

## Enhanced Long Short Term Memory for Early Alzheimer's Disease Prediction

M. Vinoth Kumar<sup>1,\*</sup>, M. Prakash<sup>2</sup>, M. Naresh Kumar<sup>3</sup> and H. Abdul Shabeer<sup>4</sup>

<sup>1</sup>Department of Information Science and Engineering, Dayananda Sagar Academy of Technology & Management, Kanakapura Main Road, 560082, Bangalore, India

<sup>2</sup>Department of Data Science and Business Systems School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, 603203, Tamilnadu, India

<sup>3</sup>Department of ECE, Vardhaman College of Engineering, Hyderabad, 501218, India

<sup>4</sup>IBM India Pvt Ltd., DLF IT Park, Chennai, 600125, India

\*Corresponding Author: M. Vinoth Kumar. Email: mailvinoji@gmail.com

Received: 29 November 2021; Accepted: 13 February 2022

**Abstract:** The most noteworthy neurodegenerative disorder nationwide is apparently the Alzheimer's disease (AD) which has no proven viable treatment till date and despite the clinical trials showing the potential of preclinical therapy, a sensitive method for evaluating the AD has to be developed yet. Due to the correlations between ocular and brain tissue, the eye (retinal blood vessels) has been investigated for predicting the AD. Hence, an enhanced method named Enhanced Long Short Term Memory (E-LSTM) has been proposed in this work which aims at finding the severity of AD from ocular biomarkers. To find the level of disease severity, the new layer named precise layer was introduced in E-LSTM which will help the doctors to provide the apt treatments for the patients rapidly. To avoid the problem of overfitting, a dropout has been added to LSTM. In the existing work, boundary detection of retinal layers was found to be inaccurate during the segmentation process of Optical Coherence Tomography (OCT) image and to overcome this issue; Particle Swarm Optimization (PSO) has been utilized. To the best of our understanding, this is the first paper to use Particle Swarm Optimization. When compared with the existing works, the proposed work is found to be performing better in terms of F1 Score, Precision, Recall, training loss, and segmentation accuracy and it is found that the prediction accuracy was increased to 10% higher than the existing systems.

**Keywords:** Alzheimer's disease; enhanced LSTM; particle swarm optimization; OCT image; boundary detection

### 1 Introduction

Alzheimer's disease (AD) is the worse condition disease as 10% of aged peoples nervous systems were getting affected by this disease and the dementia was caused due to the AD globally. Every year about 47 million populations of people were getting affected by the dementia disorder and this count may get increased to 131 million populations of people by the year 2050 as the numbers of aged people were



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

increasing every year. The accumulation of extracellular amyloid-amyloid- $\beta$  ( $A\beta$ ) plaques and the intracellular hyper phosphorylated tau were collectively called as Neurofibrillary Tangles (NFTs).

Some of the main diagnostic tools available with higher accuracy level are clinical evaluation/invasive tests, and the subjective methods. Postmortem detection of extracellular intra-neuronal NFTs and  $A\beta$  plaques will confirm the definite diagnosis of AD. For about 10%–15% of the existing clinical methods were found to have some inaccurate degrees of specificity, clinical diagnoses, and sensitivity. The above tools will be highly invasive and expensive for the purposes of diagnosis and screening as their applicability were found to be limited and hence, earlier definitive diagnosis of AD and less expensive methods, and new non-invasive methods have been required to serve best for the people in trouble.

The ocular examination section is found to be a promising one to develop a diagnostic tool and a novel non-invasive screening and the origin of the brain and eyes were found to be similar in embryological manner. The front neural tube forms the eyes and gives a raise to the forebrain and in several aspects, retinal neurons were found identical to neurons in the cerebral brain. Similar to those in the cerebral cortex, retinal neurons will also form the most complicated neural networks. AD will produce the functional and structural modifications not only in the vasculature and the retinal neurons but also in the retinal neurons thereby developing the abnormalities in the brain that are neurodegenerative. As suggested, the ocular manifestations might be used as an early biomarkers of AD.

The similarity between cerebral tissue and ocular is that the ocular symptoms might be employed as the early indicators for AD. Variations of ocular tissues both in neural and non-neural tissues including the degeneration of retinal axonal and neural tissue and an accumulation of  $A\beta$  were demonstrated in different part of studies in AD patients. The existing methods considered the evaluation of retinal vasculature and pupillary responses along with the AD biomarkers depending upon the blood flow. A sensitive screening biomarker for AD measured earlier in the illness's progression would pave the way for new possible therapeutics thereby helping to identify the population at risk, and improve the care of AD patients. The scan results of Optical Coherence Tomography Angiography (OCTA) were observed in this paper which showed that when compared with the healthy people, the individuals suffering from AD have exhibited the loss in the smaller retinal blood vessels with the reduced perfusion density (PD) and superficial capillary plexus vessel density (VD). The macular ganglion cell-inner plexiform (GC-IPL) layer of the retina is found thinner in AD patients rather than in the control group. The PD and VD, imaging biomarkers which are useful in screening for AD in symptomatic individuals has shown the difference in the densities of the blood vessels. The Changes in AD or the changes occurred in the retinal microvasculature might be a mirror small-vessel cerebrovascular and hence, these parameters might be serving as a surrogate non-invasive biomarkers for the diagnosis of the AD [1,2].

One of the biggest challenges faced by the Alzheimer's experts is that no reliable treatment has been found available for AD diagnosis till date. The Computer-Aided System (CAD) is used for accurate and early detection of AD to avoid the high care costs of the AD patient which are expected to rise dramatically. The traditional machine learning techniques have been used in the earlier diagnosis of the AD that typically takes the advantage of two types of features, namely, the region of interest (ROI)-based features and the voxel-based features. More specifically, they were highly relying upon the basic assumptions such as regional cortical thickness, and gray matter volume regarding the structural or functional anomalies in the brain. The traditional methods will depend upon manual feature extraction which is heavily relying upon the technical experience and repetitive attempts that appears to be time-consuming and subjective [3]. As a result, deep learning model is considered to be an effective way to overcome these problems and hence, this work uses Convolutional Neural Network (CNN), LSTM, and DenseNet for the diagnosis of AD. The contributions of the work are as follows. The contributions of the work are as follows.

- The boundary detection should be accurate while segmenting the OCTA scan image and hence, PSO is used for accurate retinal boundary detection as this PSO is more suitable and generic for all OCTA images.
- The Enhanced Long Short Term Memory (E-LSTM) is proposed to find the severity of the disease from ocular biomarkers and in the proposed E-LSTM, a new layer named precise layer was introduced to find the level of severity of the disease which in turn helps the doctors to provide the patients with the necessary treatments rapidly. Also, a dropout has been added to LSTM for avoiding the problem of over fitting.
- Convolutional network layers are used for extracting the features of an eye in robust manner thereby putting back the fully connected layers in the convolutional layer. One of the advantages of this is that the numbers of parameters to be adjusted were decreased as the weights are shared in the convolutional layer.

This work was planned as follows in which the Section 2 discusses about the literature survey of the previous works whereas the proposed system which predicts the AD from ocular biomarkers is described in the Section 3. The results were discussed in Section 4 and presented for conclusion in Section 5.

## 2 Related Works

A framework for classifying the medical image and detecting the AD was proposed by Healy et al. [3] based on the deep-learning CNN architectures. Four stages of AD are multi-classified which has achieved a promising accuracy of 93.61% and 95.17% for 2D and 3D multi-class AD stage classifications. The VGG19 pre-trained model is fine-tuned with an accuracy of 97% for multi-class AD stage classifications. Alloui et al. [4] studied the work for automating the detection of disease diagnosis and the damaged areas and also, evaluated the methods for demonstrating whether the proposition performed was effective and reliable or not by using the ample cases from public databases. But, thus work doesn't evaluated the results on the larger datasets.

Chitradevi et al. [5] introduced an algorithm for identifying the AD automatically by segmenting the cerebral sub regions. Hao et al. [6] proposed a multi-class study where the AD, NC, ADNI, and Mild Cognitive Impairment (MCI) samples are used for the purpose of differentiating. This method has attained an accuracy of 95% for the classification of multimodal AD with the applied selection points and the threshold. A multi-class analysis for predicting AD was introduced by Chihun Park et al. [7] to predict the AD by utilizing the huge expressions and DNA data samples and found that this method has attained 82.3% performance accuracy for their findings. Depending upon the hybrid feature extraction and CNN utilization, Arifa et al. [8] proposed a method for diagnosing the AD and the findings has attained better performance accuracy. Hong et al. [9] constructed and introduced a LSTM related prediction approach having the network of fully activated layers and the connected layers for encoding the temporal relationship between the stages of AD and the features.

Querques et al. [10] investigated the retinal vessels of the disease disorders by using the OCTA analysis, and dynamic vessel analyzer (DVA). Also, the author demonstrated the important impairment of the retinal neurovascular coupling to characterize the subjects of MCI and the AD. A predicting technique for AD has been introduced by Soliman et al. [11] by utilizing a deep 3D-CNN in which the generic attributes can be learned. The attributed like classifying Alzheimer's brain from normal healthy brain depending upon the Magnetic Resonance Imaging (MRI) scan of the brain capturing the biomarkers of AD. CNN has been utilized for feature extraction by [12,13]. Some of the existing state-of-art works for detecting AD and classifying it was analyzed by Hazarika et al. [14] depending upon the different feature extraction approaches. The dataset was re-enhanced separately with fuzzy color image enhancement, Deep Dream,

and the hyper column techniques [15]. Combined with the deep features, a Visual Geometry Group-16 (VGG-16) deep learning model is used in the enhancing process.

With a string focus laid on the problem of the functional brain network classification for AD detection, Bi et al. [16] found and introduced two different types of deep features such as the adjacency positional features, and the regional connectivity positional features. For the diagnosis of AD, Duc et al. [17] evaluated a novel deep learning method and for the classification task, a 3-D CNN architecture was developed and the obtained results proved that a functional brain feature dataset will not only help in the earlier diagnosis of AD but also in predicting Mini Mental State Examination (MMSE) scores of the individuals. Buvaneswari et al. [18] introduced a deep learning-based segmenting method using SegNet for detecting the features of AD pertinent brain parts from the Structural MRI (SMRI). ResNet-101 is used for accurate classification of the AD and the dementia conditions. An effective machine learning model has been proposed by Kumari et al. [19] that successfully diagnosed AD, ncMCI, ncMCI and CN that are being detected during the earlier stage itself. The effective segmentation and classification methods for classifying the abnormalities and normality of AD were proposed by Suresha et al. [20]. A method for automatically segmenting the retina layers in the in the Spectral Domain Optical Coherence Tomography images has been demonstrated by Stephanie et al. [21] with the graph theory along with a dynamic programming. In the demonstrated method, an appropriate value was chosen for limiting the search regions which may result in inaccurate results of boundary detection of retinal layers. Though the above process outperforms well both in the primary and the multiclass classifications for predicting AD, it sill have certain limitations like the over fitting and the class imbalance which will be the key concern for the diagnosis of the AD. Additionally the limited data samples make efficient categorization of the AD disorders and hence, PSO and LSTM are required for addressing these limitations. The datasets, techniques, identified disease types, advantages and disadvantages of previous works are shown in [Tab. 1](#).

**Table 1:** Summarization of previous works

Technique [citation]	Dataset	Identified disease type	Advantages	Disadvantages
CNN [3]	ADNI dataset	AD	Reduce computational complexity, memory requirements, overfitting.	-
CNN & PSO-FCM approach [4]	OASIS, SLIM, INDI datasets	AD	Achieved an accuracy of 94.73%.	Lack of large annotated brain data sets availability.
Four optimization algorithms [5]	Primary dataset was collected by them.	AD	high accuracy of 98% for segmentation is achieved.	-
Multi-modal neuroimaging feature selection method with consistent metric constraint (MFCC) [6]	ADNI dataset	AD	Attained a classification accuracy of 95% .	-

(Continued)

**Table 1 (continued)**

Technique [citation]	Dataset	Identified disease type	Advantages	Disadvantages
Deep learning model [7]	Gene expression and DNA methylation profiles	AD	Integrating multi-omics into prediction models to identify molecular factor.	Focused on integrating two molecular layers and did not use the multi-omics dataset from the same sample group
LSTM [9]	ADNI dataset	AD	Carried out the future state prediction for the disease.	-
Dynamic vessel analyzer (DVA) and OCTA analysis [10]	Primary dataset generated by them	AD	-	Single time point employment for each patient and small sample size of affected subjects.
3D-CNN [11]	ADNI dataset	AD	Achieved 80.6% test data accuracy.	-
Deep CNN [12]	Open access series of imaging studies dataset	AD stages	provides better classification result even for the small dataset	-
CNN [13]	OASIS dataset	AD stages	Achieved a 99.21% accuracy	-
Survey [14]	-	AD	Provides depth analysis about feature extraction technique.	segmentation part is hasn't been focused.
VGG-16 deep learning model [15]	hybrid dataset used.	Dementia	Overall accuracy was 99.94%.	No detailed information are available about patients from whom the datasets was obtained
Extreme learning machine-boosted structure [16]	ADNI dataset	AD (Mild cognitive impairment)	Achieves 91.4% AUC for on the NC-AD problem	In the functional brain network, the performance is still vulnerable to different ROI positions.
3-D CNN [17]	Primary dataset was collected by them.	AD	The mean balanced test accuracy was 85.27% achieved.	-
ResNet [18]	ADNI dataset	AD and dementia	An accuracy of 95% is achieved.	-

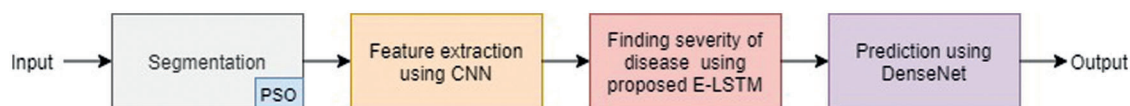
(Continued)

**Table 1 (continued)**

Technique [citation]	Dataset	Identified disease type	Advantages	Disadvantages
machine learning model [19]	ADNI dataset	AD, cMCI, ncMCI and CN	An accuracy of 90.25% is achieved.	-
Adaptive histogram equalization (AHE) and grey wolf optimization based clustering algorithm (GWOCA) [20]	NIMHANS, ADNI, MIRIAD, and OASIS datasets	AD	Achieved 2.2%–6% improvement in classification.	-

### 3 Materials and Methodology

OCTA eye image was taken for predicting the AD at its earlier stage and further for extracting the robust features of the eye, Convolutional Neural Network has been utilized by replacing the fully connected layer with the convolutional layer for extracting the features deeply. But, segmentation will be performed before extracting the features of the image and the inaccuracy of segmenting the retinal layer will be solved by using PSO owing to the advantageous feature it is having that it will be generic and well suitable for all OCTA images. Hence, the proposed E-LSTM is used for predicting the severity of AD whereas the DenseNet is used for predicting the possibilities of Ad and the flow of this process is shown in Fig. 1.



**Figure 1:** Overview of the work

#### 3.1 Data Representation

For predicting the AD disease, the datasets of Optical Coherence Tomography (OCT) test has been used. (Source: <https://www.kaggle.com/paultimothymooney/kernman2018>). As OCT is having the capability to capture the swelling/thickening of the inner retinal layers, nearly 30 million OCT scans are being taken every year across the globe and now, it has turned out to be the most popular medical imaging procedure.

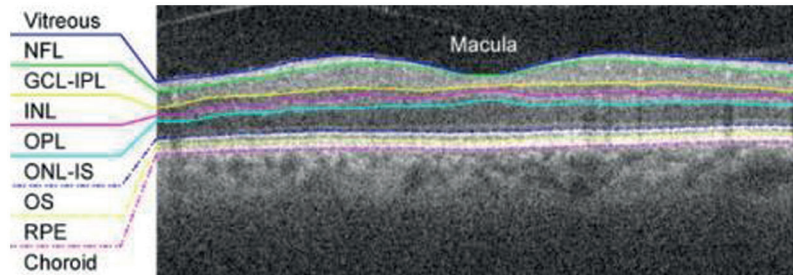
#### 3.2 Segmentation and Feature Extraction

Here in this section, the segmentation process of the OCT images are done initially which was then followed by the feature extraction process using CNN.

##### 3.2.1 Segmentation

It is always important to segment the anatomical structures and the pathological structures in OCT image to diagnose the AD but, executing the segmentation via manual process will be a prolonged and difficult process and hence, the dynamic programming and the graph theory has been used in this work to segment the retinal layers of the Spectral Domain Optical Coherence Tomography images [21]. Upon recognizing the boundaries of the eight layers, the retinal image is segmented in to seven layers as the Retinal Pigment Epithelium (RPE), the Nerve Fibre Layer (NFL), the Inner Nuclear Layer (INL), the Outer Plexiform Layer (OPL), the Outer Segment (OS), the Outer Nuclear Layer to Inner Segment

(ONL+IS), and the Ganglion Cell to Layer-Inner Plexiform Layer (GCL+IPL) and all they are illustrated in the Fig. 2.



**Figure 2:** Retinal layers of a cross-sectional SDOCT image (B-scan)

For the retinal layer boundary detection, the search space was limited in [21], and after detecting the first layer of the retinal image, the next succeeding layer or the preceding layer of the retinal image will be segmented with limiting search space to an area of specified pixels either over or beyond the segmented boundary. But, this method of allocating a particular pixel value to the next layer segmentation will end up with inaccurate results and hence, to overcome this limitation, PSO will be utilized for finding the optimal search space. Pixels of the image will be considered as swam for finding the boundary of the retinal layer and the boundary selection of the retinal layer will be executed in the following steps.

As a first step, the particles will be initialized with the random position and the velocity in which the random position will be denoting the initial position of the particle that is the pixel of the image. Every particle will be moving towards the valid boundary pixel which will be depending upon the boundary of the retinal layer. Using the fitness function, the fitness value will be calculated for each particle during iteration and then, the pbest (each particle best or local solution) and gbest (Global best for entire swarm or global solution) will be found and updated. Then, the position and the velocity of the particles will be updated. Then, the ensuring process will be executed to find whether the particles have reached the stopping condition or not. If the stopping condition is reached, the optimal solution (the layer boundary) will be obtained as shown in Fig. 3 and the new position and the new velocity for every particle can be given in Eqs. (1) and (2) and derived using [22].

$$V_i(Q+1) = V_i(Q) + y_{1i}(p_i + X_i(Q)) + y_{2i}(Gbest - X_i(Q)) \quad (1)$$

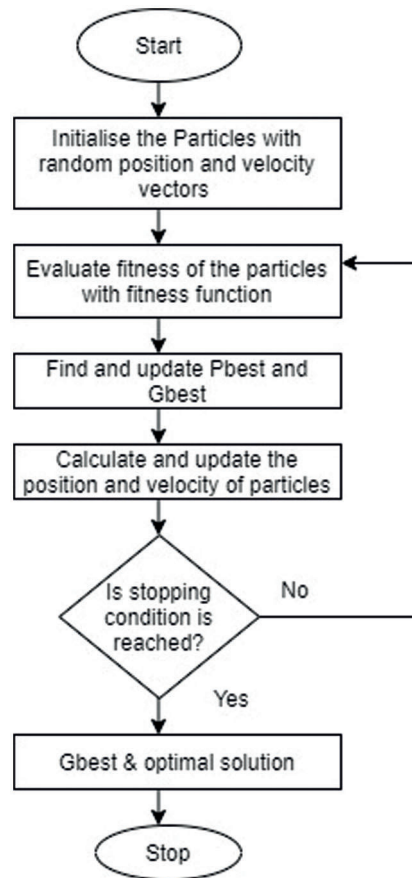
$$X_i(Q+1) = X_i(Q) + V_i(Q+1) \quad (2)$$

where,

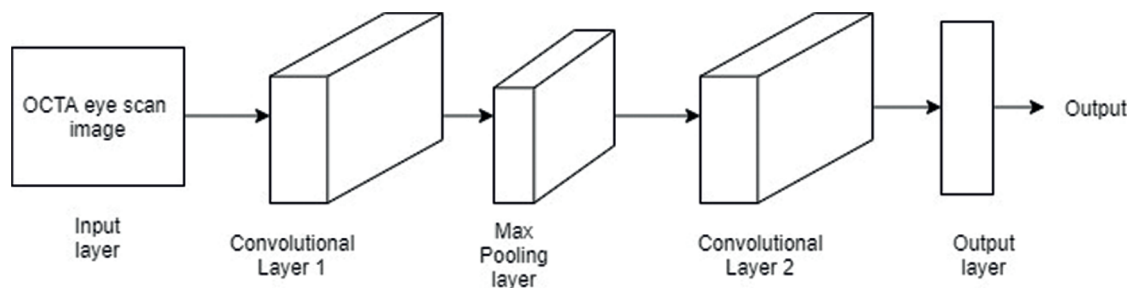
$X_i$ —Position of ith particle; Gbest—best position found by swarm; Q—discrete time index,  $p_i$ —Best position found by ith particle;  $V_i$ —Velocity of the ith particle; i—particle index.

### 3.2.2 Feature Extraction Using CNN

The eye features can be extracted directly from an OCTA scan image while passing through the layers of the CNN which consists of five layers as the input layer, convolutional layer 1, max pooling layer, convolutional layer 2, and the output layer as shown in Fig. 4 [23,24]. In this proposed work, the fully connected will be replaced with another convolutional layer to have a deep feature extraction. The five layers are explained in detail as follows.



**Figure 3:** PSO for retinal layer boundary detection



**Figure 4:** CNN for feature extraction

- **Input layer**

The input layer is the first layer in CNN and the CNN input will be in the dimension of  $H \times W \times D$  neighbourhoods of a pixel where  $D$  denotes the depth of the image and  $H$  &  $W$  denotes the height and width of the image respectively.

- **Convolution layer1**

Multiple layers will be used by the convolutional layer 1 for extracting the local features and also to ensure it (Conv1). Though plentiful filters were being used, the features are extracted using the learning



best filter where every filter will be moved to the input image. Stride is defined as the filter moves with number of pixels and when computing the dot product between the filter weight and the pixels in the image, an activation map will be produced as a result. Different set of image features will be captured by every filter and the convolutional layer can be represented as,

$$C = a(f * i + b) \quad (3)$$

In Eq. (1), C represents output of Conv1, a denotes an activation function, f denotes filters and b denotes bias of conv1 and i represents input image.

- **Max Pooling layer**

The output from Conv1 will be passed to the next pooling layer and in this max pooling layer, computational complexity will be reduced through generalisation of features from previous convolutional layer. The feature maps will turn robust while passing through the pooling layer even though it shrinks. The rectified activation map will contain only the significant information and at the same time, the depth information cannot be altered.

- **Convolution layer2**

The fully connected layer will be replaced with the convolutional layer 2 (Conv2) for extracting the more depth features and one of the main advantages of replacing the fully connected layer with the convolutional layer 2 is that the adjustments will be made to reduce the number of parameters owing to the reason that the weights are shared in a convolutional layer. This indicates the faster and more robust learning and in addition, max pooling layer will be used just after the convolutional layer for reducing the dimensionality of the layer. As a result, the robustness to the distortions in input image was improved greatly thus showing a better overall performance.

- **Output layer**

The extracted depth features will be passed to the last output. The output of this layer will be the features of eye image and the extracted final features are maximally stable extreme regions (MSER) features, Min-Eigen features, Regional minima, Moment, curvature variance, elasticity, Texture feature, entropy, Red average value, mean, contrast, and saliency variance [25].

### 3.3 Proposed E-LSTM

An enhanced technique named Enhanced LSTM (E-LSTM) is proposed to find the severity of the AD with highest accuracy and lower error rate and it is illustrated in Fig. 5. The proposed E-LSTM technique comprises of one input layer, and five succeeding sections of LSTM layers with dropout layer, 01 fully connected layer, 01 precise layer and 01 output layer [26]. By preceding the classifications based on the input features, the disease severity will be calculated in which the input features like the test results and the symptoms are fed to the first layer which is the input layer.

The OCTA scan is taken as the test results and the symptoms are the differences between the normal and abnormal subjects. The over fitting problems can be overcome with the E-LSTM method by adding the dropout layer below the LSTM layer with a substantial probability of 0.6. The input from the input layer will be forwarded to the embedding layer which will convert the input features into an embedding of a specific size. All the features are considered and embedded in the embedding layer for improving the accuracy level of the prediction and the size and dimensionality of the embedding features are the two factors where each of the input features taken are represented as vector. Then, the hidden state dimensions and the number of layers are defined in LSTM layer and the number of hidden units is assumed to be 512 in this layer.



**Figure 5:** Architecture of proposed E-LSTM

In the proposed E-LSTM method, the iterations were set to 350 and 5 sub blocks are present in which the first block has the LSTM layer that is followed by the dropout layer with the probability of 0.6. The sub block next to this block will comprise a LSTM layer which will be followed by the dropout layer with the probability of 0.5 and in the next block, a LSTM layer is followed by the dropout layer with the probability of 0.4. Similarly, the next block comprises a LSTM layer which is followed by the dropout layer with probability of 0.3 whereas the last sub block comprises a LSTM layer which is followed by the dropout layer with a probability of 0.2. In each sub blocks, the drop out value will be reduced progressively whereas several neurons will be shielded in the LSTM network to increase the model robustness of the dropout layer. The accuracy of the proposed model classification is increased by overcoming the problem of over fitting in which the data will be retrieved into the fully connected layer. The output of the LSTM layer will be mapped into a specific output size by a fully connected layer where the counts of the input features and the hidden units will be equal in the LSTM network. When compared with the other existing methods, the training loss and error rate has been reduced greatly.

In the proposed work, when the epochs are increased, the classification accuracy of the model will also be improved. The cell state will be updated after completing iteration and for every input; the cell state will be carrying the related information. The gates will decide what information should be permitted in the cell state and during the training process, the data about the hold and forget are studied by the gates and the data can be added or removed in the cell state by the gates. Following the fully connected layer, the new layer named precise layer is added to the LSTM network for finding the severity level of the disease which will help the doctors to initiate the treatments rapidly for the patients.

Two operations will be performed in the precise layer and the severity of the disease will be analysed at first by using the features of the symptoms. Then next, the severity of the disease is denoted by a range of values between 0 and 1 and this operation is performed by using a sigmoid function which will convert all the output in series between 0 and 1. The advantage of using this sigmoid function is that even the small changes in the input will be reflected in the output i.e., when the count of symptoms increase or decrease in the input, the severity level will get changed in the output simultaneously. This will increase the overall accuracy of the model.

The severity level of the disease can be described by sigmoid function in the following equation,,

$$S(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

In the above equation, 'x' denotes the symptoms and the test result range whereas S(x) denotes the severity level of the disease. The output of sigmoid function will be in range between 0 and 1 and it will be continuous. When the number of symptoms of the disease is low, the sigmoid function output will be low thereby indicating low level of the severity of the disease. On the other hand, when the number of disease symptoms is high, then the output of the sigmoid function will also get increased indicating the highest severity level of the disease. When the severity level of disease is high, it alarms that the patient is in dangerous condition and immediate attention will be required. This output will be forwarded to the output layer.

### 3.4 *Densely Connected Convolutional Network (DenseNet)*

When the layers of the convolutional neural networks will increase, it may create problems like inconsistent training results or disappearance in the gradients. Vanishing gradient problems will occur while training the Artificial Neural Networks with gradient based approaches. Hence, it will be very difficult to learn and tune the parameters in the previous layers of the network and when the numbers of architecture layers gets increased, this issue will turn out to be more severe. During the training calculations, the deep learning network may lose valuable knowledge on the features of the input. Hence, the ResNet network with enhanced CNN called DenseNet has been adopted to overcome this issue this is proposed work [27].

Via ResNet, the DenseNet will be having the enhanced network efficiency thus becoming the benchmark for the image classification network. It will decrease the numbers of parameters until a point thereby preventing the gradient disappearance problem of the neural networks. Hence, the network efficiency will be increased in the dense block thereby multiplexing the features of the network. Using [27], the Eqs. (5)–(7) will be derived.

The layer n output of the ResNet networks will be computed as,

$$X_n = H_n(X_{n-1}) + X_{n-1} \quad (5)$$

The layer  $n - 1$  and non-linear transformation of layer  $n - 1$  outputs can be used as combined outputs of layer  $n$ . Using DenseNet, the binding strength of the  $n$  layer can be calculated and it is given as,

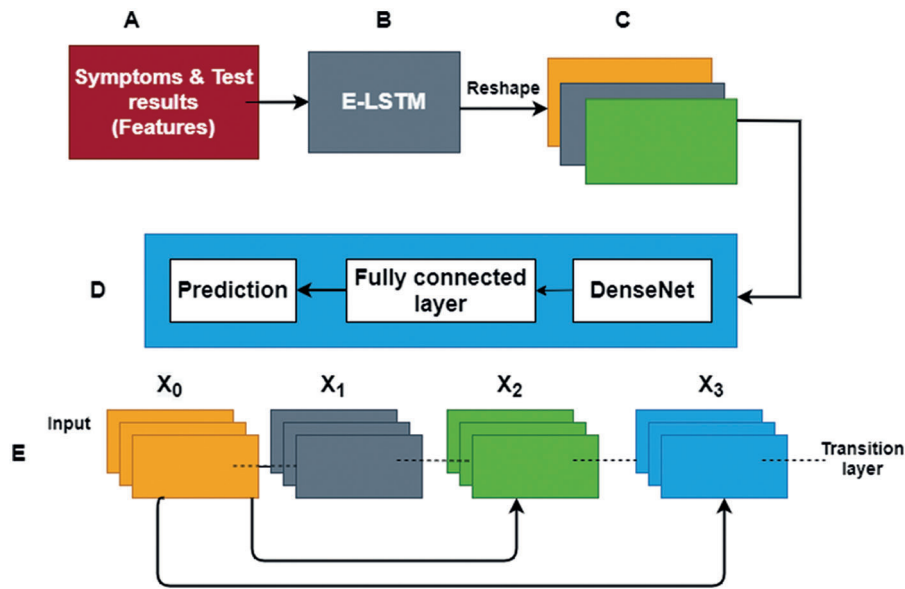
$$X_n = H_n[X_0, X_1, \dots, X_{n-1}] \quad (6)$$

The sequence of output ranges from 0 to  $n - 1$  layers and it can be denoted as  $[X_0, X_1, \dots, X_{n-1}]$ . The structural feature maps are made up of channels and they are connected and less parameter will be merged in DenseNet for enhancing the generalization effect of the proposed model.

### 3.5 Prediction Using E-LSTM and DenseNet

While predicting the disease using the CNN, the combined features such as test results and symptoms are used as inputs for training and in this case, the context features for every feature will not be obtained. Hence, to solve the context between data and over fitting, the proposed E-LSTM method is used. Being well trained with the extracted features, the proposed E-LSTM method will help in reducing the various combinations of the feature data. Then, the disease prediction will be done by combining and forwarding the trained features into the DenseNet structure.

The architecture for predicting the disease is illustrated in the Fig. 6 and the architecture is consisting of five parts. Part A will comprise the two types of data input whereas E-LSTM with the input features will be used in Part B for the training model to find the severity level of the disease. E-LSTM consists of input gate, output gate, and forget gate to find the dependencies between the data. The unit size is set to 128 where the training will be executed for one dimensional vector that comprises of three layers measuring 256 lengths. Reshape and concatenate functions are used in Part C which will be used for transforming one-dimensional vector into a 3-channel  $16 \times 16$  matrix. The features of the channel will be progressed to DenseNet training network for achieving the easiest training and efficient parameters. The network will be consisting of a ReLU activation function, convolution layer, linearization, and a pooling layer. In part D, the output of DenseNet is connected to a 128-node fully connected layer and the corresponding output will be obtained by prediction in the last layer and the DenseNet in Part E will split the whole network into several densely connected blocks. The continuous tuning will be done whereas the optimal prediction result will be obtained using a 3-layer dense block.



**Figure 6:** Architecture for AD prediction

The mean absolute percentage error (MAPE) can be utilized as loss function, and it can be given as,

$$MAPE = \sum_{i=1}^n [(f_i - y_i) / f_i] \times \frac{100}{n} \quad (7)$$

Whereas  $y_i$  denotes a predicted value of  $i^{\text{th}}$  sequence,  $f_i$  denotes a true value of  $i^{\text{th}}$  sequence, and  $n$  indicates the size of each training batch. By a separate discrete data point, the errors will be created

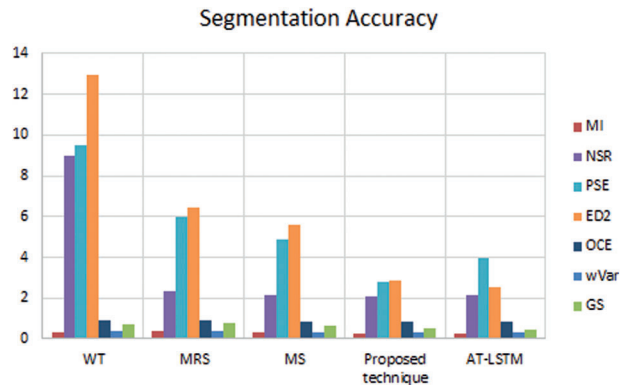
which can be then reduced by normalization of error data and this function will be carried out using MAPE. By applying L2 regularisation to the loss function, the sum of the squares of the structure's weights is added to minimize the loss function. To decrease the problem of over fitting, the dropout was set as 0.2.

## 4 Result and Discussion

Accurate boundary detection has been a challenging issue in the existing methods and to overcome this issue, the proposed method is evaluated using three parameters like training loss, training time, and segmentation accuracy, and the performance of the disease severity classification was measured in terms of Recall, Precision, and F1 Score and they are illustrated in detail in below sub-sections.

### 4.1 Segmentation Accuracy

The segmentation accuracy of the proposed work is compared with the other algorithms like the Watershed Transform Segmentation (WTS) method, the Multi-resolution Segmentation (MRS) method, and the Mean Shift (MS) method. For assessing the segmentation accuracy, the terms like the normalized global Moran index (MI), Global Score (GS), the normalized weighted variance (wVar), the Number Of Segments Ratio (NSR), Potential Segmentation Error (PSE), Object Level Consistency Error (OCE), and the Euclidean distance 2 (ED2) are used. As illustrated in Fig. 7, the proposed method has attained the optimum segmentation accuracy other than wVar term. The global calculation is carried out on the basis of segmented regions for creating a large segment that has more effect on the overall homogeneity and this is called as the wVar. The reflection will be made by global score (GS) for evaluating the best scale selection for intra and inter segment heterogeneity information. Since more detailed items were preserved by their segmentation outputs, the WTS and MS approaches were found to contain high values of wVar. It seems that the MRS segmentation output is having the highest MI rate thus suggesting that the segmentation wasn't really adequate thus making the segmentation to be inaccurate.



**Figure 7:** Segmentation accuracy of different methods

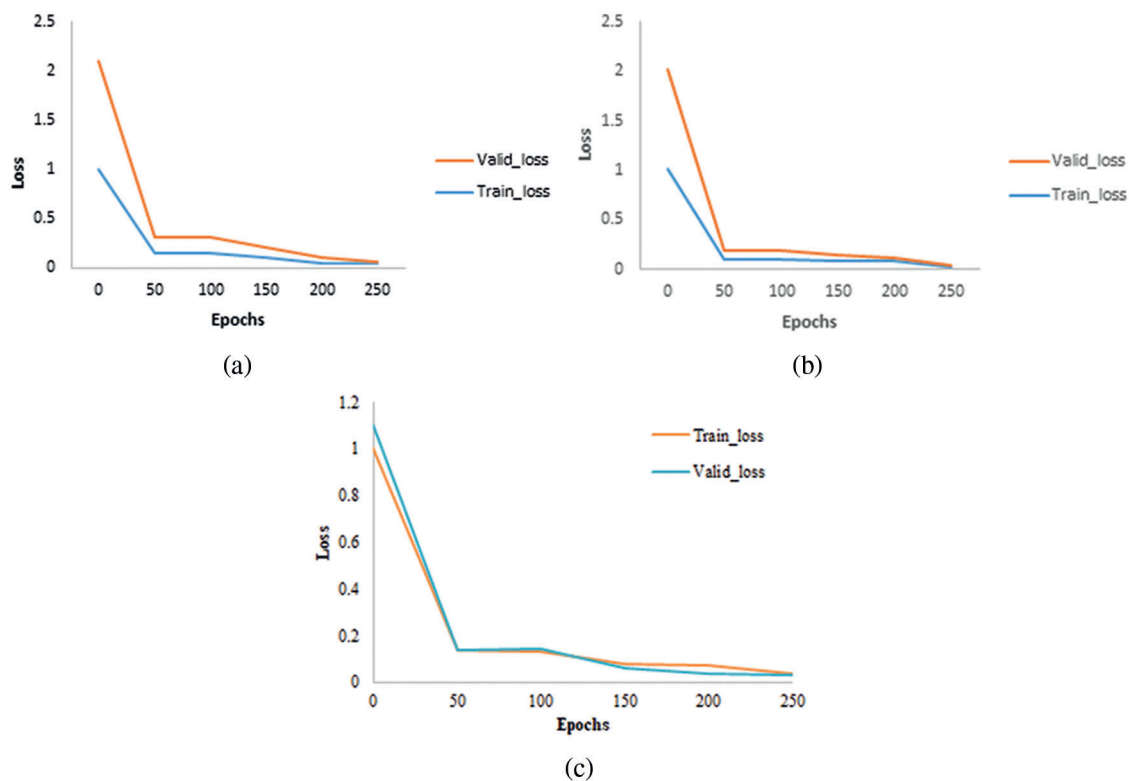
PSE can be defined as the proportion of overall under segmented region and the overall reference polygons region and the value of the PSE is in between in the range of 0 and  $+\infty$ . If the best value is found to be 0, then there will be no under segmentation in the reference object. The absolute variance in the number of reference polygons and the number of corresponding segments that are divided by the number of reference polygons will be referred to as NSR. When the value of NSR is 0, then all the matching objects and the reference objects will experience a one-to-one correspondence. This will end up resulting in the best segmentation where the one-to-many or vice versa correspondence will be experienced by all matching objects and reference objects if the value of NSR is set large and this will

result in under or over segmentation. ED2 can be computed by using NSR and PSE. The index of ED2 will become the most accurate in a two-dimensional PSE-NSR space, if NSR and PSE contains the same magnitude, and the good segmentation efficiency will be reflected with the small ED2 value. The value of PSE will be high in WTS algorithm which indicates the under segmentation. Simultaneously, the highest value of NSR will shows an over segmentation. Compared to other techniques, the proposed method performs well in terms of Ed2, and OCE. PSO is the reason for accuracy retinal layer segmentation in the proposed method and this PSO is suitable for all input images of various sizes as it can detect the correct optimal boundary in the input image.

#### 4.2 Training Loss

Let's fix, the iterations as 350, learning rate (D) as 0.00128, embedding D as 1, batch size number D as 64, output size D as 3, layers D as 2, and hidden size D as 64.

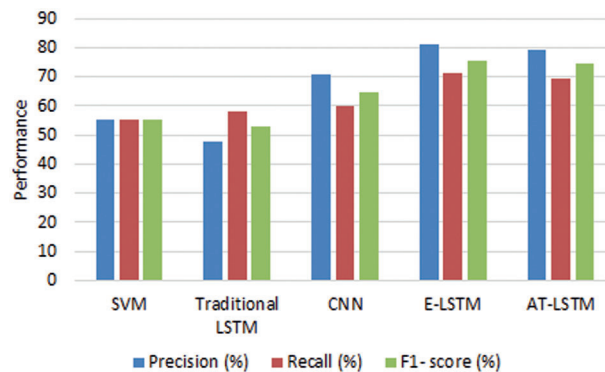
The model will be saved with the minimal validation loss and the epochs. The training loss of traditional LSTM, E-LSTM and Attention based LSTM (AT-LSTM) is illustrated in Figs. 8a–8c respectively. During the LSTM model training, the verification loss will start increasing to 150 times and it'll not be decreasing at all thus, indicating that the model is suffering from over fitting problem. When compared with the traditional LSTM, the accuracy of E-LSTM will be enhanced whereas the text loss and verification losses will be decreased. Using the above experiment, The E-LSTM model is enhanced with the iterations count set to 350. The model is saved with low validation loss and in the proposed model, no over fitting was found from training process and shows that the model is learning all the time.



**Figure 8:** (a) Training loss comparison of LSTM (b) Training loss comparison of E-LSTM (c) Training loss comparison of AT-LSTM

### 4.3 Performance of Severity Disease Classification as Measured Using Precision, Recall, and F1-Score

For assessing the performance of the severity disease classification, the three evaluation metrics such as precision, recall and F1score are used and from the observations, it is found that the proposed E-LSTM method outperforms well in terms of precision, recall and F1-score when compared to AT-LSTM, traditional LSTM, SVM, and CNN. This is because the over fitting problem in the traditional LSTM will can be reduced by the proposed E-LSTM by adding dropout layer with a dropout set to 0.2. This is shown in Fig. 9. A new layer in the name of precise layer will be added to the LSTM network for the process of disease severity classifications.



**Figure 9:** Precision, recall, and F1-score of all methods

## 5 Conclusion

Recognizing the health related brain data will be very useful for detecting the anomalies and also to save the lives of humans and especially, predicting the brain disease will be very helpful in the medical areas. So, the earlier diagnosis with the necessary preventive measures must be implemented which will help to control the death rate. Hence, a novel method named E-LSTM is proposed in this paper which is aimed at forecasting the chance of AD from eye disease. PSO is utilized for detecting the boundary of the accurate retinal layer which will also help to segment the retinal layers more accurately. When compared with the existing methods, the proposed E-LSTM method along with DenseNet is found to be most accurate in predicting the brain disease with increased prediction accuracy of 10%. Hence, the proposed E-LSTM method is found to be performing well than the other existing works in terms of Precision, Recall, Segmentation accuracy, F1 Score, and the Training loss.

**Funding Statement:** The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] A. K. Singh and S. Verma, "Use of ocular biomarkers as a potential tool for early diagnosis of Alzheimer's disease," *Indian Journal of Ophthalmology*, vol. 68, no. 4, pp. 555–561, 2020.
- [2] "Alzheimer's disease signs detected by non-invasive eye scan. GEN-genetic engineering and biotechnology news," 11 Mar 2019. [Online]. Available: [www.genengnews.com/news/alzheimers-disease-signs-detected-by-noninvasive-eye-scan/](http://www.genengnews.com/news/alzheimers-disease-signs-detected-by-noninvasive-eye-scan/).
- [3] H. A. Helaly, M. Badawya and A. Y. Haikal, "Deep learning approach for early detection of Alzheimer's disease," *Cognitive Computation*, 2021, <https://doi.org/10.1007/s12559-021-09946-2>.

- [4] H. Alliou, M. Sadgal and A. Elfazziki, "Utilization of a convolutional method for Alzheimer disease diagnosis," *Machine Vision and Applications*, vol. 31, no. 4, 2020.
- [5] D. Chitradevi and S. Prabha, "Analysis of brain sub regions using optimization techniques and deep learning method in Alzheimer disease," *Applied Soft Computing*, vol. 86, pp. 105857, 2020.
- [6] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang *et al.*, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease," *Medical Image Analysis*, vol. 60, pp. 101625, 2020.
- [7] C. Park, J. Ha and S. Park, "Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset," *Expert Systems with Applications*, vol. 140, pp. 112873, 2020.
- [8] A. Shikalgar and S. Sonavane, "Hybrid deep learning approach for classifying Alzheimer disease based on multimodal data," *Advances in Intelligent Systems and Computing*, pp. 511–520, 2020.
- [9] X. Hong, L. Rongjie, Y. Chenhui, Z. Nianyin, C. Chunting *et al.*, "Predicting Alzheimer's disease using LSTM," *IEEE Access*, vol. 7, pp. 80893–80901, 2019.
- [10] G. Querques, E. Borrelli, R. Sacconi, L. De Vitis, L. Leocani *et al.*, "Functional and morphological changes of the retinal vessels in Alzheimer's disease and mild cognitive impairment," *Scientific Reports*, vol. 9, no. 1, pp. 63, 2019.
- [11] S. A. Soliman, E. A. El-Dahshan and A. M. Salem, "Predicting Alzheimer's disease with 3D convolutional neural networks," *International Journal of Applications of Fuzzy Sets and Artificial Intelligence*, vol. 1, pp. 125–146, 2020.
- [12] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informatics*, vol. 5, no. 2, pp. 2, 2018.
- [13] H. Nawaz, M. Maqsood, S. Afzal, F. Aadil, I. Mehmood *et al.*, "A deep feature-based real-time system for Alzheimer disease stage detection," *Multimedia Tools and Applications*, vol. 80, pp. 35789–35807, 2020.
- [14] R. A. Hazarika, A. K. Maji, S. N. Sur, B. S. Paul and D. Kandar, "A survey on classification algorithms of brain images in Alzheimer's disease based on feature extraction techniques," *IEEE Access*, vol. 9, pp. 58503–58536, 2021.
- [15] M. Ogacar, Z. Comert and B. Ergen, "Enhancing of dataset using DeepDream, fuzzy color image enhancement and hypercolumn techniques to detection of the Alzheimer's disease stages by deep learning model," *Neural Computing and Applications*, vol. 33, pp. 9877–9889, 2021.
- [16] X. Bi, X. Zhao, H. Huang, D. Chen and Y. Ma, "Functional brain network classification for Alzheimer's disease detection with deep features and extreme learning machine," *Cognitive Computation*, vol. 12, no. 3, pp. 513–527, 2020.
- [17] N. T. Duc, S. Ryu, M. N. Qureshi, M. Choi, K. H. Lee *et al.*, "3D-Deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, 2020.
- [18] P. R. Buvanewari and R. Gayathri, "Deep learning-based segmentation in classification of Alzheimer's disease," *Arabian Journal for Science and Engineering*, vol. 46, pp. 5373–5383, 2021.
- [19] R. Kumari, A. Nigam and S. Pushkar, "Machine learning technique for early detection of Alzheimer's disease," *Microsystem Technologies*, vol. 26, pp. 3935–3944, 2020.
- [20] H. S. Suresha and S. S. Parthasarathy, "Detection of Alzheimer's disease using grey wolf optimization based clustering algorithm and deep neural network from magnetic resonance images," *Distributed and Parallel Databases*, 2021.
- [21] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt *et al.*, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Optics Express*, vol. 18, pp. 19413–19428, 2010.
- [22] Particle Swarm Optimization. Share and Discover Knowledge on SlideShare, 22 May 2008. [Online]. Available: [www.slideshare.net/stelabouras/particle-swarm-optimization](http://www.slideshare.net/stelabouras/particle-swarm-optimization).
- [23] V. Kudva, K. Prasad and S. Guruvare, "Automation of detection of cervical cancer using convolutional neural networks," *Critical Reviews in Biomedical Engineering*, vol. 46, no. 2, pp. 135–145, 2018.



- [24] “What Are the benefits of converting a fully connected layer in a deep neural network to an equivalent convolutional layer?. quora,” [Online]. 2016. Available: <https://tinyurl.com/bxymmdne>.
- [25] C. S. Sandeep, A. Sukesh Kumar, K. Mahadevan and P. Manoj, “Extracting the features of retinal OCT images for the early diagnosis of Alzheimer's disease,” in *2019 5th Int. Conf. on Advanced Computing & Communication Systems (ICACCS 2019)*, Coimbatore, India, 2019.
- [26] S. Agrawal, Reading between the layers (LSTM Network). Medium, 2019. [Online]. Available: <https://towardsdatascience.com/reading-between-the-layers-lstm-network-7956ad192e58>.
- [27] Z. Li, J. Zhu, X. Xu and Y. Yao, “RDense: A protein-RNA binding prediction model based on bidirectional recurrent neural network and densely connected convolutional networks,” *IEEE Access*, vol. 8, pp. 14588–14605, 2020.