tanh output. Here, $W_i$denotes weight matrices, and bi represents the input gate bias of LSTM. The LSTM's forget gate then allows the selective passage of information using a sigmoid layer and a dot product. The decision about whether to forget related information from a previous cell with a certain probability is executed using (6), in which W_f refers to the weight matrix, bf is the offset, and σ is the sigmoid function.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (35)$$

The LSTM's output gate determines the states that are required for continuation by the h_{t-1} and $x_t$inputs following (7) and (8). The final output is obtained and multiplied by the state decision vectors that pass new information, C_t, through the tanh layer

$$O_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (36)$$

$$?_t = O_t \tanh(C_t) \quad (37)$$

where W_o *and* b_o are the output gate's weighted matrices and LSTM bias, respectively

Combined DCNN-LSTM network

In this study, a combined method was developed to automatically classifyEEG data using their features. The structure of this architecture was designed by combining CNN and LSTM networks, where the CNN is used to extract complex features from input, and LSTM is used as a classifier. The network has 20 layers: 12 convolutional layers, five pooling layers, one FC layer, one LSTM layer, and one output layer with the softmax function. Each convolution block is combined with two or three 2D CNNs and one pooling layer, followed by a dropout layer characterized by a 25% dropout rate. The convolutional layer with a size of 3 × 3 kernels is used for feature extraction that is activated by the ReLU function. The max-pooling layer

with a size of 2 × 2 kernels is used to reduce the dimensions of an input. In the last part of the architecture, the function map is transferred to the LSTM layer to extract time information. After the convolutional block, the output shape is found to be (none, 7, 7, 512). Using the reshape method, the input size of the LSTM layer has become (49, 512). After analyzing the time characteristics, the architecture sorts the EEG through a fully connected layer to predict whether they belong under which categories (Presence of Alzheimer's Disease, moderateor normal).

Based on the quantity of EEG datasets, the following five categorization issues have been studied: Two classes of characteristics: Control in comparison to symptoms of mild and moderate AD Three classes of AD characteristics were compared: mild against moderate, regulate versus weak and moderate, and controls vs weak vs intermediate (3-class). The hyperparameter definition has a significant impact on the network's efficiency. The most significant issue is that there is no accepted paradigm for determining the ideal set of DCNN-LSTM hyperparameters and creating the architecture of the network. Using the ACS algorithm in this work, the authors created an automated method for hyperparameter optimization and structural design to solve this problem. The DCNN-LSTM model's chosen hyperparameters are batch size, number of epochs, number of filters, pooling size, and kernel size.

Adaptive clonal selection (ACS) algorithm for hyperparameter optimization

At present, many cloning selection algorithms have been studied, and improvements have been made to the cloning, mutation and selection steps. However, there is a lack of research on the optimization of the updating operation steps. The clonal selection algorithm is traditionally updated through a random complement of antibodies, which is a blind and uncertain process. The added antibodies may gather near a local optimal solution, resulting in the need for more iterations to obtain the global optimal solution. To solve this problem, our improved algorithm introduces a crowding degree factor in the antibody updating stage to determine whether there is crowding between antibodies. By eliminating antibodies with high crowding potential and poor afnity, the improved algorithm guides the antibodies to update in the direction of the global optimal solution and ensures stable convergence with fewer iterations.

Organisms have a variety of immunologically active cells, and immune cells undergo mitosis to produce clones, i.e., a single cell gives rise to a group of cells. This mechanism of generating antibodies that recognize and destroy invading antigens and using asexually reproducing clones to activate differentiated and proliferating antibodies to achieve an immune response and ultimately remove the antigen is called clonal selection [38].

The flow of the traditional CSA is shown in the following steps:

Step 1: Initialize the parameter. Initialize the population size D, outstanding antibody size N, dimension dim, upper and lower bounds M and m for the array of antibody values, respectively, maximum number of iterations G_{max}, and variation probability p where N < D.

Step 2: Initialize the antibody population. Random generated antibody population of antibodies A of size D, x_i is the i th antibody $x, i \in [1, D]$.

$$A = \{x_1, x_2, \dots, x_i, \dots, x_D\} \quad (38)$$

Step 3: Calculate affinity. Calculate the affinity of each antibody in the antibody population by using the objective function as the target value.

Step 4: Perform segmentation. Sort the calculated antibody affinities and select the top N antibodies as the outstanding antibody population B.

$$B = \{x_1, x_2, \dots, x_i, \dots, x_N\} \quad (39)$$

Step 5: Clone. Clone each antibody at different scales, depending on affinity, to obtain the clonal population B'.

$$B' = \{x_1, \dots, x_1, x_2, \dots, x_2, x_i, \dots, x_N \dots, x_N\} \quad (40)$$

Step 6: Mutate. Mutate the B′ antibody population with probability p to obtain the mutated antibody population B″, where $x_i'$is the i th mutated antibody $x, i \in [1, N]$.

$$B'' = \{x_1', \dots, x_1', x_2', \dots, x_2', x_i', \dots, x_N' \dots, x_N'\} \quad (41)$$

Step 7: Select. For the cloned mutated antibodies, the antibody with the best affinity is selected to replace the original antibody. Finally, the new top N antibodies with good affinity are obtained as a new population of outstanding antibodies C.

$$C = \{x_1', x_2', \dots, x_i', \dots, x_N'\} \quad (42)$$

Step 8: Stock update. Randomly replenish d antibodies combined with the outstanding antibody population C as the initial population for the next generation.

$$d = D - N \quad (43)$$

Step 9: Repeat step 3 if the maximum number of iterations $G_{max}$is not satisfed; otherwise, output the result.

Update operations with a targeted update policy

The directed update strategy is an improvement on the traditional random update method. The integration of the directed update strategy into the update operation of the clonal selection algorithm regulates the balance between randomness and determinism, thus avoiding the problems of low accuracy in fnding the best solution and slow convergence rate due to complete randomness in the local search phase [38]. It can also avoid the blindness and uncertainty of antibodies generated by complete randomness, which may gather near a local optimal solution and produce crowding, so that the global optimum is never found.

The DUSCSA uses a threshold σ to determine whether to add a random complement of antibodies to the antibody population. The threshold calculates the farthest distance between antibodies and introduces a crowding degree factor (cdf), where $cdf \in (0,1)$, which is used as the defning value for determining "crowding". If the difference in position between the current antibody and the best antibody in the current iteration is greater than the threshold σ, crowding has not occurred and can be added to the antibody population; if the difference in position between the current antibody and the best antibody in the current iteration is less than the threshold σ, crowding has occurred. If the diference between the current antibody and the best antibody in this iteration is less than a threshold σ, crowding has occurred. In this case, the affinity between the two needs to be further determined, and if the current antibody has a higher affinity, this antibody can also be added to the antibody population. Otherwise, the algorithm continues to generate antibodies at random. The expression for the directed update strategy is as follows:

$$\sigma = cdf\sqrt{\sum_{i=1}^{dim}(M(i) - m[i])^2} \quad (44)$$

$$div = \sqrt{\sum_{i=1}^{dim}\left(x_i^g - b_i^g\right)^2} \quad (45)$$

where M denotes the array of upper limits of antibody values, m denotes the array of lower limits of antibody values, dim denotes the antibody dimension, div denotes the distance between the current antibody and the optimal antibody during the current iteration, x_i^g denotes the i-th variable of antibody x during the g-th iteration, b_i^g denotes the i-th variable of the optimal antibody b during the g-th iteration, $1 \le i \le dim$, and $g \le G_{max}$.

$$\begin{cases} A \Leftarrow x_i^g = r(M[i] - m[i] + m[i] + div > \sigma \| Aff(x^g) > Aff(b^g) \\ \qquad\qquad r(M[i] - m[i] + m[i] \qquad\qquad ot\square ers \end{cases} \quad (46)$$

where r denotes a uniformly distributed random number between [0, 1]. $Aff(x_i^g)$ denotes the affinity of the i-th variable of antibody x during the g-th iteration, and $Aff(b_i^g)$ denotes the affinity of the i-th variable of the optimal antibody b during the g-th iteration.

This process was repeated ten times; each time, one subset was transmitted into the testing classifier. Finally, the results were averaged to produce overall classification accuracy.

EXPERIMENTAL RESULTS AND DISCUSSION

The sensitivity, specificity, receiver operating characteristic curve area, and classifier accuracy may all be used to evaluate a diagnostic system. A multitude of classifiers were developed implemented and evaluated in order to obtain the greatest diagnostic efficiency and accuracy in classification. In this study, the proposed DCNN-LSTM with ACS is compared with various methods, such as RF, ELM, QDA, KNN, SVM, ANN, and MCNN classifiers, that produced the best outcomes. The k-fold cross-validation approach is used in the classification process, which randomly divides all EEG features into k equal groups [39]. The balance of subsets is used for training, and one subset is chosen for testing (validation). A selected number of each has been used once in the k repetitions of this procedure. We used 10-fold cross-validation in the current study, putting all of the detected EEG signal characteristics into a feature matrix that was obtained using feature extraction techniques, and then applying 10-fold a cross-valid to the feature matrix. Following that, these qualities were split into two distinct subsets: 10% for testing and 90% for training. The examination subset was given into the model after it was successfully trained each time, and the training subset was used for training the classifier with the aim to build and preserve the database of features of the learned classifier.

As mentioned before, the EEG data utilized in this study were divided into 3 categories. Thirty-one moderate AD patients, twenty mild AD patients, as well as 30 Normal subjects submitted the EEG records for the initial, next, and final groups, respectively. The electroencephalogram (EEG) dataset was processed utilizing a band-pass elliptic digital filter having the threshold frequency of 0 and 60 Hz in order to improve the signal-to-noise ratio. Next, various statistical parameters and nonlinear analysis parameters are used to improve diagnostic performance and generate the EEG feature vectors. Ultimately, several classifier types were employed to divide the EEG data into groups based on these

classifications, and the classification accuracy of each group was calculated and contrasted. An in-depth evaluation of the recommended techniques was made possible by plotting the receiver's operating characteristic curves and calculating the regions beneath them.

Its definition is a ratio of correctly segmented samples to the total number of samples. This represents one of the most frequently utilized metrics of the effectiveness of classification, as seen below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (47)$$

Where True Positive (TP): This denotes the accurate detection of AD. True Negative (TN): This status shows that a feature was appropriately identified as non-AD. FP: It means that a feature was mistakenly identified as AD. False Negative (FN): This signal means that feature was missed and classified as an AD.

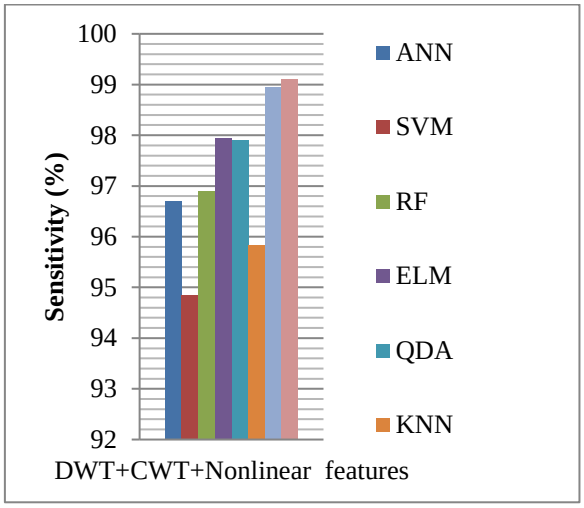

Fig.4. Sensitivity performance comparison

Figure 4 compares the sensitivity of the existing RF, ELM, QDA, KNN, SVM, ANN, OMCNN and that of the proposedDCNN-LSTM with ACS. Figures comparing the proposed strategy to existing methods indicate that it can achieve a high sensitivity rate when used for DPWT based feature extraction. When compared to other existing approaches such as RF, ELM, QDA, KNN, SVM, ANN, OMCNN for DWT and, the proposedDCNN-LSTM with ACSachieves a high sensitivity rate of 99.12%, indicating its effectiveness in AD detection. The recommended solution displays a significantly lower end error and a substantially faster and less variable training curve when compared to the preceding strategies. These results provide additional evidence that the suggested method has a greater rate of AD detection.

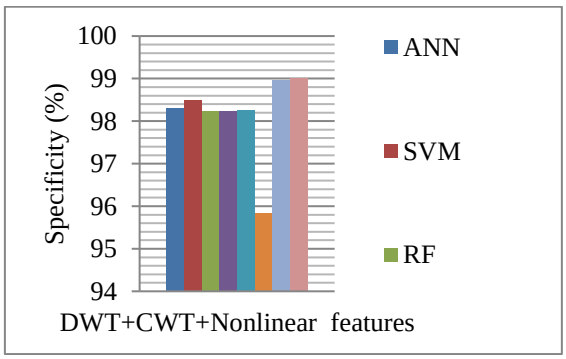

Fig.5. Specificity performance comparison

The results of a specificity comparison between the present RF, ELM, QDA, KNN, SVM, ANN, KNN,OMCNN and the proposed DCNN-LSTM with ACSare shown in Figure 5. According to the results, the suggested DCNN-LSTM with ACSobtains specificity rates of 99.02% for DWT-based feature extraction approaches.

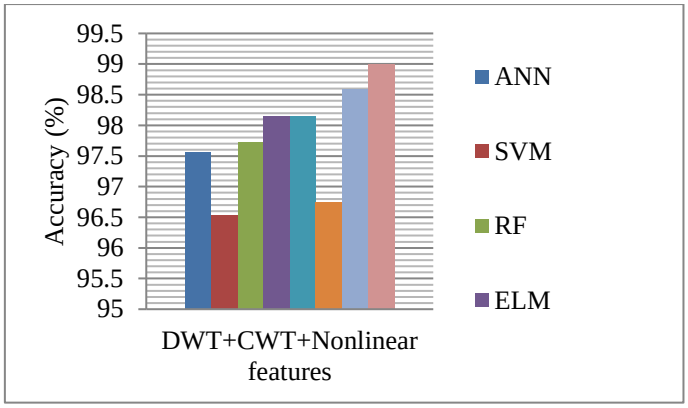

Fig.6. Accuracy performance comparison

Fig. 6 shows the accuracy comparison of the proposed DCNN-LSTM with ACSwith the current RF, ELM, QDA, KNN, SVM, ANN and OMCNN. Based on the data, a high accuracy rate value was obtained by the suggested DCNN-LSTM with ACS, with a high rate of AD detection. With an efficient prediction model, the training time would be shortened by the filtering and feature extraction based on DPWT in the suggested scheme. Accuracy value is about 99% for DPWT along with nonlinear analysis, respectively, show that the suggested study can provide superior AD detection findings than current approaches. Medical professionals and physicians can benefit from the suggested method's automated, quick, simple, accurate, and efficient identification of Alzheimer's disease. The suggested method may cut down on the restricted number of neurologists, speed up diagnosis, and increase diagnosis precision.

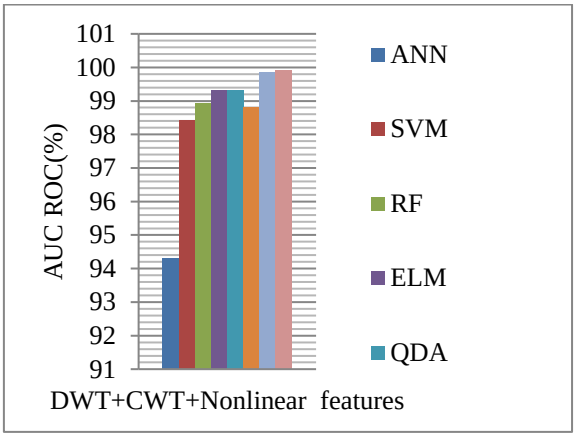

Fig.7. AUCROC performance comparison

The AUCROC comparison for AD prediction is shown in Fig. 7. The findings demonstrate that, in terms of improved liquid particle identification results with high accuracy rates, the suggested DCNN-LSTM with ACStechnique outperforms the current algorithms. DCNN-LSTM with ACSlearning approaches improve accuracy without causing the local optima issue because they are very resilient to distortion in data used for training. These results imply that the suggested method is better capable of reliably and effectively differentiating not just AD however their sorts.

CONCLUSION AND FUTURE WORK

The main aim of this work is to develop an Alzheimer's disease diagnosis tool based on EEG data processing. The creation of an automated diagnostic system which can interpret signals from the brain on itself would quicken up and boost the reliability of the diagnosis process. The gathered EEG datasets during this investigation were filtered using BEF. After that, many signal features were combined using the DPWT methodology to improve diagnosis performance, and it helps to divide the signal that was filtered into high and low frequency bands. DCNN-LSTM with ACSthen looked into how EEG characteristics were categorized into groups. This study aims to evaluate many approaches and determine the best combination approach for Alzheimer's disease diagnosis. Datasets related to manageable, minor, and moderate Alzheimer's illnesses were utilized. The proposed system obtained better accuracy for classification. Additionally, the accuracy of the present OMCNN was enhanced by 3% when compared to the traditional model. This makes the recommended approach for processing EEG

signals very important, and the best classification would result from using a customized hyper-parameter model to predict the severity range. In future several preprocessing and feature selection approaches will be used to identify characteristics that might be helpful for future early Alzheimer's disease detection.

REFERENCES

1. N.N. Kulkarni, V.K. Bairagi, Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier, IETE J. Res., 63 (1) (2016), pp. 11-22.
2. Y. Chen, et al. DCCA cross-correlation coefficients reveals the change of both synchronization and oscillation in EEG of Alzheimer disease patients, Phys. A, 490 (2018), pp. 171-184.
3. V. Bairagi, EEG signal analysis for early diagnosis of Alzheimer disease using spectral and wavelet based features, Int. J. Inf. Technol., 10 (3) (2018), pp. 403-412.
4. L.R. Trambaiolli, N. Spolaôr, A.C. Lorena, R. Anghinah, J.R. Sato, Feature selection before EEG classification supports the diagnosis of Alzheimer's disease, Clin. Neurophysiol., 128 (10) (2017), pp. 2058-2067.
5. S.J. Ruiz-Gómez, et al., Automated multiclass classification of spontaneous EEG activity in Alzheimer's disease and mild cognitive impairment, Entropy, 20 (1) (2018), pp. 1-15.
6. R. Cassani, M. Estarellas, R. San-martin, F.J. Fraga, T.H. Falk, Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment, Dis. Markers, 2018 (2018), pp. 1-26.
7. J.P. Amezquita-sanchez, N. Mammone, F.C. Morabito, S. Marino, A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals, J. Neurosci. Methods, 322 (2019), pp. 88-95.
8. P. Dai, F. Gwadry-Sridhar, M. Bauer, M. Borrie, A hybrid manifold learning algorithm for the diagnosis and prognostication of Alzheimer's disease, AMIA Annual Symposium Proceedings, vol. 2015 (2015), p. 475
9. N. Kulkarni, Use of complexity based features in diagnosis of mild Alzheimer disease using EEG signals, Int. J. Inf. Technol., 10 (1) (2018), pp. 59-64
10. Rodrigues, P. M., Bispo, B. C., Garrett, C., Alves, D., Teixeira, J. P., and Freitas, D. (2021) Lacsogram: A new EEG tool to diagnose Alzheimer's disease. IEEE Journal of Biomedical and Health Informatics, 25(9), 3384-3395. doi: 10.1109/JBHI.2021.3069789
11. Kulkarni, N. (2018) Use of complexity-based features in diagnosis of mild Alzheimer disease using EEG signals. International Journal of Information Technology, 10(1), 59-64. doi: 10.1007/s41870-017-0057-0
12. Klepl, D., He, F., Wu, M., De Marco, M., Blackburn, D. J., and Sarrigiannis, P. G. (2021) Characterising Alzheimer's disease with EEG-based energy landscape analysis. IEEE Journal of Biomedical and Health Informatics, 26(3), 992-1000.
13. Tosun, M. (2021). Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning. Physical and Engineering Sciences in Medicine, 44(3), 693-702. doi: 10.1007/s13246-021-01018-x
14. Amezquita-Sanchez, J. P., Mammone, N., Morabito, F. C., Marino, S., and Adeli, H. (2019) A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals. Journal of Neuroscience Methods, 322, 88-95. doi: 10.1016/j.jneumeth.2019.04.013
15. F. Duan, Z. Huang, Z. Sun, et al. Topological network analysis of early Alzheimer's disease based on resting-state EEG, IEEE Trans. Neural Syst. Rehabil. Eng., 28 (10) (2020), pp. 2164-2172.
16. M.A. Kramer, F.L. Chang, M.E. Cohen, et al.Synchronization measures of the scalp electroencephalogram can discriminate healthy from Alzheimer's subjects, Int. J. Neural Syst., 17 (2) (2007), pp. 61-69
17. C. Ieracitano, N. Mammone, A. Bramanti, et al. A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings, Neurocomputing, 323 (2019), pp. 96-107
18. Z. You, R. Zeng, X. Lan, et al. Alzheimer's disease classification with a cascade neural network, Front. Public Health (2020).
19. E. Mazrooei Rad, M. Azarnoosh, M. Ghoshuni, et al. Diagnosis of mild Alzheimer's disease by EEG and ERP signals using linear and nonlinear classifiers, Biomed. Signal. Proces. (2021)
20. M. Amini, M. Pedram, A. Moradi, et al. Diagnosis of Alzheimer's disease by time-dependent power spectrum descriptors and convolutional neural network using EEG signal, Comput. Math. Methods Med. (2021)

21. C.J. Huggins, J. Escudero, M.A. Parra, et al. Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing, J. Neural. Eng., 18 (4) (2021), pp. 1-14
22. S. Brucki, R. Nitrini, P. Caramelli, P. Bertolucci, and I. H. Okamoto, ‘‘Suggestions for utilization of the mini-mental state examination in Brazil,’’ Arquivos de Neuro-Psiquiatria, vol. 61, no. 3B, pp. 777–781, 2003.
23. G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, ‘‘Clinical diagnosis of Alzheimer’s disease: Report of the NINCDS-ADRDA work group* under the auspices of Department of Health and Human services task force on Alzheimer’s disease,’’ Neurol., vol. 34, no. 7, p. 939, 1984.
24. Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association, Washington, DC, USA, 1980, vol. 3.
25. P. A. M. Kanda, L. R. Trambaiolli, A. C. Lorena, F. J. Fraga, L. F. I. Basile, R. Nitrini, and R. Anghinah, ‘‘Clinician’s road map to wavelet EEG as an Alzheimer’s disease biomarker,’’ Clin. EEG Neurosci., vol. 45, no. 2, pp. 104–112, 2014
26. R. Cassani, T. H. Falk, F. J. Fraga, P. A. Kanda, and R. Anghinah, ‘‘Towards automated EEG-based Alzheimer’s disease diagnosis using relevance vector machines,’’ in Proc. 5th ISSNIP-IEEE Biosignals Biorobot. Conf., Biosignals Robot. Better Safer Living (BRC), May 2014, pp. 1–6.
27. X. Perrin, ‘‘Semi-autonomous navigation of an assistive robot using low throughput interfaces,’’ Ph.D. dissertation, ETH Zurich, Zürich, Switzerland, 2009.
28. F. J. Fraga, T. H. Falk, L. R. Trambaiolli, E. F. Oliveira, W. H. L. Pinaya, P. A. M. Kanda, and R. Anghinah, ‘‘Towards an EEG-based biomarker for Alzheimer’s disease: Improving amplitude modulation analysis features,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., May 2013, pp. 1207–1211.
29. Prommee, P., Karawanich, K., Khateb, F., & Kulej, T. (2021). Voltage-mode elliptic band-pass filter based on multiple-input transconductor. *IEEE Access*, *9*, 32582-32590.
30. Chavan, M. S., Agarwala, R. A., & Uplane, M. D. (2006). Digital elliptic filter application for noise reduction in ECG signal. WSEAS transactions on Electronics, 3(1), 65-70.
31. Sairamya, N.; Subathra, M.; Suviseshamuthu, E.S.; George, S.T. A new approach for automatic detection of focal EEG signals using wavelet packet decomposition and quad binary pattern method. Biomed. Signal Process. Control 2021, 63, 102096
32. Hu, L.; Zhang, Z. EEG Signal Processing and Feature Extraction; Springer: Singapore, 2019. [CrossRef]
33. Faust, O.; Bairy, M. Nonlinear analysis of physiological signals: A review. J. Mech. Med. Biol. 2012, 12, 1–21.
34. Albertovich, T.D.; Aleksandrovna, R.I. The Fractal Analysis of the Images and Signals in Medical Diagnostics. In Fractal Analysis—Applications in Health Sciences and Social Sciences; InTechOpen: London, UK, 2017
35. R.L. Galvez, A.A. Bandala, E.P. Dadios, R.R.P. Vicerra, J.M.Z. Maning Object detection using convolutional neural networks,IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, Institute of Electrical and Electronics Engineers Inc. (2019), pp. 2023-2027,
36. Avcı Kutlu, A novel method for classifying liver and brain tumors using convolutional neural networks, discrete wavelet transform and long short-term memory networks,Sensors, 19 (2019), p. 1992,
37. S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions,Int J Uncertain Fuzziness Knowledge-Based Syst, 6 (1998), pp. 107-116,
38. Yang, C., Huang, Z., Jiang, B., Zhu, M., Luo, A., & He, J. (2023). Improved clonal selection algorithm based on the directional update strategy. The Journal of Supercomputing, 79(17), 19312-19331.
39. P. Refaeilzadeh, L. Tang, H. Liu, L. Angeles, and C. D. Scientist, ‘‘Crossvalidation,’’ Encyclopedia Database Syst., vol. 5, pp. 532–538, Jan. 20