Biomedical Signal Processing and Control 86 (2023) 105318

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

EEGAlzheimer'sNet: Development of transformer-based attention long short term memory network for detecting Alzheimer disease using EEG signal

Dileep kumar Ravikanti *, Saravanan S.

Computing Technology, SRM Institute of Science and Technology, Chennai, India

ARTICLE INFO

Keywords:
Alzheimer's Disease
EEG signal
3-level Lifting Wavelet Transform
Enhanced Wild Geese Lemurs Optimizer
Weighted Stacked Features Optimized
Transformer-based Attention Long Short Term Memory

ABSTRACT

A previous diagnosis of Alzheimer's disease (AD) in its initial stages is needed for patient care because it helps patients adopt preventative measures before irreversible brain damage occurs. Several studies have used computers to detect AD, although hereditary results limit most computer detection methods. There is no straightforward method to screen for AD, partly because the condition is difficult to diagnose and sometimes requires costly and occasionally intrusive testing that is uncommon outside of highly specialized clinical settings. Therefore, this study implements a deep learning strategy for detecting AD with the help of the "Electroencephalogram (EEG)" signal. Initially, the required EEG signal is obtained from traditional online databases and then applied to the 3-level "Lifting Wavelet Transform (LWT)" decomposition to decompose the signal into many wavelets. From the decomposed signal, the temporal features are retrieved by a "Recurrent Neural Network (RNN)", and the spatial features are extracted from a "Multi-scale dilated Convolutional Neural Network (CNN)". Further, the Enhanced Wild Geese Lemurs Optimizer (EWGLO) algorithm is implemented to find the optimal weight value for acquiring the weighted stacked features. These resultant weighted stacked features are applied to the semi-detection stage, where the "Optimized Transformer-based Attention Long Short Term Memory (OTA-LSTM)" model is utilized to detect AD. In the detection stage, parameter optimization takes place to increase the performance of the detection process using the same EWGLO. The designed model is validated with various performance metrics to show the effective outcome. Moreover, the developed model attains 96% and 98% in terms of accuracy and MCC. Throughout the validation, the offered model shows enriched performance when compared with other-state-of-art methods.

1. Introduction

AD is a brain condition that manifests as progressive dementia in middle-aged or old age. Neurotic plaques, neurofibrillary tangles, and the degeneration of certain nerve cells were the pathologic features [1]. It was challenging to determine whether an individual had AD, particularly when the disease was still developing [2]. On the other hand, the stage at which the condition was accurately discovered and diagnosed significantly impacts how well-known therapies work [3]. To maintain cognitive abilities and quality of life to have the most therapeutic effect. It was crucial to identify the individuals before the emergence of major clinical symptoms [4]. Memory loss and other cognitive impairments gradually beginning and progressing are clinical indicators of AD. The initial AD stage had no sensory, coordination, or noticeable motor abnormalities. Laboratory tests can't establish the diagnosis. This method [5] of employing a combination of motor and psychological ability tests to diagnose AD was not entirely accurate [6]. Since AD is a brain disorder, measuring brain activity may be able to detect its presence. In this way, EEG signal analysis has been frequently utilized to identify the existence of AD.

The average electrical activity of the brain at various sites throughout the skull is what the EEG represents [7]. More specifically, it was the total of a sizable set of neurons' extracellular current fluxes. EEG recordings were made using high-conductivity electrodes placed throughout the head [8]. These measurements were mostly made using special electrodes inserted into the brain following surgery or on the scalp [9]. Early AD identification by the EEG signal has generated more attention in the field because of the signal's nonlinear fluctuations and

* Corresponding author.
E-mail address: dr7128@srmist.edu.in (D. Ravikanti).

https://doi.org/10.1016/j.bspc.2023.105318
Received 21 April 2023; Received in revised form 13 July 2023; Accepted 31 July 2023
Available online 15 August 2023

dynamic nature. Learning abstract EEG representations for better generalization is still challenging due to the dynamic nature and variety of accessible EEG data [10]. Also, most studies show that the machine-learning system cannot modify the recovered attributes [11]. EEG changes are complex and nonlinear in high-dimensional space. Scientists have studied the use of EEG for AD detection throughout the last ten years [12]. EEG is a neurophysiology technique that uses scalp electrodes to record electrical impulses from the brain. This method is transportable, non-intrusive, and inexpensive. As a result, it offers a promising alternative for identifying neurological illnesses. Researchers have recently used machine learning techniques and EEG processing to identify AD [13]. These studies often aim to accomplish two key objectives: a study of AD progression and identification of AD cohorts, as the authors of this thorough review indicate. Detection studies aim to categorize clinical cohorts with clear boundaries, including Mild Cognitive Impairment (MCI) and AD [14]. In contrast, progression studies look for persons who develop AD following a long-term assessment [15].

Better results may be seen due to the EEG signal's dynamic nature and nonlinearity in return networks, and its spectrum and frequency characteristics have been described. The frequency band features were used as data processing tools [16]. Comparing this group to healthy older persons, the alpha frequency band was lowered, whereas the delta frequency band showed distinct results. These results arise from comparing brain signal anomalies in dementia patients with Levy Bodies and moderate AD patients with healthy subjects. Compared to healthy individuals, the amplitude of this frequency range was higher in mildly unwell individuals. Yet, the power spectrum can now be used as an input for categorization because of CNN's capability to learn signals and images. On the other hand, a 2D or 3D network can be created using the Fourier feature. Signal intensity and amplitude are regarded as network inputs [17]. In [23], the author has developed the novel network-based Takagi-Sugeno-Kang (N-TSK) model for identifying AD. Moreover, the features of weighted and unweighted networks were extracted with the construction of a functioning network. Hence, it does not provide accurate results for identifying AD in EEG signals. In [37], the author has investigated the modulatory efficacy of acupuncture in brain activity. During the process of acupuncture, the power spectral efficiency was initiated to analyze the EEG power. At various states, the metrics in game theory have been extracted to explore the changes in functional connectivity. Based on the heterogeneous nature, the frequencies have varied in the EEG signal. Thus, it minimizes the system's performance. In [38], the author has designed an effective classification model for various acupuncture manipulations using game theory and machine learning models. Based on the game theory, the topological features were extracted with the construction of a brain network. It suffers from handling a large number of data. To overcome these limitations, a novel method is developed to diagnose AD using EEG signals. The end-to-end learning method was designed with neural networks, which have the benefit of automatically learning complicated features from vast volumes of data.

The following is a description of the primary goals of the established classification method.

- To improve the detection of AD disease using a hybrid heuristic-aided DL model with EEG signal that helps to easily identify the tumor disorder and assists the doctor in treating the patients effectively.
- To decompose the original EEG signal using 3-level LWT for getting the significant characteristics of the signal, which aids to removes some noise factors and helps to increase the performance.
- To acquire the weighted stacked features with the help of optimal weight factors determined using the EWGLO approach, in which the RNN features and multiscale dilated CNN features are included.
- To develop the EWGLO approach for tuning the weights and epochs, optimizer, and batch size, hidden neuron count of AT-LSTM for providing promising performance.
- To frame the novel OTA-LSTM model for classifying the AD, where it infers the concept of transformer layer, attention mechanism, and LSTM. Further improvement, the parameters are optimally selected by EWGLO.
- To determine the model's performance with divergent measurements and compare it among traditional approaches. Thus, the results demonstrate how well the system classifies signals.

The parts showed the following are designated as such. Part II explains the literature review of the most recent AD categorization models. Part III covers the new category framework's architecture and its use with the database. Moreover, it provides examples of the hybrid methodology and feature selection method. In Section IV, the simulation and its results are presented. The overview of the providing infrastructure is represented in Section V.

2. Existing works

In 2021, Mazrooei *et al.* [18] have proposed a deep learning method for identifying AD in an EEG signal. Here, 40 individuals' brain circuits were captured in the Pz, Cz, and Fz modes of stimulation, recall, and open eyes. A variety of EEG signal parameters were retrieved after preprocessing. Studies have demonstrated that with advancing age and the stages of AD, the latency and amplitude of the P300 component vary after stimulation. In addition, theta band activity increased, and beta band activity dropped in the early stages of moderate Alzheimer's patients, which has been linked to a decline in alpha-band activity, and linear discriminate analysis was used to classify the participant's attributes after choosing the appropriate features. The findings demonstrated that among the three EEG signals and the four recording stages, the stimulation mode and Pz channel both displayed more accuracy than the others. Using Pz channel characteristics increased Elman neural network accuracy, improved CNN accuracy, and increased Linear Discriminant Analysis (LDA) accuracy and recall over excitation modes. According to the results, extracting suitable linear and nonlinear characteristics improved the classifiers' accuracy. Moreover, the convolutional neural network outperformed LDA and Elman because of the dynamic nature of the brain output.

In 2022, Morillas *et al.* [19] have reported the preliminary evaluation of a sixteen-channel commercial" EEG acquisition device-based self-driven AD multi-class discrimination approach". A self-driven analytic pipeline was then built to differentiate between the three groups. Automated artifact rejection techniques were first used in the EEG recordings. Finally, retrieved entropy, complexity, and features for power using the preprocessed epochs. They tested a multi-class classification problem using a multi-layer Perceptron and cross-validation with one subject left out. The preliminary results suggested that AD might be detected utilizing a self-driven strategy utilizing commercial EEG and machine learning, when supplemented with the best available literature. Additional investigation may help pave the way for AD to be identified in a single visit, lowering the expense of AD screening and maybe advancing medical therapy.

In 2019, Xiaojun and Haibo [20] have implemented a way to detect AD using deep learning. A dramatic rising trend in the incidence rate also accompanies this hot topic. This prominent hot subject area has been depicted the early identification of AD using EEG as a distinctive signature. Given the small size of the accessible EEG spectrum pictures, it's still quite difficult to figure out how to extract more abstract characteristics for improved generalization. It demonstrated the potential of multi-task active learning built on a discriminative convolutional rising Boltzmann engine with mixed feature maps to effectively address the problem. Compared to the previous design, the "Contractive Slab and Spike Convolutional Deep Boltzmann Machine (CssCDBM)" performs

Table 1
Features and challenges of existing deep learning-assisted Alzheimer disease detection methodologies.

Author [citation]	Methodology	Features	Challenges
Mazrooeiet *al.* [18]	CNN	• It's a cheap procedure that yields extremely precise outcomes. This technique allows for the initial curing of mild AD.	• It is challenging since the condition is complicated and the brain waves are dynamic.
Morillaset *al.* [19]	MLP	This approach often involves non-portable EEG gear or manual processing, which may pave the way for the adoption of forthcoming accessible and mobile early AD screening tools. .	• These factors reduce portability and necessitate more workers, making them less than ideal for automated diagnosis.
Xiaojun and Haibo [20]	CssCDBM	It is a multi-task learning strategy, It is superior to prior-based regularization techniques, with results that surpass multiple sophisticated models.	The combination of improving early AD diagnostic performance with early Stages of ad recognition, validation, and timely detection is a first effort.
V.Podgorelec [21]	STFT	• This method of EEG data analysis using the ML approach demonstrates that a resilient and adaptive ML algorithm induces well-balanced classification models, and it is a practical method for identifying AD.	• By introducing tree structure into our evolutionary method, we obtained approximate accuracy.
Nansonaet *al.* [22]	RNN	• The strategy may help analyze LT continuous EEGs for early AD identification with less computing time because it offers higher performance in terms of sensitivity and specificity.	• Additional information is needed to determine the proposed product's accuracy in considering different disease stages and taking note of treatment.It offers some prognosis indicators. It is significant to highlight that this work supports the usefulness of the proposed RNN method of training for early AD diagnosis.
Liu *et al.* [23]	N–TSK	• From the perspectives of functional networks and fuzzy systems, it demonstrates how the technique offers new concepts automatically diagnosing neurological illnesses. It also demonstrates how it may be used to discover possible markers.	To enhance the accuracy of identification and detection methods, combine quasi or unattended learning approaches with complex network methods. This will reduce their reliance on acquiring data labels.
Kulkarni [24]	WEKA	• It might introduce new technologies since there isn't a gadget comparable to this AD in the current market.In terms of classification rates and diagnosis accuracy, it provides better results.	The accurate prediction of AD is a challenging but doable undertaking.
Akşahinet *al.* [25]	DWT PSD Coherence	• It provides high-performance measurements and produces consistent and objective results without case-specific sensitivity; it would be highly economical. It is more appropriate for the elderly population.	• It is noted that even the most experienced specialists may fail in some circumstances, It may fail in some circumstances, making reviewing these medical records a lengthy process requiring competent doctors.

EEG spectrum image analysis by inducing a labeling layer, resulting in a Racist and discriminatory variation CssCDBM, known as DCssCDBM. Compared to the past, the demonstrated DCssCDBM may be expanded into a classification algorithm instead of employing extracted features alone. Then, the key innovation of the approach was to coach our DCssCDBM with a multi-task learning system using EEG spectral image-based verification and identification tasks for sweeping generalization decrease in the first example. The suggested approach outperformed various state-of-the-art techniques regarding its capacity to extract high-level representations and advanced outcomes.

In 2012, V. Podgorelec [21] reported that the early diagnosis of AD was crucial for having the biggest treatment impact. They provided a technique for using a machine learning methodology to analyze EEG signals to identify AD. It demonstrated how to combine EEG recordings data with a machine learning algorithm to create a classification model for AD. The results were quite encouraging.

In 2001, Nansona *et al.* [22] have explored the channels of 10 early-onset AD patients, and 10 control subjects were gathered, and the RNNs were utilized for training as well as testing. The RNNs was chosen because they have state memory, which was essential for the task under consideration and can implement exceedingly non-linear decision limits. The best training/testing results were attained using a three-layer Recurrent Neural Network (RNN) on level four high wavelet subbands in the left frontal channel. After being trained on recordings from three AD patients and three controls, the resulting RNN performed well when tested on all remaining control systems and Five of Seven AD patients. This demonstrated a performance that was much better than random, with roughly 80% sensitivity and 100% specificity.

In 2019, Liu *et al.* [23] have proposed the "Network-based Takagi-Sugeno-Kang (N-TSK) fuzzy classifiers' capacity". Patients with closed eyes could attain an accuracy of 97.3%, and those with eyes wide open could obtain an accuracy of 94.78%. Moreover, N-TSK performed

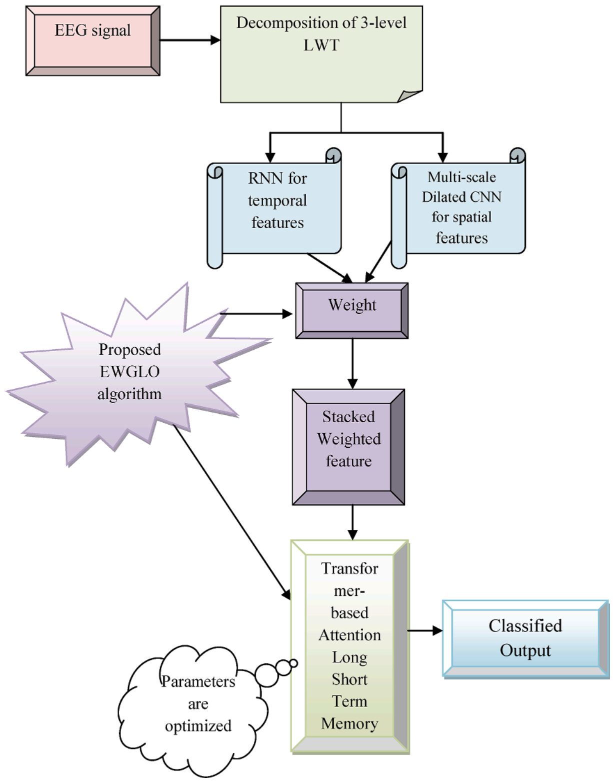


Fig. 1. A schematic diagram of novel detection for Alzheimer's Disease using EEG signal.

significantly best compared to unweighted N-TSK. Additional special tuning for the N-TSK fuzzy filters revealed that the clustering coefficient and local efficiency was the most useful parameters for AD detection. This research contributed to diagnosing and diagnosing neurological illnesses.

In 2019, Kulkarni [24] have analyzed using spectral and complexity properties; the EEG signal can be used to diagnose AD. The study used an experimental database that was obtained from a hospital. "Support Vector Machine (SVM)" and "K-Nearest Neighbor (KNN)" classifiers were used to classify the EEG relative power, assess, compare, and spectrum entropy, spectral centroid, and spectral flux to divide the data into two groups. The acquired results showed that EEG signals could be used as a baseline for diagnosing AD due to the severity shown in AD patients.

In 2021, Akşahin *et al.* [25] have proposed a method for automatically diagnosing EEG waves. Using the established ad-hoc technique, this study evaluated the EEGs of 35 participants. Three different analysis techniques made up the algorithm: coherence, "Power Spectral Density (PSD)", and Discrete Wavelet Transform (DWT). It was first applied to the signals to create the primary EEG sub-bands. Then, using Burg's method, the PSD of every post was determined. In the next method, cross-coherence values were determined. The characteristics were determined using the amplitude summations of the coherence values corresponding to the primary sub-bands and each sub-PSD band's amplitude and variability summations. Of the investigated classifiers, Packaged Trees was chosen as a classifier.

2.1. Research gaps and challenges

Memory loss and other cognitive impairments with gradual onset and progression are symptoms of AD. In most cases, it's already too late to identify the root cause of AD, and prompt treatment also requires little effort. As such, the better solution would have been the identification of AD at a preliminary phase. The challenges and features of the existing deep learning based AD disease is tabulated in Table 1. A reduced approach that produces good precision, CNN [18], can detect mild AD at a preliminary phase. Yet, it is hard to pinpoint because of the cerebral impulses' dynamic nature and the condition's intricacies. Multilayer Perceptron (MLP) [19] helps advance the upcoming acceptance of practical, portable, and expensive early AD screening methods. However, because this method typically relies on manual operations, these elements were not ideal for computer-aided diagnosis. However, this method typically relies on manual operations, and all these traits were insufficient for computer-aided diagnosis. They prevent portability and need more staff. CssCDBM [20] is a multi-task educational approach that outperforms various sophisticated ones in performance. It is superior to prior-based methods. However, this is the initial effort at enhancing early AD diagnostic performance with early AD verification, diagnosis, and detection. The ML technique for EEG data analysis is a viable method for diagnosing AD.

According to Short-Time Fourier Transform (STFT) [21], a balanced classification system can be created by utilizing a robust and adaptive ML algorithm. But with this iterative method, were able to approximate accuracy through the formation of judgment trees. RNN [22] offers greater specificity and sensitivity performance. To analyze single sustained EEGs for initial AD identification, a hybrid wavelet/RNN technique has been proposed. It's important to remember that this inquiry is a preliminary step in proving the effectiveness of the suggested RNN training method for initial AD detection.

Further studies are needed to examine the responsiveness of the proposed technique in identifying monitoring treatment, offering some prognostic indicators and multiple disease stages. The approach presents new concepts for autonomously diagnosing neurological illnesses from the standpoint of growth and advancement and fuzzy systems. N-TSK [23] is used to explore potential biomarkers. Nevertheless, unsupervised learning methods or semi-supervised ones can be combined with sophisticated network methods to enhance the efficacy of the identification and detection method and lessen its reliance on the acquisition of data labels. Since there's not a single gadget on the market for AD, Waikato Environment for Knowledge Analysis (WEKA) [24] introduces some new technologies and offers greater classifying rates and diagnostic precision. Yet, while it is feasible, providing an exact AD prediction is challenging. The DWT [25] achieves high-performance measurements, produces reliable and objective results without case-specific sensitivity, and is more applicable to the older population. Yet, analyzing these medical records requires long consultations with experienced clinicians. It is said that even the most skilled experts can make mistakes. It is said that even the most highly skilled experts can occasionally falter. The recently created deep learning-based AD detection method overcomes these difficulties.

3. Architecture representation of Alzheimer's disease detection using EEG signals: transformer-aided deep learning method

3.1. Novel detection system for Alzheimer disease

AD is a neurological ailment characterized by behavioral changes, difficulties performing everyday tasks, and cognitive problems. The computerized diagnosis of AD may be well suited for EEG signal analysis. Using spectral and complexity-based characteristics, EEG recordings in the frequency and time domain are used to diagnose AD in its early stages by using a suitable machine learning technique of EEG. EEG abnormalities in AD are a reflection of the structural as well as functional deficiencies of the diseased cerebral cortex. Dementia patients' EEG signals are less complex and have fewer functional connections, as shown by the EEG's "Nonlinear Dynamic Analysis (NDA)". The EEG has proven to be a reliable diagnostic and research tool for dementia. There are many people interested in the use of EEG in AD. The EEG helps in differential diagnosis and disease progress prediction. AD is difficult to diagnose medically, and signs are frequently written off as inevitable

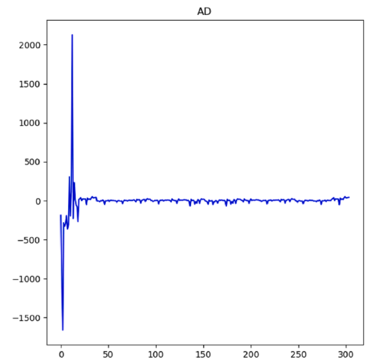
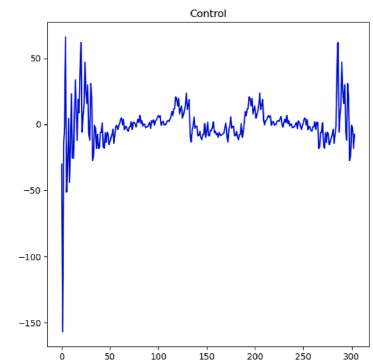
Description	Normal signal	Abnormal signal
Signal 1		

Fig. 2. Sample images of normal and abnormal EEG signals for AD detection.

side effects of age. Typically, diagnosis is carried out by a mix of in-depth examinations and the ruling out of alternative explanations. Blood tests, spinal fluid tests, neurologic exams, imaging methods, and psychological testing like the "Mini-Mental State Examination (MMSE)" are all used to assist in identifying the illness. In detecting AD using EEG signals in recent years, researchers have seen problems like increased signal complexity and disruptions in EEG synchrony. But given the high degree of diversity among AD patients, these effects are not always obvious. Hence, none of those occurrences now allow for the early, accurate diagnosis of AD. To overcome these issues, a novel deep learning based detection approach is proposed. The architecture of the designed AD detection model using EEG signal is illustrated in Fig. 1.

Initially, the EEG signal is fetched from the data sources, and the signal decomposition phase follows it. Here, the decomposed signal is attained by adopting the 3-level LWT to reduce irrelevant signal features. Subsequently, the resultant decomposed signal is given as input to the feature extraction process, where the temporal features are retrieved from RNN and spatial features are acquired from multi-scale dilated CNN. With the assistance of two features, the weighted stacked feature fusion takes place, in which the EWGLO approach optimally selects the weight factor. Finally, the fused features are subjected to the OTA-LSTM for classifying the AD. For further attaining the optimal results, the parameters like an epoch, optimizer, batch size, and hidden neuron are tuned by EWGLO, which helps increase the system efficiency. The performance is estimated using divergent measures and compared across conventional methods. Thus, the outcome demonstrates that the proposed work attains better results in diagnosing AD.

3.1.1. *Implemented dataset description*

The source EEG signal belongs to the link "https://github.com/tsyoshihara/Alzheimer-s-Classification-EEG/tree/master/data: Access Date: 2023–03-17". The data is collected from the Alzheimer-s-Classification-EEG dataset. One of the most common neurodegenerative diseases is AD. Moreover, the EEG is the most prominent non-invasive diagnostic tool for AD. Here, the patients with AD, Mild Cognitive Impairment (MCI) and healthy controls are considered for experimentation. Various frequencies were not able to localize the different regions of the brain. The collected signal is denoted I_s here $s = 1, 2, 3.....S$. In turn, the term S signifies the signal which is collected. Here, the data is splitted into two phases like, training and testing. Here, 75% of the data is given in the training phase, and the rest of the 25% of the data is utilized in the testing phase. Fig. 2 depicts the sample images of normal and abnormal EEG signals for AD detection.

3.1.2. *3-Level LWT-based signal decomposition*

After the signal is collected, the 3-level LWT is deployed for the decomposition phase. In contrast with the first-generation transform, the LWT coefficients are integers free of quantization defects. LWT has been used in some studies to increase the reliability of the signal copyrighting system. Using all of the benefits above, the copyrighted information embedding process uses LWT [26]. LWT reduces to the polyphone version of the DWT method without extra coefficients and in a zero-padding elongation mode. Since the wavelet can be used in the spatial or time domain, the Fourier transform information is unnecessary for the scheme. The LWT's construction method is the opposite of how it is broken down. It covers the unique LWT breakdown flow in this section. The precise LWT decomposition has three steps to make up the process, which are.

Stage 1: Split: It is described as a lazily implemented wavelet that separates the signal L_m into odd and even bits, denoted by the letters $Lo(m)$ and $Le(m)$ accordingly. Split is mathematically expressed in Eq. (1)

$$Le(m) = L(2m) \tag{1}$$

$$Lo(m) = L(2m+1) \tag{2}$$

From the above equation, the term m is the sample numbers and non-negative integers.

Stage 2: Predict: Both even samples and odd sampling should be predicted by even samples. The following is an expression for the discrepancy between the prediction value of $I[Le(m)]$ and the actual value of $I[Lo(m)]$. The exact equation is mathematically shown in Eq. (3)

$$y(m) = Lo(m) - I[Le(m)] \tag{3}$$

Here the term $I[.]$ is the prediction operator, and the high-frequency component is $I(m)$.

Stage 3: Update: Utilize $y(m)$ to update $I[Le(m)]$:

$$b(m) = Le(m) + Uy(m) \tag{4}$$

From this above equation $U[.]$, the operator of the updater $y(m)$ is the component of the low frequency of $L(n)$, corresponding to a low pass filter. Finally, the output of the LWT is decomposed signal L_s^{dec}. Fig 3 shows the signal decomposition diagram for 3-level LWT for detecting AD.

3.2. *Feature extraction techniques and weighted stacked features for Alzheimer's disease detection*

3.2.1. *Rnn-based temporal feature*

In this phase, the decomposed signal I_s^{dec} is the input to the RNN [22]. Since the original study vibration represents time series data, RNN is a good processing option. Before giving the input to RNN, the data are

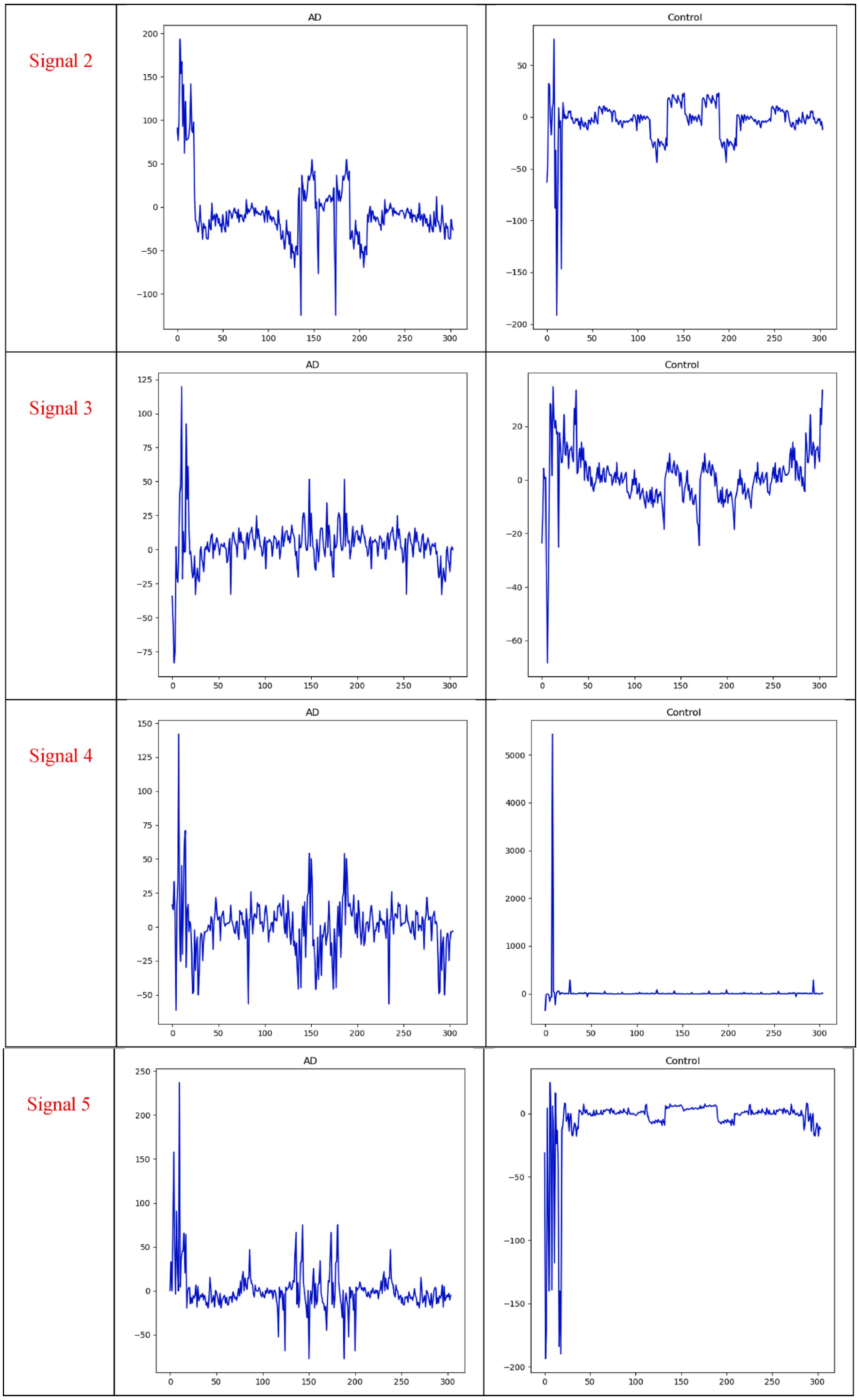


Fig. 2. (*continued*).

converted into two dimensions using the dimensions of time windows. Also, to maintain previously memorized data. With the input sequence y_m and the previous hidden state n_m, each node in this architecture retains a present hidden state n_{m-1} and produces an output. This equation is shown in Eqs. (5) and (6).

$$Y_m = I_s^{dec}(R_n n_{m-1} + f_n y_m + a_n) \tag{5}$$

$$N_m = I_s^{dec}(R_E n_m + a_E) \tag{6}$$

Here the term a is the bias for output states and hidden, R and f are the weights in recurrent connections of the hidden layers. When modeling the dynamics of a continuous data sequence, an RNN is quite successful; however, If modeling lengthy sequences, it may experience the issue of explosion and gradient disappearance. Although the RNN can maintain the relevance of the data, a longer time step increases the likelihood of gradient disappearance or explosion owing to weight reuse. Finally, temporal features are extracted from the RNN, denoted as F_s^{temp}.

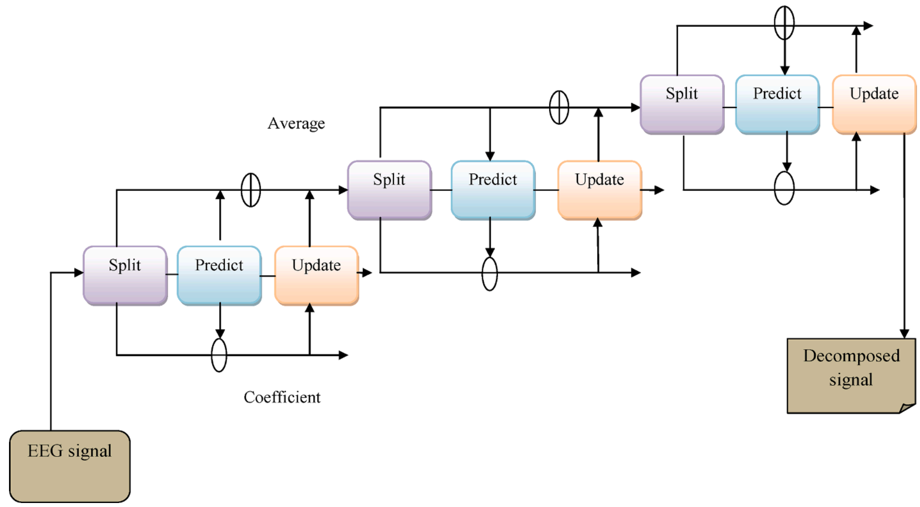


Fig. 3. A schematic diagram for 3-level LWT for signal decomposition.

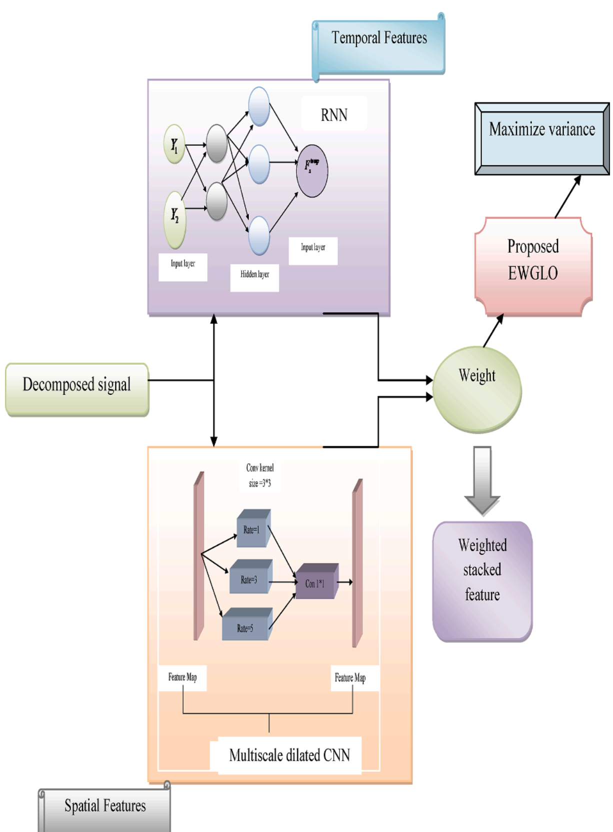


Fig. 4. A schematic diagram for a weighted stacked feature for signal decomposition.

3.2.2. Multiscale dilated CNN-based spatial features

In this phase, the decomposed signal I_s^{dec} is the multi-scale dilated CNN [27] input. For further processing, the audio signal can be transformed into two-dimensional time frequency representations. The one targeted speaker is assumed to create the received aural signal, and it is possible to converge the voice signal with the "Head-Related Impulse Responses (HRIR)" in the temporal domain. This temporal feature is shown in Eq. (7)

$$S = I_s^{dec}(m) \otimes s_m(n) + h_m(n), n \in (k, d) \tag{7}$$

Here the sign $\otimes$ represents convolution, m is the period number. Here, the term k and d represent the right and left microphone, and n is for the audio microphone. The speech signal is denoted by $l(m)$, and the head-related impulse response is denoted by $s_m(n)$ binaural signals divided into 32 auditory channels. Transfer function filters have logarithmically equal separation across their center frequency, which ranges from 80 Hz to 8 kHz on the corresponding rectangle width scale. Interaural data are processed, and the brings-to-bear value between the left and right signals for each frequency post is individually calculated. The right and left signals' correlation further generalizes +e CCF. The derivation of +e CCF for time-varying delay is defined as τ. The Cross Correlation Function (CCF) equation is shown in Eq. (8)

$$CCF_{m,g}(\tau) = \frac{M_{kd}(m,k,d)}{\sqrt{M_{kk}(m,g,0)M_{dd}(m,g,0)}} \tag{8}$$

Here g is the number of the harmonic thread, and $M_{kd}(m,k,d)$ refers to the cross-correlation between both the left and right impulses. The correlation of $m, g,$ and 0 is indicated by $M_{kk}(m,g,0)$ and $(m,g,0)$ similarly at $\tau = 0$ and are the left and right signals. The dummy head's prosthetic ears measure 15 to 17 cm in diameter. The minimum delay time is 2 ms to compensate for helmet shadowing under realistic circumstances. For instance, a matrices CCF of 32*65 is created by the brings-to-bear function of stereo sounds recorded at 16 kHz over a spectrum of center delays of 2 ms. Here, the auditory cue energy Interaural Level Difference (ILD) transfer in logarithm between signals of binaural, which would be defined as Eq. (9)

$$lld_{m,g} = 10\log_{10}\frac{\sum y \in \delta inx_r(m,k,d)}{\sum y \in \delta inx_1(m,k,d)} \tag{9}$$

Here δin is the set of all sampling indices in the nth frames, beginning with the index y. The result would be no energy since the binaural sounds are structured into brief but stable voice signal frames. This state from making any special frames would not be considered. In all frequency sub-bands, the interaural volume differential of binaural signals creates vectors ILD with a size of 32* 1. Finally, spatial features are extracted from the multiscale-dilated CNN, denoted as F_s^{spat}.

3.2.3. Weighted stacked features

The spatial and temporal features are extracted via multi-scale dilated CNN and RNN. The weighted stacked feature is accomplished with the aid of two features. The temporal and spatial features are further multiplied by their weight. The feature fusion takes place by determining the optimal weights with the help of the EWGLO approach. The number of features picked wrongly is reduced in half while the

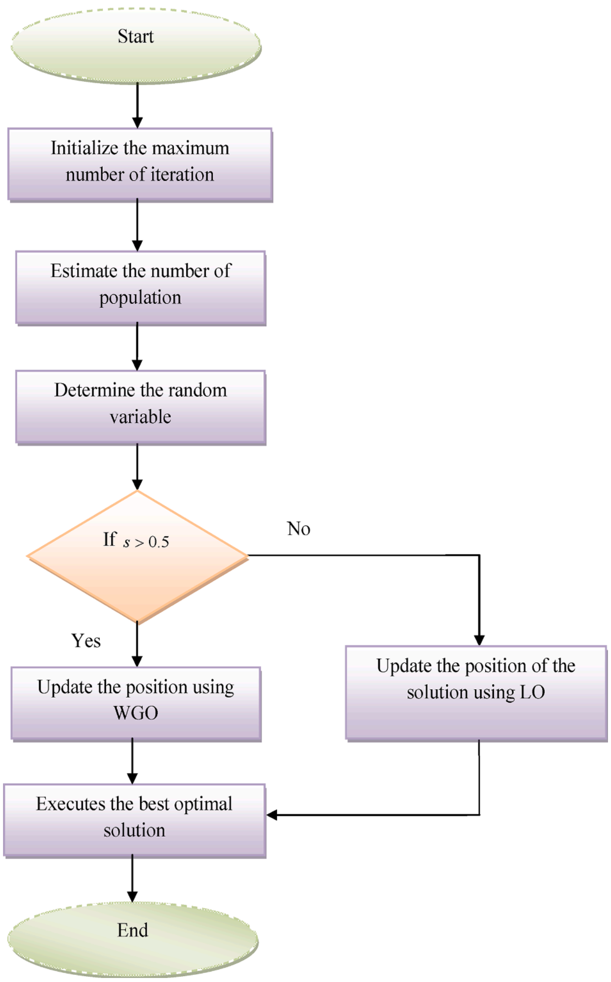


Fig. 5. Flowchart for proposed EWGLO.

actual positive rates are kept constant. This indicates that it is more effective at selecting the appropriate variables, and the final model is simpler, easier to understand, and more accurate. Combine the optimal features to create the stacked weighted feature. Which is mathematically represented and shown in Eq.(10)

$$WSF_f = \left(wt1 * F_s^{temp}\right) + \left(wt2 * F_s^{spat}\right) \tag{10}$$

The term $wt1$ and $wt2$ are the weighted features. The ranges vary from 0.01 to 0.99. The term WSF_f is the stacked weighted feature. Which is mathematically represented as Eq.(11)

$$Obj = \underset{\{wt_1 wt_2\}}{\text{argmax}}[VF] \tag{11}$$

The term $wt1$ and $wt2$ is representing the optimal weights. The term VF defines the variance. It is computed by Eq. (12).

$$variance = \sum \frac{(y_i - \bar{y})^2}{n-1} \tag{12}$$

The term y_i and $\bar{y}$ are the value of mean observation. The term n is the number of multiple observations. Fig.4 shows the schematic diagram for a weighted feature for signal decomposition.

3.3. Heuristic improvement and Optimized Transformer-based attention long Short-Term Memory for Alzheimer's disease detection

3.3.1. Hybrid heuristic Approach: EWGLO

The recently created EWGLO algorithm optimizes the parameter and boosts performance. The parameters and the optimized TA-LSTM method's hyper-parameters should be tuned primarily. This optimization can increase the effectiveness of their classification.

Grey Wolf Optimization (GWO) is a newly invented pack intelligence optimization technique extensively used in several risks. It largely impersonates the grey wolf race pack's hierarchical system and hunting tactics to achieve efficiency through these behaviors. GWO has advantages over common optimization algorithms like Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) in that it has fewer parameters, has simpler concepts, and can be applied rapidly. GWO has limitations, such as a slow convergence pace and a tendency to enter the local optimum quickly. The newly invented formula in EWGLO for the parameter s is determined by Eq. (13).

$$s = \frac{bf - cf}{Wf} \tag{13}$$

The phrase wf signifies the worst fitness and the best fitness bf; s represents the random variable, and the term cf denotes current fitness.

From the above equation, the term s represents the random variable; the term bf describes the best fitness and cf is considered as the current fitness and also wf denotes the worst fitness.

Wild Geese Migration Optimization Algorithm (WGO) [28]**:** The WGO algorithm's starting population is produced randomly in the solution space, and a predetermined number of wild geese are chosen to serve as the first head geese. The number of head geese, the size of the WGO algorithm's population, and the swarm of wild geese are the direction of the head geese as you migrate. The migration group's starting radius size is fixed, and the equation is shown in Eq. (13).

The migrating groups are reestablished with each iteration phase based on the location of the head geese. Within the s-radius, the head of the geese serves as the centre. It is mathematically represented and shown in Eq. (14)

$$\begin{cases} y_a^b = y_c^b & ada = m^*(c-1)+1 \\ y_a^b = y_c^b - s + 2s \ \ randm\ (1, \text{dim}) & else \end{cases} \tag{14}$$

Here, the term y_a^b represents the iteration position at the a_{th} individual where $a = (1, 2.....A)$ and $b = (1, 2.....K)$ represents the iteration position at the C_{th} individual where $a = (1, 2.....L)$ and $b = A/B$.

The WGO employs a synchronous flying technique, with equal flight steps for every goose in the migrating group, to mimic the flight characteristics of wild geese. The head goose provides current information for the migrating group since it bases its choices mostly on the optimal position and references to other head geese's location information. The mathematical model of the following equation is shown in Eq. (15)

$$y_a^{b+1} = y_a^b + f_1(y_{best}^b - y_c^b) + f_2(y_v^b - y_c^b) \tag{15}$$

Here, the phrase y_{best}^b represents the optimal individual global and y_v^b and y_c^b are the individual goose selection. The flight step size $f_1 \in [0, 1]$ c1 and then f_2 evaluated by using Eq. (16)

$$\begin{cases} f_2 = \exp\dfrac{f(a) - f_{avg}}{f_{worse} - f_{best}} f(a) \leqslant f_{avg} \\ f_2 = \exp\dfrac{f(a) - f_{avg} - 2f_{best}}{f_{worse} - f_{best}} f(a) > f_{avg} \end{cases} \tag{16}$$

Here $f(a)$ is the value of fitness, fitave, fittest, and head goose, and show the average, best, and worst fitness value of the head geese, respectively, and f_2 is mostly used to proportion control the experience of head geese's information. If $f(a) \leqslant f_{avg}$, It demonstrates that the value $f(a)$ is low, indicating that y_c^b is a superior head goose and does not require further information from other head geese. When the term $f(a) > f_{avg}$ the inverse is true.

Equation reduces the migrating group radius s once all the head geese have been replaced Eq. (13). The goal is to improve the group's composition by increasing the density of its members and the algorithm's accuracy in exploration. This is mathematically shown in Eq.

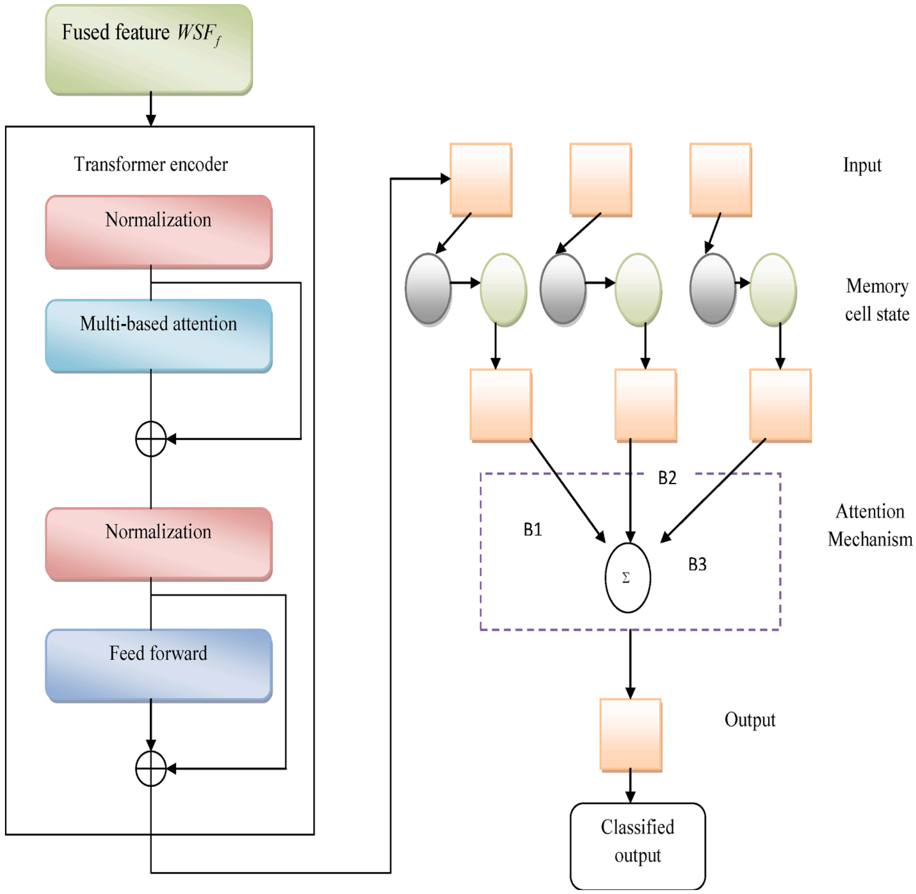


Fig. 6. Schematic representation of TA-LSTM model for AD detection.

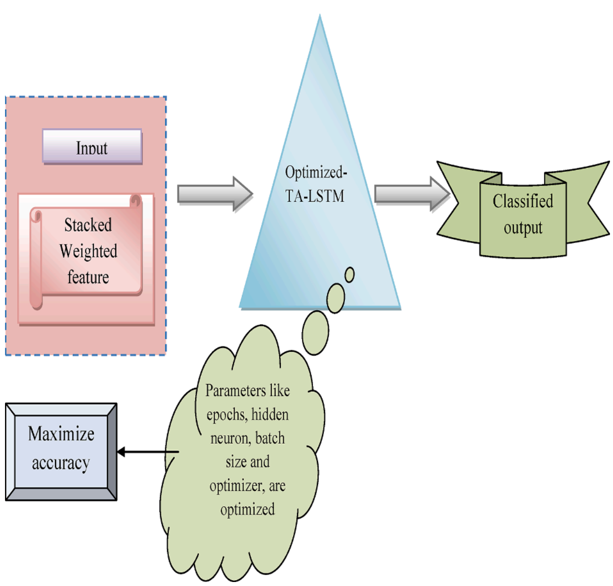


Fig. 7. Diagram for OTA-LSTM for AD detection with parameter optimization of EWGLO algorithm.

(13).

The term d is the iterations number in maximum and b is the number of iterations in current.

Lemur Optimization (LO) [29]: As a population-based algorithm, the LO algorithm represents the group of lemurs in a matrix. The following steps are taken to accomplish this. Assuming that the

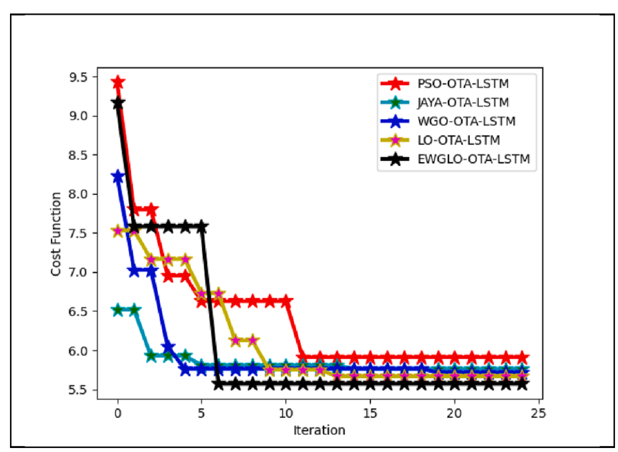


Fig. 8. Convergence performance of proposed AD detection model using EEG signal compared over other *meta*-heuristic algorithms.

population is represented by the matrix below Eq. (17)

$$M = \begin{bmatrix} s_1^1 & s_1^2 & \dots & s_1^d \\ s_2^1 & s_2^2 & \dots & s_2^d \\ \vdots & \vdots & \vdots & \vdots \\ s_m^1 & s_m^2 & \dots & s_m^d \end{bmatrix} \tag{17}$$

Here M is a population matrix of size $m \times d$, d containing a collection of lemurs d for decision variables and s for potential solutions.

The best lemur in the area was used to determine the value. This is

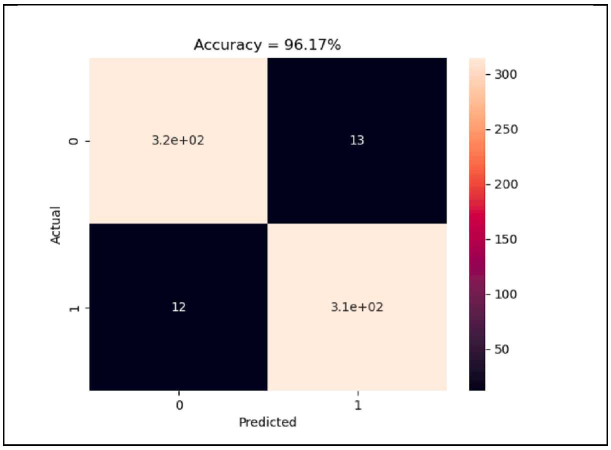


Fig. 9. Confusion matrix of proposed AD detection and classification model using EEG signal.

formatted as Equation shows. Eq. (18).

$$s_a^c = \begin{cases} s(a,c) + abs(s(a,c) - s(bnl,c)^*(randm - 0.5)^*2 \\ \qquad randm \geqslant FRR \end{cases} \tag{18}$$

Here $s(a,c)$ denotes the a value, and the current lemur is denoted as $s(bnl,c)$ denotes c is the nearest lemur, identifies the best lemur globally. The "Free Risk Rate (FRR)" shows the level of risk for all the lemurs, *randm* is a random number $[0,1]$. FRR is mathematically expressed in Eq. (19)

$$FRR = FRR(HRR) - CI \times (HRR - LRR)/MI \tag{19}$$

Here, the term *FRR* shows the free risk rate; the term *HRR* denotes the high-risk rate *LRR* as the low-risk rate. The phrase *CI* and *MI* are the maximum iteration and current iteration. Fig 5 shows the flowchart for the proposed EWGLO algorithm.

Algorithm 1: EWGLO
Consider the maximum iteration i
Find the number of population j
The parameter s is evaluated using Eq. (13)
Do
If $s > 0.5$
WGO approach
WGO position is determined
Else
LO approach
LO position is determined
End
Returns the result

3.3.2. Attention long Short-Term Memory

Combining the LSTM and the attention mechanism yields the ALSTM [30]. ALSTM networks are well suited for analyzing, making predictions, and classifying using data from time series since there may be lags of varying lengths between significant events in a time - series data. LSTM is going to be utilized to solve the vanishing gradient issue.

LSTM network: Stacked weighted feature WSF_f is the input to this phase. Using sequential data, such as EEG period, RNNs can extract high-dimension dependencies. RNN units are connected to the layers to gather data from earlier inputs. Traditional RNNs struggle to learn long-term dynamics because they worry about growing and disappearing gradients, but they can learn brief dynamics quickly.

Cells in an LSTM network provide outputs that change over time in response to information stored in the past. Long-term dependencies are maintained along the whole LSTM chain of cells thanks to a shared cell state among the cells. The course. The network can then choose whether to update the recent state L_{c-1} with the prior state L_c due to the input and forget gates M_c. The output of the search cell is controlled by an output gate U_c, enabling it to calculate its output using the current cell as input. The following LSTM cell architecture formulas are displayed in

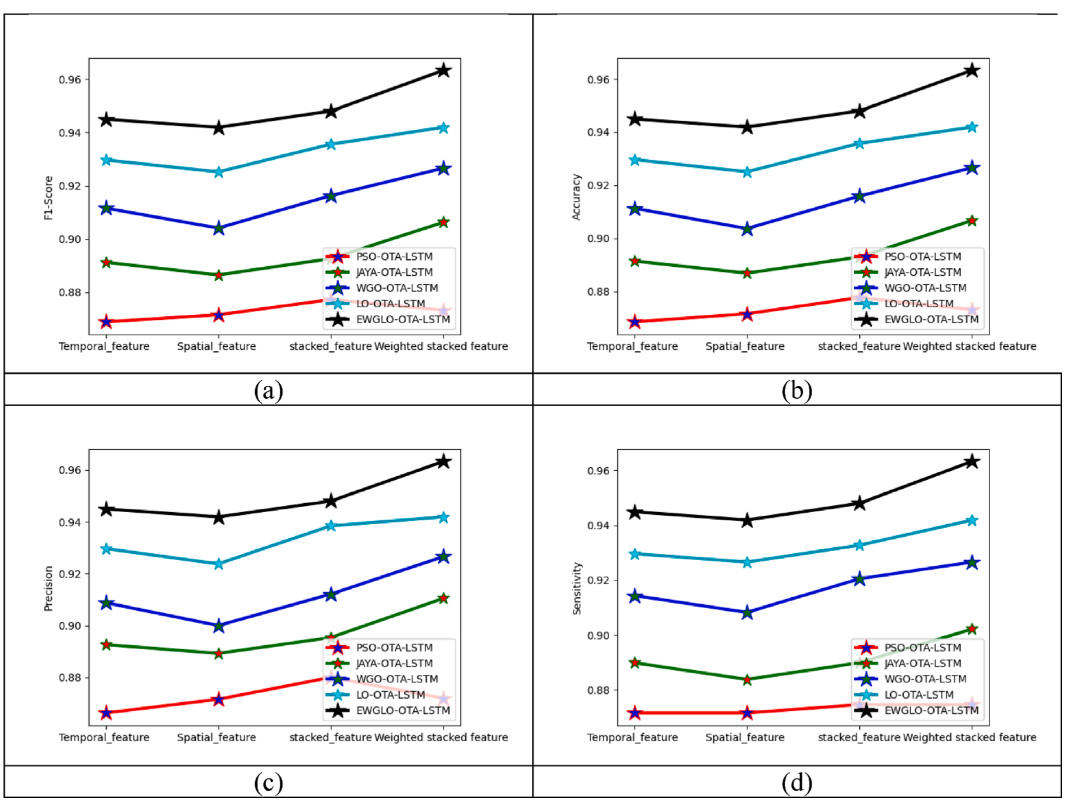


Fig. 10. Feature analysis of proposed AD detection and classification model using EEG signal compared with *meta*-heuristic algorithm models in terms of "(a) F1-score (b) Accuracy (c)Precision (d)Sensitivity".

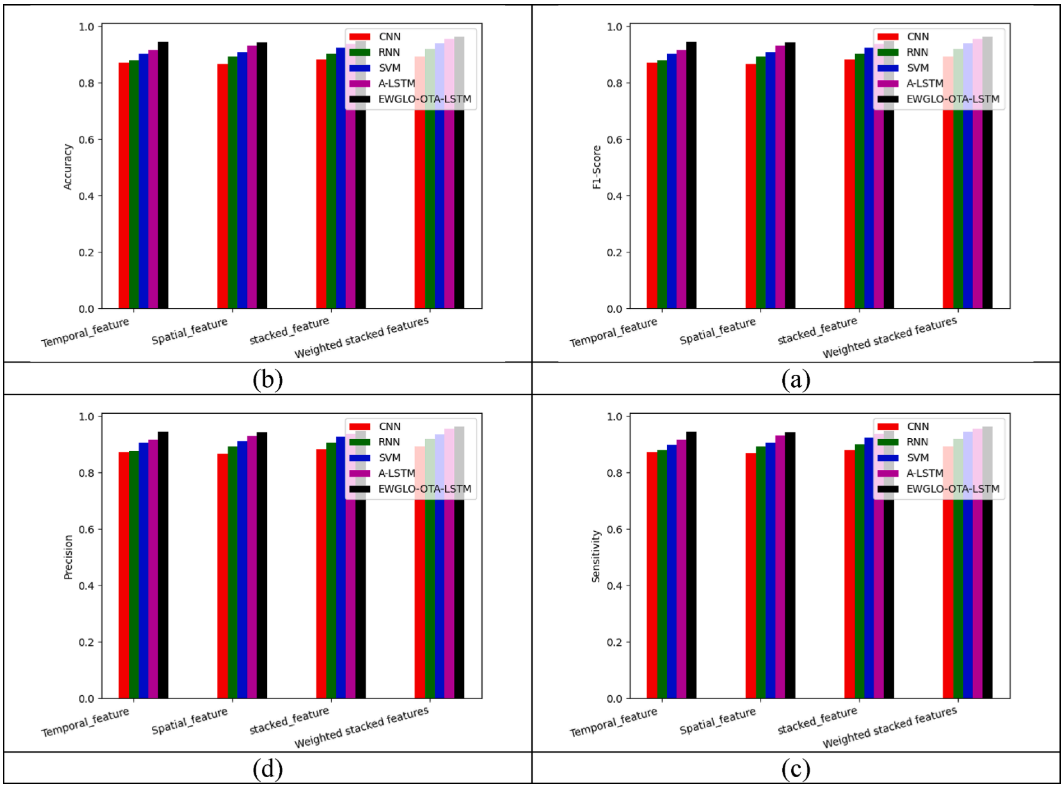


Fig. 11. Feature analysis of proposed AD detection and classification model using EEG signal compared with classifier algorithms in terms (a) Accuracy (b) F1-score (c)Precision (d)Sensitivity".

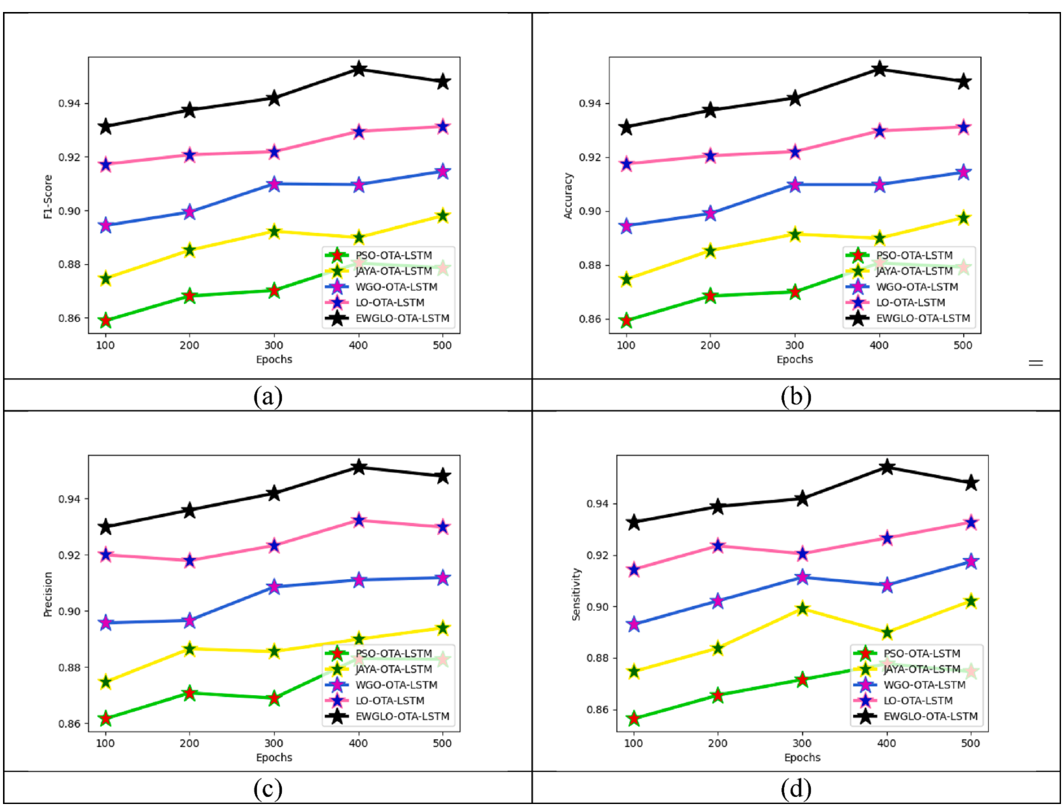


Fig. 12. Convergence analysis of proposed AD detection and classification model using EEG signal compared with *meta*-heuristic models regarding epochs in terms "(a)F1-score (b)Accuracy (c)Precision (d)Sensitivity".

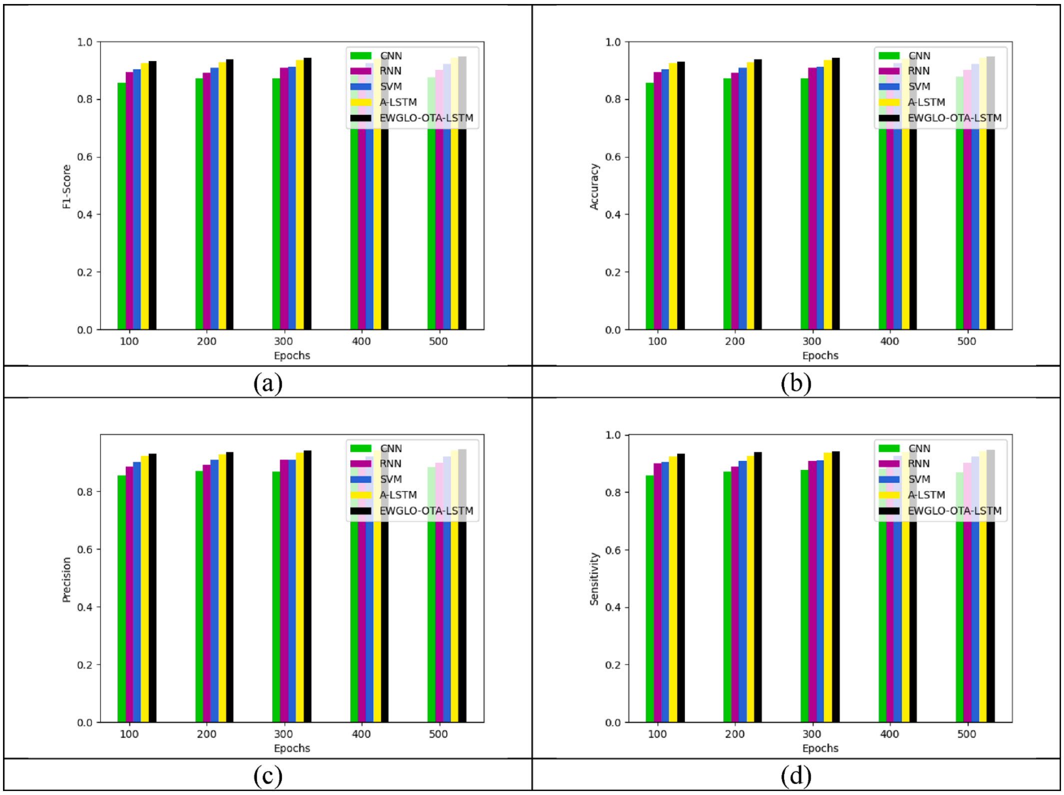


Fig. 13. Feature analysis of proposed AD detection and classification model using EEG signal compared with classifier algorithms regarding epochs in terms “(a)F1-score (b)Accuracy (c)Precision (d)Sensitivity.”.

Table 2
Statistical analysis of the proposed AD detection and classification model using EEG signal against conventional algorithms.

TERMS	PSO- OTA-LSTM[32]	JAYA-OTA-LSTM [33]	WGO-OTA-LSTM [28]	LO- OTA-LSTM[29]	EWGLO-OTA-LSTM
Worst	9.430461	6.518863	8.224892	7.531048	9.167706
Best	5.910825	5.763593	5.716434	5.668725	5.570506
Mean	6.457841	5.860187	5.963261	6.1321	6.116579
Median	5.910825	5.808632	5.763668	5.751239	5.570506
Std	0.831445	0.200883	0.578043	0.66297	1.013983

Table 3
Overall estimation of proposed AD detection and classification model using EEG signal against conventional algorithms.

TERMS	PSO- OTA-LSTM[32]	JAYA-OTA-LSTM [33]	WGO-OTA-LSTM [28]	LO - OTA-LSTM[29]	EWGLO-OTA-LSTM
Accuracy	88.3792	91.2844	92.0489	93.8838	96.1774
Sensitivity	88.685	91.1315	92.0489	0	96.0245
Specificity	88.0734	91.4373	92.0489	93.8838	96.3303
Precision	88.1459	91.411	92.0489	93.8838	96.319
FPR	11.9266	8.5627	7.9511	6.1162	3.6697
FNR	11.315	8.8685	7.9511	6.1162	3.9755
NPV	88.6154	91.1585	92.0489	93.8838	96.0366
FDR	11.8541	8.589	7.9511	6.1162	3.681
F1-Score	88.4146	91.2711	92.0489	93.8838	96.1715
MCC	76.7598	82.5692	84.0979	87.7676	98.9775

Table 4
Overall estimation of recommended proposed AD detection and classification model using EEG signal against conventional classifiers.

TERMS	CNN [18]	RNN [22]	SVM [35]	A-LSTM [30]	EWGLO-OTA-LSTM
Accuracy	89.4495	91.2844	93.2722	95.2599	96.1774
Sensitivity	89.2966	91.1315	93.2722	95.4128	96.0245
Specificity	89.6024	91.4373	93.2722	95.107	96.3303
Precision	89.5706	91.411	93.2722	95.122	96.319
FPR	10.3976	8.5627	6.7278	4.893	3.6697
FNR	10.7034	8.8685	6.7278	4.5872	3.9755
NPV	89.3293	91.1585	93.2722	95.3988	96.0366
FDR	10.4294	8.589	6.7278	4.878	3.681
F1-Score	89.4334	91.2711	93.2722	95.2672	96.1715
MCC	78.8995	82.5692	86.5443	90.5203	92.3552

Eq.(20), Eq.(21), Eq.(22), and Eq.(23).

$$j_c = \sigma(WSF_f[p_{c-1}, y_c] + a_c) \quad (20)$$

$$M_c = \sigma(WSF_f[p_{c-1}, y_c] + a_M) \quad (21)$$

$$L_c = M_t{}^*L_{c-1} + WSF_f{}^*\tan P[p_{c-1}, y_c] + a_c \quad (22)$$

$$P_c = U_c + \tanh(M_c) \quad (23)$$

The term L_{c-1} is the state of the cell at the time step t. The weight is K_l and K_m is the bias that can be derived over the training algorithm through time a_v.

Attention: The attention method can develop LSTM performance by concentrating on specific time steps with the most racist and discriminatory data. The result illustrates how, in contrast to normal LSTM networks, an LSTM network with a long short - term memory combines the broadcast hidden states by the beginning of the training. This can be stated as follows and shown in Eq. (24)

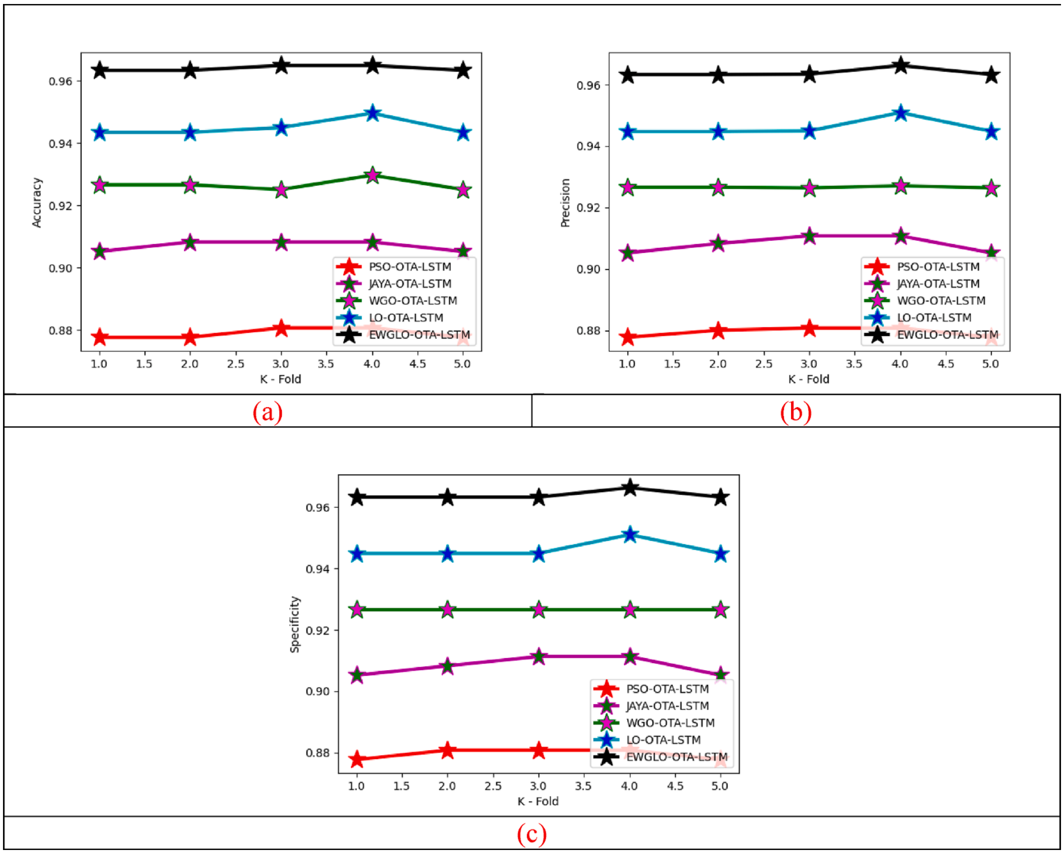


Fig. 14. Performance analysis of the designed AD model with existing algorithms regarding (a) Accuracy, (b) Precision and (c) Sensitivity.

$$M_c = LSTM(d_i), i \in [1, L] \tag{24}$$

Here M_c is the *ith* LSTM's output hidden state vector. L is the number of cells, and the following definition of attention mechanism encapsulates the significance of each concealed state:

3.3.3. Transformer-based attention long Short-Term Memory

AD is a quite common disorder, and also clinicians need to find out the AD at an initial stage. Various existing approaches are emerged to detect AD disorder. But, they are not well performed for the identification of AD. While considering CNN, the long term dependency problem cannot be solved thus, it affects the system's performance. Moreover, the existing approaches contain a limited number of factors which is ineffective to understand the complex and nature of the disease. Inorder to resolve these challenges, the Transformer-based Attention Long Short Term Memory is selected based on its effective performance. Moreover, this model helps to solve the long term dependency problem. Thus, the computational expensive is high, and also it requires more time to train the data. So, our research work optimizes the parameters of hidden neuron count, weights, optimizer, batch size and epochs in the AT-LSTM with the help of the EWGLO algorithm. That's why we improved an OTA-LSTM model for detecting AD disorder. Thus, it reduces the computation burden. In this phase weighted stacked feature WSF_f is the input of the Transformer based LSTM [31]. Here, the long-term signal dependencies can be solved by LSTM, and Transformer layers are coupled. Identification of expression in Spoken Uncompressed Audio Using a Hybrid LSTM-Transformer Model.

In this study, the positional code in the transformer architecture is swapped out for the LSTM layer. Moreover, the transformer's design utilizes spatial compression. It must re-learn every detail at each time step, which raises the computation cost. The LSTM's recurring operation maintains the input features' hidden state.

Moreover, LSTM was created to address the short attention span issue, but it is inadequate to solve the problem of longer-term dependence. Hence, the Multi-Head Attention feature of the inverter coder layer improves the model's capacity for recognizing the long-term relationships that have been modeled and considered. Multi-Head Attention can cooperate at different stages in the sequence to pay attention to data from the generated attributes. A feed-forward network layer with kernel function and layer normalization is also part of the Transformer coder layer.

In contrast to LSTM, execution is allowed in the transformer layer. This layer makes it easier for the model to pick up long-term connections. It is anticipated that combining both systems will enhance comprehension of long-term connections. Because speech empathy mostly depends on temporal information, it is appropriate for the SER model. Using linear and softmax layers yields the desired emotional response. Dropouts are added to the LSTM, Transformer encoder, and final linear layer to prevent overfitting. Get the categorized output, and lastly. Fig. 6 depicts the transformer A-LSTM model for the detection of AD disorder.

3.3.4. Optimized Transformer-based attention long Short term Memory

Identification of LSTM is a performance that learns long-term dependencies. By identifying previously unresearched long-term correlations, the performance of the LSTM-Transformer model in recognizing emotions in audio data was further examined. The position coding in the transformer architecture is swapped out for a recurrent LSTM process in the suggested LSTM-Transformer model to learn the input features' hidden layer. Also, the transformers decoder layer coupled the LSTM with the Multi-Head Long Short-Term Memory instead of employing a single attention layer on the LSTM. The disadvantages of this transformer-based LSTM include the inability to encode object orientation, and thus it requires a large amount of training data. They struggle to categorize signals with diverse locations, as described numerically, as

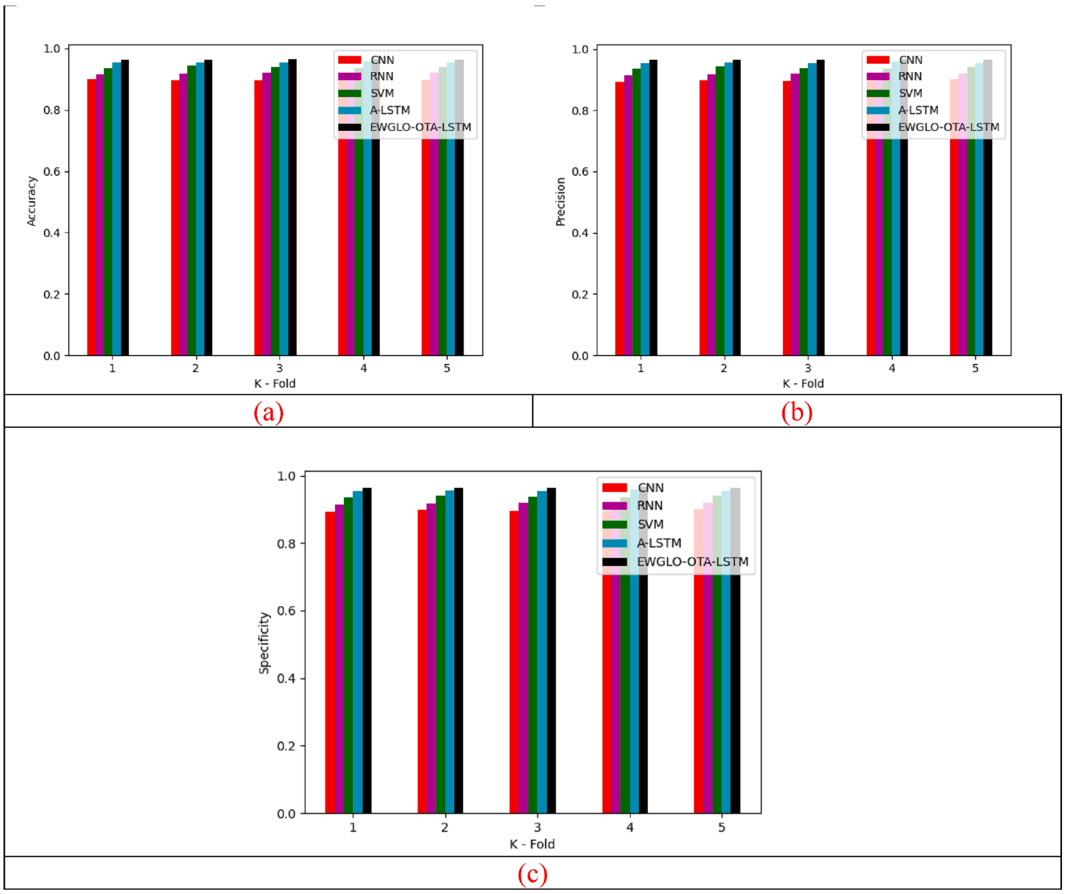


Fig. 15. Performance analysis of the designed AD model with existing classifiers regarding (a) Accuracy, (b) Precision and (c) Sensitivity.

Table 5
Comparative analysis of the recommended model for AD disorder.

TERMS	3D-CNN [39]	KNN-CNN [40]	FDN-ADNet [41]	PROPOSED
Accuracy	90.52	92.51	94.34	96.18
Recall	90.52	92.35	94.19	96.02
Specificity	90.52	92.66	94.50	96.33
Precision	90.52	92.64	94.48	96.32
FPR	9.48	7.34	5.50	3.67
FNR	9.48	7.65	5.81	3.98
NPV	90.52	92.38	94.21	96.04
FDR	9.48	7.36	5.52	3.68
F1-Score	90.52	92.50	94.33	96.17
MCC	81.04	85.02	88.69	92.36

shown in Eq. (25).

$$Obj = \underset{\{EP,HN,OP,BS\}}{\arg\max}\ [AK] \tag{25}$$

Some parameters include epochs, the optimizer, batch size, and hidden neurons. Epoch is symbolized by the terms EP “hidden layer is HN”,“ optimizer is, OP“and ”batch size is BS“taken into consideration during optimization. The hidden neuron extends from 4 to 256, while the number of epochs extends from 5 to 100. The batch size spans from 5 to 255, and the optimizer from 0 to 4. The term’s intended accuracy value is AK “closed value relating to true measure.” It is calculated by Eq. (26).

$$AY = \frac{ErA + ErB}{ErA + ErB + MaA + MaB} \tag{26}$$

The phrase nrP and mrQ represent “False Positive (FP) and False Negative (FN)” in the expression above, the term saP and saQ denotes the “True Positive (TP) and True Negative (TN)” rates. Fig.7 shows the designed OTA-LSTM for AD detection with parameter optimization.

4. Experimental results

4.1. Simulation setup

The Python tool was used to carry out the indicated AD detection and associated discoveries. Ten population estimations were used to determine the proposed action method, and 25 iterations were permitted. The performance was validated using a variety of metrics. Current algorithms include Particle Swarm Optimization (PSO)-OTA-LSTM [32], JAYA-OTA-LSTM [33], WGO-OTA-LSTM [28], and LO-OTA-LSTM [29]. Similar to this, additional models employed in the classifier comparison were CNN [18], RNN [22], Support Vector Machine (SVM) [34], and A-LSTM [30,36].

4.2. Performance measures

Accuracy: The accuracy value is equated by Eq. (26).

Specificity: It defines the average of fault data omitted through the signal detection which is shown in Eq.(27)

$$speci = \frac{ErB}{ErB + MaB} \tag{27}$$

FTR and False Negative Rate (FNR): It expressed the measure of fault responses that is calculated using Eq. (28).

$$FTR = \frac{MaA}{MaA + EaB} \tag{28}$$

$$FNP = \frac{ErB}{EaB + MrA} \tag{29}$$

Recall*:* It denoted as the correctly detected outcome among the diverse signals which is calculated using Eq. (30).

$$Reca = \frac{ErA}{ErA + MaB} \tag{30}$$

False Discovery Rate (FDR): It depicts the ratio of fault positives within the actual values and also "the FP, TP, and FP ratio is employed to calculate the false alarm rat".

$$FDR = \frac{MaA}{ErA + EaA} \tag{31}$$

Net Present Value (NPV): It refers the fault values from the detection model which is calculated using Eq. (32).

$$NPV = \frac{ErA}{MrB + MaB} \tag{32}$$

Matthews Correlation Coefficient (MCC): It is a correlation coefficient of the variables that is calculated using Eq. (33). ".

$$MCC = \frac{ErA \times ErB - MaA \times MaB}{\sqrt{(ErA + MaA)(ErA + MaB)(ErA + MaB)(ErA + MaB)}} \tag{34}$$

4.3. Comparative analysis of the proposed AD detection and classification model using EEG signal over existing methodologies

Fig. 8 compares the cost analysis of the suggested classification model to that of different algorithms with variable cost percentages. The suggested EWGLO-OTA-LSTM is acquired the less cost function than 10%, 0.8%, 0.4%, and 4.1% for PSO-OTA-LSTM, JAYA-OTA-LSTM, WGO-OTA-LSTM, and LO-OTA-LSTM, respectively, when the iteration is 10. Thus, the more positive results assist in diagnosing the disease significantly.

4.4. Confusion matrix of the proposed AD detection and classification model using EEG signal

Fig. 9 compares the suggested categorization model versus several techniques with varying accuracy percentages. The confusion matrix's AD signal categorization is 96.17%. As a result, the lower error rates show that the suggested model has better categorization capabilities.

4.5. Stacked feature analysis on the proposed AD detection and classification model using EEG signal over existing methodologies

The novel model's performance analysis compared to traditional optimizations is shown in Fig. 10, and Fig.11 shows the performance analysis compared to classifier optimization. The F1-score performance of the suggested technique is shown in Fig. 10 (a). The system acquires a higher value as the model estimates weighted stacked features for classification. The F1-score of the proposed system attains maximum value rather than 10% of PSO-OTA-LSTM, 2.5% of JAYA-OTA-LSTM, 5.3% of WGO-OTA-LSTM and 3.4% of LO-OTA-LSTM, respectively. Thus, the more positive results assist in diagnosing the disease significantly.

4.6. Comparative analysis of the proposed AD detection and classification model using EEG signal over existing methodologies

The novel model's performance analysis compared to traditional optimization regarding epochs is shown in Fig. 12, and Fig.13 shows the performance analysis compared to classifier optimization. The F1-score performance of the suggested technique is shown in Fig. 12(a). The suggested work's F1-score is 30% higher than PSO-OTA-LSTM, 40% higher than JAYA-OTA-LSTM, 39% higher than WGO-OTA-LSTM, and 45% higher than LO-OTA-LSTM during the 100th epoch. These findings assure the system gains more positive values to diagnose AD disease.

4.7. Statistical performance of proposed AD detection and classification model using EEG signal over existing methodologies

The statistical analysis of the suggested model in comparison to conventional optimizations is explained in Table 2. Five different measures are used to measure these things. "The maximum and minimum rates are the best and worst values. The mean refers to the average value, whereas the median refers to the midpoint. The degree of variance between each execution is the final definition of the standard deviation". The table findings assure that the system gains more positive values to diagnose AD disease.

4.8. Overall performance of proposed AD detection and classification model using EEG signal over existing optimization models

Table 3 provides the total estimation of the proposed AD detection model utilizing the current optimization strategies. The suggested method's accuracy value in the dataset is greater than 4% of PSO-OTA-LSTM, 1% of JAYA-OTA-LSTM, 2% of WGO-OTA-LSTM, and 1.7% of LO-OTA-LSTM, respectively. Furthermore, the remarkable results make it possible for the suggested method to be more effective at detecting diseases.

4.9. Overall performance of proposed AD detection and classification model using EEG signal over existing methodologies

Table 4 provides the total estimation of the proposed AD detection model utilizing the current optimization strategies. The sensitivity for the suggested technique is greater than 4.18% of CNN, 3.11% of RNN, 2.03% of A-LSTM, and 1.07% of SVM, respectively; thus, the outstanding results make it possible for the suggested approach to effectively diagnose AD disorders.

4.10. Cross validation of the designed AD model

The cross validation of the developed AD model using different algorithms and classifiers are shown in Fig. 14 and 15. The performance of the designed model shows 9.3%, 6.2%, 4.1% and 2.2% superior performance than PSO, JAYA, WGO, and LO regarding accuracy. Thus, the simulation outcome of the developed model attains better performance when compared with traditional methods.

4.11. Comparative analysis of the offered method

The comparative analysis of the recommended EWGLO-OTA-LSTM model is validated with recent existing approaches, and it is tabulated in Table 5. Moreover, the performance of the offered model attains 6.4%, 3.9% and 1.9% enriched performance than 3D-CNN, KNN-CNN and FDN-ADNet regarding precision. Throughout the analysis, the developed model performs better than existing recent methods.

5. Conclusion

This paper has developed a novel development of transformer-based attention long short-term memory network for detecting AD using EEG signal. The necessary EEG signal was initially sourced from conventional internet sources, and it was then subjected to the 3-level LWT decomposition to divide the signal into many wavelets. RNN and multi-scale dilated CNN were used to extract the spatial features and temporal information from the decomposed signal. The developed EWGLO algorithm was then used to determine weights for the fusion process. The OTA-LSTM model was used to detect AD in the final detection stage using these weighted stacked features, where the parameters were tuned by using EWGLO. Lastly, the performance was compared using traditional approaches and assessed across several performance parameters. According to the results, the accuracy of the suggested EWGLO-OTA-LSTM is 4% greater than CNN, 2.6% greater than RNN, 2.5% greater than SVM, and 0.2% greater than A-LSTM. The improved outcomes thus show that the suggested approach has produced the desired results for preventing AD illnesses. Thus, the designed model helps to detect AD effectively. Consequently, the result provides an accurate outcome thus; the clinicians treat the patients who are affected by the AD disorder. A few limitations of the developed model are listed below. Here, the designed model has limited with the number of subjects while detecting the AD disorder. Thus, the emission of signals from EEG is quite cause harm to the patient while detecting AD. In future work, advanced techniques will be developed with the help of ensemble models to enhance the performance. Additionally, the MRI and PET images will be considered and evaluated in the future for detecting AD more effectively.

6. Clinical applications

In clinical practice, AD becomes the most neurodegenerative disorder, and also various clinical researches have been performed early. Moreover, the designed model provides some clinically applied treatments that can diagnose AD symptoms in a short period of time. In the era of the clinical setting, AD identification can be differentiated in the form of neurodegenerative dementia like frontotemporal dementia (FTD) and dementia with Lewy bodies (DLB). Hence, effective treatment should be provided for AD patients in an accurate manner.

Declaration of Competing Interest

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

Data availability

No data was used for the research described in the article.

References

[1] G. Defazio, F. Federico, V. Lepore, P. Livrea, F. Roselli, B. Tartaglione, Rate of MMSE score change in Alzheimer's disease, Clinical Neurol. Neurosurg 111 (2009) 327–330.
[2] B. Dogan, B. Tasci, Primate brain pattern-based automated Alzheimer's disease detection model using EEG signals, Cognitive Neurodynamics (2022).
[3] S. Berent, K.A. Frey, N.L. Foster, B. Giordani, D.E. Kuhl, S. Minoshima, Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease, Ann. Neurol. 42 (1) (1997) 85–94.
[4] R. Cuingnet, E. Gerardin, J. Tessieras, et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI, Database, "NeuroImage 56 (2) (2011) 766–781.
[5] R. Cassani, T.H. Falk, Alzheimer's Disease Diagnosis and Severity Level Detection Based on Electroencephalography Modulation Spectral "Patch" Features, IEEE J. Biomed. Health Inform. 24 (7) (July 2020) 1982–1993.
[6] H. Aghajani, E. Zahedi, M. Jalili, A. Keikhosravi, B.V. Vahdat, Diagnosis of Early Alzheimer's Disease Based on EEG Source Localization and a Standardized Realistic Head Model, IEEE J. Biomed. Health Inform. 17 (6) (Nov. 2013) 1039–1045.
[7] ÖmerFarukAlçin SiulySiuly, EnamulKabir, AbdulkadirŞengür, Hua Wang, Yanchun Zhang, and Frank Whittaker, "A New Framework for Automatic Detection of Patients With Mild Cognitive Impairment Using Resting-State EEG Signals,", IEEE Trans. Neural Syst. Rehabil. Eng. 28 (9) (Sept. 2020) 1966–1976.
[8] S. Khatun, B.I. Morshed, G.M. Bidelman, A Single-Channel EEG-Based Approach to Detect Mild Cognitive Impairment via Speech-Evoked Brain Responses, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (5) (May 2019) 1063–1070.
[9] G. Henderson, E. Ifeachor, N. Hudson, C. Goh, N. Outram, S. Wimalaratna, C. Del Percio, F. Vecchio, Development and assessment of methods for detecting dementia using the human electroencephalogram, IEEE Trans. Biomed. Eng. 53 (8) (Aug. 2006) 1557–1568.
[10] H.A. Gonzalez, S. Muzaffar, J. Yoo, I.M. Elfadel, BioCNN: A Hardware Inference Engine for EEG-Based Emotion Detection, IEEE Access 8 (2020) 140896–140914.
[11] A.R. Aslam, M.A.B. Altaf, An On-Chip Processor for Chronic Neurological Disorders Assistance Using Negative Affectivity Classification, IEEE Trans. Biomed. Circuits Syst. 14 (4) (Aug. 2020) 838–851.
[12] H.A. Gonzalez, R. George, S. Muzaffar, J. Acevedo, S. Höppner, C. Mayr, J. Yoo, F. H.P. Fitzek, I.M. Elfadel, Hardware Acceleration of EEG-Based Emotion Classification Systems: A Comprehensive Survey, IEEE Trans. Biomed. Circuits Syst. 15 (3) (June 2021) 412–442.
[13] H. Kassiri, A. Chemparathy, M.T. Salam, R. Boyce, A. Adamantidis, R. Genov, Electronic Sleep Stage Classifiers: A Survey and VLSI Design Methodology, IEEE Trans. Biomed. Circuits Syst. 11 (1) (Feb. 2017) 177–188.
[14] P. Campisi, D.L. Rocca, Brain waves for automatic biometric-based user recognition, IEEE Trans. Inf. Forensics Secur. 9 (5) (May 2014) 782–800.
[15] H.-S. Chiang, S.-C. Pao, An EEG-Based Fuzzy Probability Model for Early Diagnosis of Alzheimer's Disease, J. Med. Syst. vol. 40, Article number: 125 (2016).
[16] A.A. SamanFouladi, N.M. Safaei, FoadGhaderi & M. J. Ebadi,, Efficient Deep Neural Networks for Classification of Alzheimer's Disease and Mild Cognitive Impairment from Scalp EEG Recordings, Cogn. Comput. 14 (2022) 1247–1268.
[17] B. Deng, LihuiCai, Shunan Li, Ruofan Wang, Haitao Yu, Yingyuan Chen & Jiang Wang, "Multivariate multi-scale weighted permutation entropy analysis of EEG complexity for Alzheimer's disease,", Cogn. Neurodyn. 11 (2017) 217–231.
[18] M. Rad, M. Azarnoosh, M. Ghoshuni, M.M. Khalilzadeh, Diagnosis of mild Alzheimer's disease by EEG and ERP signals using linear and nonlinear classifiers, Biomed. Signal Process. Control 70 (2021), 103049.
[19] C.M. Gutiérrez, E. Perez-Valero, I. Carrera-Muñoz, M.Á. Lopez-Gordo, R. M. Vílchez-Carrillo, A self-driven approach for multi-class discrimination in Alzheimer's disease based on wearable EEG, Comput. Methods Programs Biomed. 220 (2022), 106841.
[20] W.H. BiXiaojun, "Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning, "NeuralNetworks 114 (2019) 119–135.
[21] V. Podgorelec,"" Analyzing EEG Signals with Machine Learning for Diagnosing Alzheimer's Disease," Faculty of Electrical Engineering and Computer, Science 18 (8) (2012) 1392–11215.
[22] W. Lajara-Nansona, A.A. Petrosiana, D.V. Prokhorov, R.B. Schiffer, Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG, Clin. Neurophysiol. 112 (2001) 1378–1387.
[23] C. Liu, Y.u. Haitao, J. Wang, X. Lei, Z. Song, Supervised Network-based Fuzzy Learning of EEG Signals for Alzheimer's identification, Transactions on Fuzzy Systems (2019) 2903753.
[24] Nilesh Kulkarni, "EEG Signal Analysis for Mild Alzheimer's Disease Diagnosis using Spectral- and Complexity-Based Features and Machine Learning Techniques," Advances in Intelligent Systems and Computing, pp.828, 2019.
[25] BurcuOltu Akşahin, M. Feyzi, SedaKibaroğlu,, A novel electroencephalography based approach for Alzheimer's disease and mild cognitive impairment detection, Biomed. Signal Process. Control 63 (2021), 102223.
[26] M. Talbi, R. Baazaoui, and M. S. Bouhlel, "A New Technique of ECG Denoising based on LWT and Total Variation Minimization ECG Denoising," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-6.
[27] Li, Fei Qi, Guangming Shi, and Chunhuan Lin" A multiscale dilated dense convolutional network for saliency prediction with instance-level attention competition"Journal of Visual Communication and Image Representation. Vol. 64, October 2019.
[28] W.u. Honggang, X. Zhang, L. Song, Y. Zhang, LidongGu, and Xiaonan Zhao, "Wild Geese Migration Optimization Algorithm: A NewMeta-Heuristic Algorithm for Solving Inverse Kinematics of Robot," Computational Intelligence and Neuroscience, Article ID 5191758 (2022) 38.
[29] A.K. Abasi, S.N. Makhadmeh, M.A. Al-Betar, O.A. Alomari, M.A. Awadallah, Z.A. A. Alyasseri, I.A. Doush, A. Elnagar, E.H. Alkhammash, M. Hadjouni, Lemurs Optimizer: A NewMetaheuristic Algorithm for Global Optimization, Appl. Sci. 12 (2022) 10057.
[30] H. Zhang, Q. Zhang, S. Shao, T. Niu, X. Yang, "Attention-Based LSTM Network for Rotatory Machine Remaining Useful Life Prediction, "IEEE, Access 8 (2020) 132188–132199.
[31] F. Andayani, L.B. Theng, M.T. Tsun, C. Chua, Hybrid LSTM-Transformer Model for Emotion Recognition From Speech Audio Files, IEEE Access 10 (2022) 36018–36027.
[32] R. Poli, James Kennedy, and Tim Blackwell"Particle swarm optimization"Swarm, Intelligence 1 (2007) 33–57.
[33] H. Das, B. Naik, H.s., Behera"A Jaya algorithm based wrapper method for optimal feature selection in supervised classification"Journal of King Saud University -, Computer and Information SciencesVol. 34, Issue 6, Part B (June 2022) 3851–3863.
[34] Corinna Cortes and Vladimir Vapnik"Support-vector networks"Machine Learning volume 20, pages273–297 1995.
[35] Y. Li, Y. Yang, C. Yang, and B. Zhang, "A-LSTM Model for Predicting the Deaths Caused by COVID-19 in U.S.," 2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Yibin, China, 2021, pp. 287-290.
[36] S. Rajeyyagari, Automatic speaker diarization using deep LSTM in audio lecturing of e-Khool Platform, Journal of Networking and Communication Systems 3 (4) (2020) pp.
[37] Y.u. Haitao, W.u. Xinyu, L. Cai, B. Deng, Jiang Wang "Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation,", IEEE Trans. Neural Syst. Rehabil. Eng. 26 (5) (2018) 977–986.
[38] Y.u. Haitao, X. Li, X. Lei, Jiang Wang "Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals,", IEEE Trans. Neural Syst. Rehabil. Eng. 27 (10) (2019) 1973–1984.

[39] M.A. Shakir, I.U. Khan, Y.-K. Ma, M.d. Sadek Ali, "On Improved 3D-CNN-Based Binary and Multiclass Classification of Alzheimer's Disease Using Neuroimaging Modalities and Data Augmentation Methods," Journal of Healthcare, Engineering (2022).

[40] M.G. Lanjewar, J.S. Parab, A.Y. Shaikh, Development of framework by combining CNN with KNN to detect Alzheimer's disease using MRI images, Multimed. Tools Appl. 82 (2023) 12699–12717.

[41] R. Sharma, M. Tripti Goel, R.M. Tanveer, FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans, Appl. Soft Comput. 115 (2022), 108099.