ASSESSMENT OF NEUROIMAGING DATA AND IDENTIFICATION OF ALZHEIMER'S DISEASE USING EXTREME LEARNING MACHINES

Dharini Raghavan,* K.V. Suma, & Puneeth N Ganesh

Department of Electronics and Communication, M S Ramaiah Institute of Technology, Bangalore, Karnataka 560054, India

*Address all correspondence to: Dharini Raghavan, Department of Electronics and Communication, M S Ramaiah Institute of Technology, Bangalore, Karnataka 560054, India, E-mail: dhariniraghavan2001@gmail.com

Original Manuscript Submitted: 4/4/2023; Final Draft Received: 4/14/2023

Alzheimer's disease (AD), one of the most common forms of dementia, is a cognitive disorder that is progressive in nature and causes a dynamic deterioration of the mental state of an individual. It severely damages the brain cells, neurotransmitters, and nerves, leading to irreparable damage to the brain, which is one of the major causes of dementia. Early identification, assessment, and timely diagnosis are of paramount importance to slow down the progression of the disease, which calls for the design and development of algorithms and technology-aided tools for accurate detection, diagnosis, and prediction of the severity of Alzheimer's disease. To provide a solution to this, we propose an extreme learning machine (ELM) algorithm that is trained on neuroimaging data from longitudinal MRI scans obtained from the OASIS database. We adopt an extensive feature engineering pipeline to choose the most significant features for early identification of the onset of dementia. We obtain an overall accuracy of 98.3%, sensitivity of 0.956, specificity of 0.962, and F1 score of 0.972. We also show that our proposed ELM algorithm outperforms several other contemporary classifiers based on a range of evaluation metrics. The paper also provides a feasibility analysis of the proposed model for real-time clinical deployment.

KEY WORDS: *dementia, Alzheimer's disease, magnetic resonance imaging, extreme learning machines, machine learning classifiers*

1. INTRODUCTION

Alzheimer's disease (AD), which is an irrevocable cognitive disorder, is one of the most common forms of dementia. It is known to have affected about 55 million people worldwide, among which over 60% of them belong to low- and middle-income countries. The number of cases of Alzheimer's disease is exponentially increasing, with an average of 10 million cases every year. AD is expected to affect about 150 million people by 2050, thus leading to severe economic and medical consequences. Dementia or Alzheimer's disease is caused by abnormal deposits of proteins in and around the cells of the brain. For instance, the amyloid protein produces clumps around the brain cells, and another protein called tau produces tangles around the brain cells. Early diagnosis of Alzheimer's disease is highly challenging and often involves expensive

and invasive tests that are available only in sophisticated clinical settings, which makes them infeasible for common people to adopt. For instance, biomarkers of amyloid and phosphorylated tau assessed through cerebrospinal fluid examinations, PET scans, and plasma assays are useful for Alzheimer's disease diagnosis, but these examinations are not appropriate for screening potential Alzheimer's disease in primary care or community contexts. Numerous studies have shown that there is no proven way to fully treat or stop the progression of Alzheimer's disease (De Strooper and Karran, 2016). In order to avoid irreversible memory loss, this calls for the creation of extremely reliable and accurate methods for AD early detection, particularly at the presymptomatic stages (Galvin, 2017).

To categorize the severity of AD under various conditions using images from various modalities, several researchers have recently suggested machine learning– and deep learning (DL)–based medical imaging methods (Aldhyani et al., 2020; Zarandi et al., 2011; Korolev et al., 2017; Litjens et al., 2017; Qiu et al., 2018; Zhang et al., 2001b; Jenkinson et al., 2002). Advanced neuroimaging methods, such as positron emission tomography (PET) and magnetic resonance imaging (MRI), have been extensively used to find structural and molecular biomarkers for the detection of onset of AD (Bartos et al., 2019; Fonov et al., 2011; Vidoni et al., 2012). MRI is a noninvasive imaging technique that offers crucial details about internal bodily structures, including their location, shape, and size. It offers a noticeable soft tissue contrast, enhancing the clarity of images that are captured for analysis. Functional and structural imaging are the two general categories for MRI, with tasking-state and resting-state functional imaging falling under functional imaging. T1-weighted MRI, T2-weighted MRI, and diffusion tensor imaging are the three types of structural MRI imaging.

Accurate understanding of the human brain's processes is also aided by PET imaging. The diffusion of R18 fluorodeoxyglucose (FDG), which provides information on the brain glucose metabolic rates (CMRglc), is used in PET modality to create neuroimages. CMRglc is used to distinguish between AD and other types of dementia. When it is difficult to distinguish between variations in pathological and physiological morphology, FDG PET can be especially useful.

In order to effectively identify and categorize brain disorders, deep learning–aided techniques have gained popularity due to their improved robustness and accuracy in feature extraction. Feature representation and extraction play a vital role in the analysis of medical images. Numerous pattern recognition techniques, including support vector machines (SVM), logistic regression (LR), linear program boosting method (LPBM), linear discriminant analysis (LDA), and support vector machine-recursive feature elimination (SVM-RFE), have been widely used to identify and classify AD as well as predict the course of the disease (Rathore et al., 2017). Deep learning has the advantage of automatically identifying the features from a given dataset for a given application. This is typically not feasible using traditional feature extraction techniques, which necessitate some prior knowledge for feature extraction. Deep learning–based approaches are also capable of discovering novel features that are suitable for specific uses; this is very helpful in classifying and predicting fatal disorders, which in turn helps to avoid their occurrence by early detection.

Although there has been significant improvement in medical analysis using machine learning, there are some problems that researchers still face. AD diagnosis is usually a multiclass classification problem, and the brain's structure has features that have high correlation, leading to certain disadvantages.

In this paper, we propose a robust extreme learning machine that is trained to identify the onset of Alzheimer's disease based on features obtained from longitudinal MRI scans. The proposed algorithm is compared with state-of-the-art machine learning classifiers based on metrics

such as accuracy, sensitivity, specificity, and F1 score. The paper is organized into the following sections. Section 2 presents an overview of the existing implementations and approaches pertaining to Alzheimer's disease identification and prediction of progression. Section 3 describes the methodology adopted in this paper, followed by an analysis of the results obtained in Section 4. Concluding remarks and future implementations are discussed in Section 5.

2. RELATED WORK

Recently, several computer vision algorithms (Suma et al., 2022a) have been used to carefully examine significant patterns in the brain tissues of AD patients and thereby derive methodologies to aid early disease detection and thus suppress its effects to some degree. The significance of influential biomarkers for image classification has been emphasized by numerous studies. The region-based method of feature extraction for AD classification using MRI and PET images has been suggested by Zhang et al. (2011), Liu et al. (2015), Suk et al. (2015), and Suma et al. (2022b). Human brain scans are used to map the first four moments and the entropy of histograms of anatomical regions of interest in order to find regional features. Numerous studies have used machine learning– and deep learning–based models to identify significant structural variations in the entorhinal cortex, hippocampus, and hippocampus regions of the brain by contrasting the brain tissues of AD patients and healthy individuals. Different imaging methods, including structural and functional magnetic resonance imaging, single-photon emission computed tomography (SPECT), diffusion tensor imaging (DTI), and positron emission tomography (PET), are used to detect changes in the brain tissues such as the degeneration of brain cells.

Payan and Montana (2015) proposed a pattern classification system and tested an algorithm that combines 3-D convolutional neural networks (CNNs) and sparse autoencoders for the prediction of AD. In order to combine multimodal features from MRI and PET scans and further categorize the severity of AD, Liu et al. (2015) found 83 ROIs. These ROIs will be used to train a neural network with multiple layers of autoencoders.

The quality of the data is important in computer vision applications, and smaller datasets frequently result in less accurate models. To address the issues caused by a limited dataset, researchers have turned their attention to recent developments in deep learning such as the use of residual dense neural networks and batch normalization. These also help with automatic feature generation, making them useful for training 3-D MRI images without intermediary feature extraction. A significant restriction of brain MRI data is related to its high dimensionality (Liu et al., 2020), which requires the use of significant computational power and a sizable dataset for deep neural network training in order to achieve a significant level of accuracy in classification tasks. Researchers have begun to use ensemble-based methods to overcome the constraints presented by computational complexity, high dimensionality, and lower accuracy due to the small datasets in tasks related to AD identification and classification.

A classification based on an ensemble of DL architectures is suggested by Ortiz et al. (2016) for the early detection of AD. The key feature of this method is that gray matter taken from the brain is divided into 3-D patches based on areas identified by the automated anatomical labeling atlas. The resulting patches are used to train numerous deep belief networks, and an ensemble voting method is used to carry out the final classification. To extract the resident features from MRI scans that are subsequently used for AD classification, Li et al. (2018) have suggested a multiple cluster–based DenseNet. Extraction of 3-D patches is carried out from each of the numerous local regions that make up the complete cerebral cortex. Using k-means clustering, these extracted patches are then further organized into clusters, and patch features are pulled

from each cluster using DenseNet. The classes are derived by clustering the features that were learned from each of these groupings. To determine the final classification for the image, the predictions received from each local region are combined. It is also claimed that segmentation and rigid registration as part of image preprocessing are not necessary for this technique of feature extraction.

Gorji and Kaabouch (2019) suggested a robust convolutional neural network architecture that is used to extract high-quality features from brain MRI scans and is used to categorize them as either healthy, early mild cognitive impairment, or late mild cognitive impairment (LMCI). The analysis is done in sagittal view, and the results show that LMCI and control normal (CN) groups can be classified with 94.54% accuracy. Basheera and Sai Ram (2019) suggested a method for extracting gray matter from the human brain and classifying the features that were extracted using a modified CNN architecture. The voxels were enhanced using a Gaussian filter, and then the skull stripping method was used to remove redundant tissues. A hybrid, independent, and improved component analysis was used to conduct voxel segmentation. CNN was supplied with segmented gray matter, and this method had a 90.47% accuracy rate.

Karim et al. (2017) proposed a multi-task binary classification, which is used to categorize subjects with mild cognitive impairment (MCI), Alzheimer's disease (AD), and normal control (NC) subjects. The suggested method has a 91% accuracy rate. Fusion on the fully connected layer as well as on a single-projection CNN output were both used in this technique.

Among all the ROIs, the hippocampus was found to be the most significant anatomical and impacted area in AD patients. Numerous studies (Amoroso et al., 2018; Beg et al., 2013; Chupin et al., 2009; Gerardin et al., 2009; Ho et al., 2011; Leung et al., 2010; Lindberg et al. 2012; Platero and Tobar, 2016; Shen et al., 2012; Li et al., 2012) have suggested different approaches for computing the volumetric characteristics and shape from bilateral hippocampi. Cao et al. (2018) has proposed a multi-task DL technique, which is used to segment the hippocampal region of the brain using MRI scans, along with clinical score regression. A method for performing hippocampal segmentation using anatomical and probabilistic priors was suggested by Chupin et al. (2009). A technique to gauge the hippocampal region's volumetric parameters was created by Platero and Tobar (2016). Using MRI, this data was used to separate AD/MCI patients from healthy groups. A fast, multi-atlas segmentation technique was used. For the purpose of diagnosing AD, shape analysis is frequently combined with volumetric analysis of the hippocampus to capture the complex morphology of this part of the brain.

In the following section, we describe our approach to identify and classify the onset of AD.

3. MATERIALS AND METHODS

Figure 1 represents the overall methodology adopted in this work. Initially, longitudinal MRI scans are preprocessed, and feature engineering steps are applied to obtain clinical features from the neuroimages. These features are obtained from the OASIS 2 database (Marcus et al., 2007) and are then iterated over a forward feature selection algorithm to select the most significant features that contribute to the identification of Alzheimer's disease. The features are then ranked based on priority and fed to an extreme learning machine network that achieves exceptional training speed by eliminating the gradient-based technique. The ELM network is then tuned using various hyperparameter tuning techniques for achieving robust performance. The proposed network is compared with off-the-shelf machine learning classifiers based on several evaluation metrics. In biomedical applications, sensitivity is an evaluation metric that is given the highest priority since it quantifies the ability of the algorithm to predict the proportion of patients

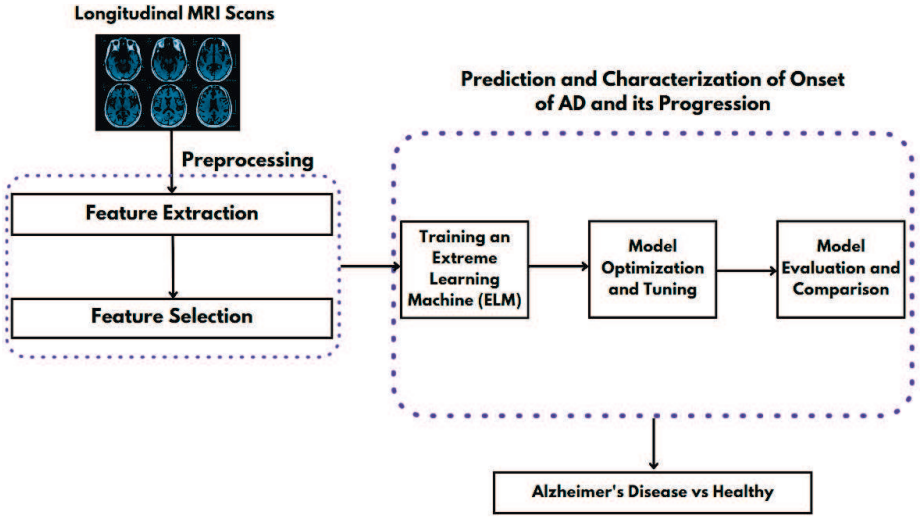

FIG. 1: Proposed methodology

possessing a particular disease. Hence, in our case, sensitivity of the model would quantify the ability of the model to rightly identify patients with the onset of Alzheimer's disease. The following subsections describe the methodology in detail.

3.1 Dataset Description

The dataset used in this paper consists of clinical parameters of 150 subjects obtained from longitudinal MRI scans. The subjects are aged between 60 and 96 years, each subject was reported to visit the hospital at least twice, and the duration of separation between the visits was at least one year, for a total of 373 imaging sessions. The subjects include both males and females, and for each of them, three or four T1-weighted MRI scans are included as a part of the dataset. Out of the 150 subjects under consideration, 72 subjects were identified as nondemented. Sixty-four subjects were characterized as having dementia during their initial visit to the hospital, out of which 51 subjects were diagnosed with Alzheimer's disease of mild to moderate severity. The remaining 14 subjects were identified as not having dementia during their initial visit but were later diagnosed with dementia during their subsequent visit. Table 1 presents an overview of the dataset adopted in this work.

3.2 Exploratory Data Analysis (EDA)

We adopted a range of data visualization techniques to obtain critical information about the distribution of classes, analyze the age groups that are most affected by Alzheimer's disease, and determine the correlation between various clinical features under consideration. A correlation analysis is performed alongside feature selection to discard redundant features. Figure 2 illustrates the distribution of classes in the dataset where most of the subjects are classified as not having dementia, followed by the demented group. The converted category represents the number of subjects that were nondemented during their initial visit but were identified as demented during their subsequent visit.

TABLE 1: Dataset description

Parameter	Subjects
Demented	64
Nondemented	72
Mild/moderate Alzheimer's disease	51
Demented during subsequent visit	14
Total	150

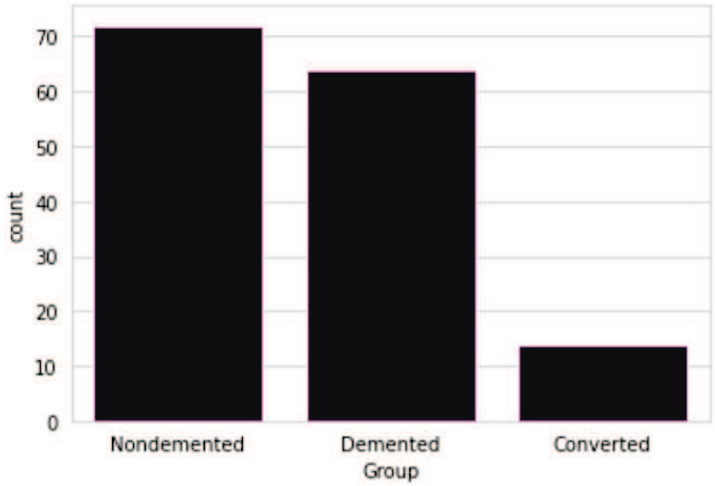

FIG. 2: Distribution of classes

Figure 3 portrays the distribution of demented and nondemented categories with respect to the gender ratio. It can be inferred that most females are nondemented as compared to males, which puts males at a slightly higher risk of dementia.

Analyzing age groups that are most likely to be affected with Alzheimer's disease is the first step toward developing algorithms for early identification of its onset. Figure 4 depicts a plot of the distribution of age groups and the corresponding number of subjects belonging to that age group to have been diagnosed with Alzheimer's disease. Subjects in the age group between 73 and 78 years are seen to be the most affected with dementia and its progression.

3.3 Feature Engineering

This subsection presents a brief overview of the different features adopted in this paper for achieving the task of early detection of the onset of Alzheimer's disease.

3.3.1 Clinical Dementia Rating (CDR)

This metric is obtained from interviews of the subject and the informant by a clinician who characterizes a subject based on six different parameters, namely: memory, orientation, problem solving, community affairs, hobbies, and personal care. The rating is based on a 5-point scale, where a CDR of 0 indicates no dementia, 0.5 indicates very mild AD, 1 indicates mild AD, and

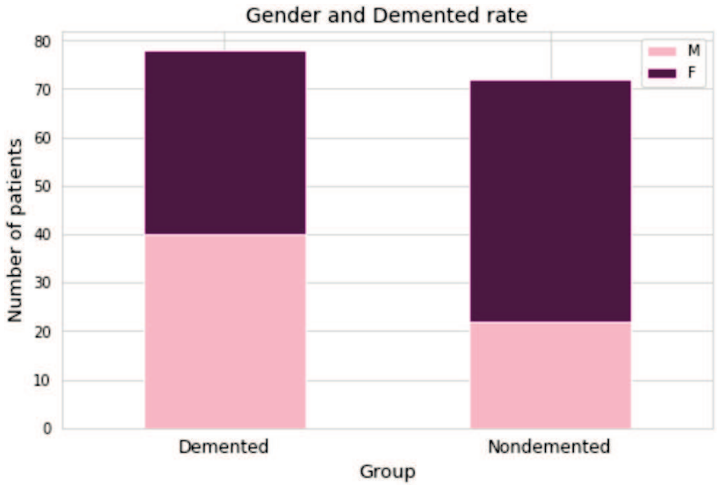

FIG. 3: Distribution of classes based on gender ratio

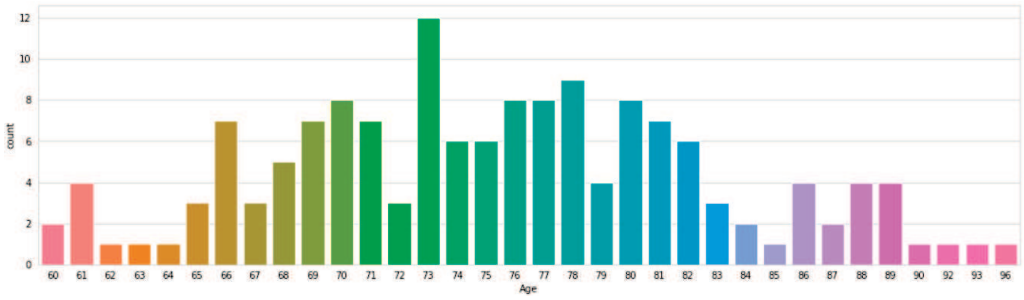

FIG. 4: Analysis of age groups affected with dementia

2 indicates moderate AD. The scores for the six individual parameters are provided initially, and a cumulative score is calculated using an algorithm. This serves as an important parameter in determining the level of dementia a subject suffers from and its progression.

3.3.2 Estimated Total Intracranial Volume (ETIV)

In the study of neurodegenerative disorders, it is important to perform volumetric analysis of the brain to estimate the maximum premorbid brain volume. But on analysis, it was found that the average value of eTIV neither varied with time nor with the subjects. On an average, men showed ∼12% larger eTIV than women.

3.3.3 Cognitive Impairment

Cognitive impairment is a state of the mind that causes retarded memory, difficulty in comprehending new information, and lack of decision-making ability. On a scale of 30, a normal person is expected to have a cognition score of 24, whereas a score of 19 to 23 indicates mild cognitive

impairment, a score of 10 to 18 indicates a moderate cognitive impairment, and a score below 9 indicates severe cognitive impairment.

3.3.4 Mini-Mental State Examination Score (MMSE)

The Mini-Mental State Examination (MMSE) is a 30-point (0 being the worst and 30 being the best) questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. It is commonly used in medicine and allied health to screen for dementia. It is also used to estimate the severity and progression of cognitive impairment and to follow the course of cognitive changes in an individual over time, thus making it an effective way to document an individual's response to treatment.

Alongside the above clinical parameters, the normalized whole-brain volume (nWBV) and the atlas scaling factor (ASF) are also considered. The nWBV is a percentage of all the volumetric pixels (voxels) in the atlas masked image that are classified as gray or white matter in the tissue segmentation process. The ASF is a computed scaling factor used for transforming the native brain space and skull to the atlas target.

To identify potential relationships between the features under consideration and to determine outliers and redundant features, a correlation matrix as shown in Fig. 5 is plotted. The matrix depicts the fact that the delay in MRI imaging sessions that a particular subject has undergone is directly related to the number of visits to the hospital. Another set of highly correlated features is the clinical dementia rating index and the category to which a subject belongs (nondemented, demented, or converted).

3.4 Comparison of Classification Algorithms

In this paper, we train various binary classification algorithms for identification of Alzheimer's disease and its onset. The algorithms are evaluated under different metrics, which are described in the following section, and the most robust model is selected based on its performance on these metrics as well as the computational complexity of the algorithm.

3.4.1 Logistic Regression (LR)

Logistic regression (LR) is a statistical model that is used in binary classification problems and helps model the target variable using the logistic or the sigmoid function, given by

$$F(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}. \tag{1}$$

Equation (1) converts the range of values that the model deals with from $(-k, k)$ to $(0, 1)$. Assuming $p(x)$ to be a linear function in x, and applying a logit transformation, we obtain the following mathematical model:

$$\log \frac{p(x)}{1-p(x)} = a_0 + a * x. \tag{2}$$

Since the output of the logistic regression algorithm is a class probability, Eq. (2) can be fit using a likelihood estimation, given by

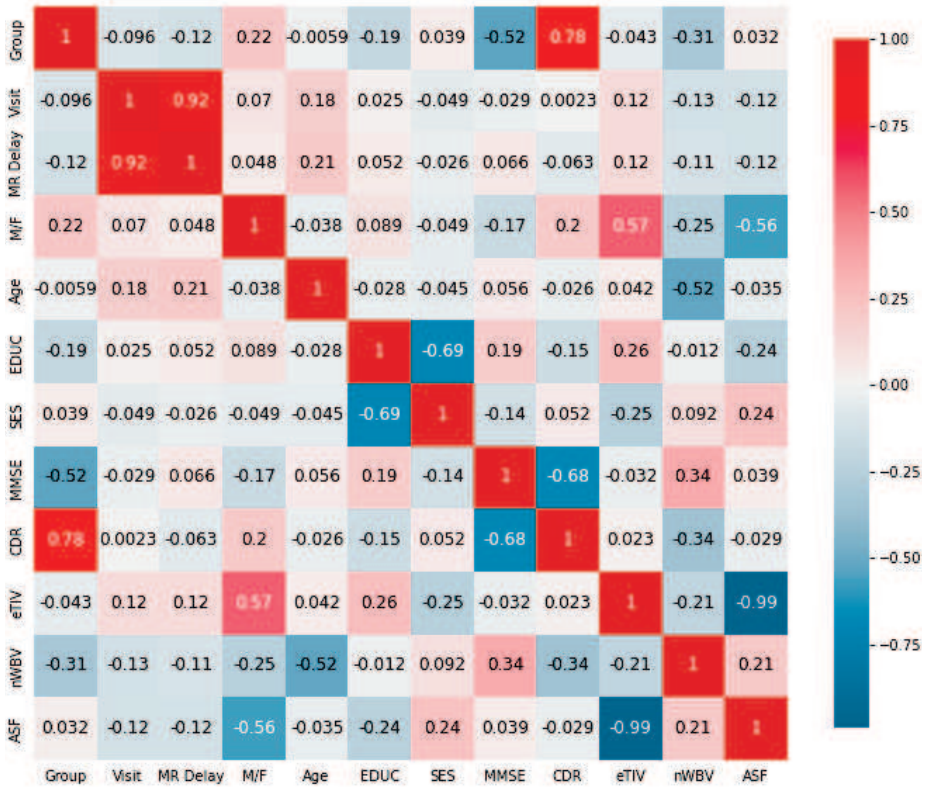

FIG. 5: Correlation matrix

$$L(a_0, a) = \prod_{i=1}^{n} p(x_i)^{y_i} [1 - p(x_i)]^{(1-y_i)}, \tag{3}$$

where the output is p if $y = 1$ or $1 - p$ if $y = 0$.

3.4.2 k-Nearest Neighbors (KNNs)

The concept behind the working of k-nearest neighbors (kNNs) can be described as the process of finding the closest point to the given input. The algorithm classifies unseen test points using majority votes from the k-nearest neighbors. The first step in the algorithm is to apply a transformation to the data points to convert them into vectors. The algorithm works by finding the distance between each data point and the test data point and then finds the probability of the points resembling the test data. Finally, classification is based on the points that share highest probabilities. After the computation of the Euclidean distance, the input x is assigned to the class with the highest probability, given by

$$P(y = j \mid X = x) = \frac{1}{k} \sum_{i \in A} I(y^i = j). \tag{4}$$

3.4.3 Naive Bayes (NB)

Naive Bayes (NB) is a supervised learning algorithm that applies the Bayes theorem to predict output class probabilities, given by

$$p(A \mid B) = \frac{p(A)p(B \mid A)}{p(B)}. \tag{5}$$

The NB algorithm finds the class of observation y_i, given the set of features. Assuming the set of features are given as $\{x_1, x_2 \dots x_n\}$, the NB algorithm can be formulated as follows:

$$p(y_i \mid y_i x_1, x_2 \dots x_n x_1, x_2 \dots x_n) = \frac{p(y_i)p(x_1, x_2 \dots x_n \mid y_i)}{p(x_1, x_2 \dots x_n)}. \tag{6}$$

3.4.4 Support Vector Machines (SVMs)

The major goal of support vector machines (SVMs) is to find the most optimal hyperplane that separates the data points with maximum margin. The algorithm orients the hyperplane at the maximum distance from the closest member of both classes. To achieve this, SVMs solve the following optimization problem:

$$\begin{gathered} \min_{w,b} \frac{1}{2}||w||^2 \\ \text{subject to } y_i(x_i + w * b) \geq 1 \\ \text{for } i = 1, 2, 3 \dots n. \end{gathered} \tag{7}$$

3.4.5 Decision Trees (DT)

The algorithm selects the best attribute and assigns it as the decision node, using the attribute selection measure (ASM) to split the given records. The ASM is a heuristic for selecting the best criterion to partition the data. The ASM ranks each feature based on a set of rules, and the best attribute or feature is chosen as the splitting attribute. In this paper, we use the information gain as an ASM to iteratively select the best set of attributes, which is defined as the difference between the entropy before split and average entropy after split, based on a selected attribute. Information gain is mathematically described as

$$I(T) = -\sum_{i=1}^{m} p_i \log_2 p_i \,, \tag{8}$$

where p_i is the probability that a tuple T belongs to class C_i.

3.4.6 Random Forest Classifier (RFC)

Random forest classifiers (RFCs) are an ensemble of decision trees, where the final prediction of the algorithm is based on the individual predictions from each decision tree. This is mathematically formulated as

$$RFf_i = \frac{\sum_{j \in T} normf_{ij}}{T}, \tag{9}$$

where T is the total number of decision trees, $norm f_{ij}$ is the normalized feature i in tree j, and RFf_i is the importance of feature i calculated over all trees in the RFC model.

3.4.7 Gradient Boosting (GB)

The gradient boosting (GB) classifier is an ensemble method that combines predictions from several base classifiers. The algorithm uses the cross-entropy loss function, which is given by

$$L = -[y_i \log_{10} p \ + (1 - y_i) \log_{10}(1 - p \,)]. \tag{10}$$

3.5 Extreme Learning Machine (ELM)

Conventional neural networks employ gradient-based training algorithms, such as back propagation, that often do not achieve the global optimal solution. These algorithms are dependent on the initialization of parameters and the complexity of the feature space and hence often converge at local extrema. Unlike traditional algorithms, extreme learning machines (ELMs) assign random weights to the nodes between the input layer and the hidden layer, and these values are kept fixed throughout the training process rather than iteratively updating the weights (Wang et al., 2022). ELMs only learn the weights between the hidden layer and the output layer, and hence they tend to converge much faster than the traditional gradient-based algorithms. ELMs achieve a comparable performance to other deep neural networks while maintaining a lower computational complexity. The process of ELM training can be categorized into two stages, namely: random parameter initialization and linear parameter solution. Random weights w_i and bias b_i are initialized in the hidden layer and remain unchanged during the training process. The input vector is then mapped into a feature space with stochastic parameter values and nonlinear activation functions, which shows superior performance to iteratively trained parameters. The ELM achieves robust generalization capability by using a piecewise approximation for a nonlinear continuous activation function. In the next step, the output weight matrix β is obtained using the Moore–Penrose inverse, and it is then formulated as a linear problem.

Consider the following training set:

$$x = \{x_1, x_2, x_3, \ldots x_N\}, \tag{11}$$

where N represents the total number of samples in the training set, and x_i represents the ith training sample.

The set of targets are represented as:

$$t = \{t_1, t_2, t_3, \ldots t_N\}, \tag{12}$$

where t_i is an array of binary values containing the class values for the x_i sample from the training set. The final output of the ELM is denoted as

$$o_j = \sum_{i=1}^{L} \beta_i g_i(x) = \sum_{i=1}^{L} \beta_i g_i(w_i * x_j + b_i), \tag{13}$$

where $j = 1, 2, 3, \ldots N$, $g_i(x)$ is the activation function in the hidden layer of the network. The objective function of the ELM is the one that minimizes the loss between the predicted and actual values. The commonly used function is the mean squared error (MSE), which is defined as

$$MSE = \sum_{i=1}^{N} (t_{ij} - o_{ij})^2, \tag{14}$$

where $j = 1, 2, 3, \ldots m$, N is the total number of training samples, and i and j are the indices for the training sample and the output node, respectively.

Initially the weights matrix is constructed for the input layer, given by the following:

$$W = \begin{bmatrix} rand & \cdots & rand \\ \vdots & \ddots & \vdots \\ rand & \cdots & rand \end{bmatrix}. \tag{15}$$

The output matrix of the hidden layer is calculated, and an activation function is applied, which introduces nonlinearities in the network.

$$H = W * X. \tag{16}$$

The Moore–Penrose pseudoinverse is computed for the output matrix of the hidden layer, which is mathematically described as follows:

$$H^{+} = \left(H^{T} * H\right)^{-1} * H^{T}. \tag{17}$$

The output weight matrix β is computed as

$$\beta = H^{+} * T. \tag{18}$$

Equation (18) is used to compute the output matrix for the testing data, and a resultant matrix is obtained, given as follows:

$$O = H * \beta. \tag{19}$$

The Softmax algorithm is used to obtain the final class probabilities, and the output matrix O is compared with the set of targets t to evaluate the performance of the algorithm.

3.6 Evaluation Metrics

To select the most robust model for Alzheimer's disease classification, a range of evaluation metrics are adopted. The following metrics are considered in this work. Accuracy is a measure of the number of total observations (both positive and negative) that are correctly classified.

In terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), accuracy can be computed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{20}$$

Specificity and sensitivity provide valuable performance benchmarks for evaluating the predictive performance of binary classifiers. Specificity quantifies the percentage of true negatives (TN), which in our case is the number of correct predictions made on the nondemented subjects. It is calculated as

$$Specificity = \frac{TN}{TN + FP}. \tag{21}$$

Sensitivity, on the other hand, quantifies the proportion of true positives (TP), or the recall rate, which is the number of correct predictions on the demented records. In this paper, we prioritize sensitivity over specificity since it is of prime importance to rightly predict a demented record rather than failing to classify a nondemented record. Sensitivity is mathematically calculated as

$$Sensitivity = \frac{TP}{TP + FN}. \tag{22}$$

$F1$ score is defined as the harmonic mean of the precision and recall that quantifies the performance of a classification algorithm. It is mathematically given as

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}, \tag{23}$$

where precision is the percentage of positive samples that are rightly predicted, which is mathematically given by

$$Precision = \frac{TP}{TP + FP}. \tag{24}$$

4. RESULTS AND DISCUSSION

We have compared the performance of several machine learning classifiers, namely: random forest classifier (RFC), decision trees (DT), support vector machines (SVM), naïve Bayes classifier (NB), k-nearest neighbors (kNNs), logistic regression (LR), gradient boosting classifier (GB), and extreme learning machines (ELMs). Table 2 presents a performance comparison of the various classifiers under consideration, and on experimentation it was found that the proposed extreme learning machines outperformed the contemporary classifiers.

As described in the previous section, the algorithms were evaluated on metrics such as accuracy, sensitivity, specificity, and $F1$ score. The performance of the algorithms based on these parameters is presented in Tables 2 and 3. From Table 2 it can be inferred that while ELMs

TABLE 2: Comparison of sensitivity and specificity of classification algorithms

Algorithm	Sensitivity	Specificity
LR	0.482	0.567
kNN	0.511	0.613
NB	0.679	0.689
SVM	0.734	0.787
DT	0.831	0.802
RFC	0.889	0.864
GB	0.901	0.897
ELMs	**0.956**	**0.962**

TABLE 3: Comparison of $F1$ score and accuracy of classification algorithms

Algorithm	$F1$ score	Accuracy
LR	0.321	49.2%
kNN	0.495	53.8%
NB	0.689	60.0%
SVM	0.786	72.1%
DT	0.851	86.4%
RFC	0.901	89.1%
GB	0.926	90.2%
ELMs	**0.972**	**98.3%**

have outperformed several other classification algorithms, their robust performance in AD identification and classification can be attributed to choice of activation function employed, features selected, and the data preprocessing methods adopted. Since ELMs do not adopt a gradient-based training algorithm, they tend to converge at the global minima rather than the local minima. This is one of the major reasons for their robust performance.

Another major observation that can be made from Tables 2 and 3 is that ensemble-based methods such as DT, RFC, and GB tend to perform better than SVM, NB, kNN, and LR.

This can be attributed to the fact that ensemble-based methods combine the predictions from several base classifiers and make a final prediction by combining the outcomes of several of these classifiers.

This is usually achieved by hard or soft voting, which are based on majority votes received from each of the base models. The accuracy of kNNs may tend to drop with higher dimensionality of data since they employ Euclidean distance to assign class probabilities. In such cases of multidimensional data, SVMs tend to perform better. The optimization problem in SVMs involves the finding of the most optimal hyperplane that lies at the maximum distance from the closest data points to it. SVMs are also capable of dealing with complex structures of data by adopting kernels of higher-order polynomials.

Figure 6 presents a visualization of the AUC–ROC curve of the ELM algorithm. The receiver operator curve (ROC) summarizes the classifier's performance based on tradeoffs between the error rates of true positives and true negatives. The area under the curve (AUC) is a performance metric that quantifies the probability of a classifier performing better in positive instances than a negative instance. The ELM algorithm presented the highest AUC of 0.84, indicating that the algorithm performs better within the positive instances (in 84% of the cases) than the negative instances.

5. CONCLUSION

Since Alzheimer's is a neurodegenerative disorder that progresses over time, no prescribed medication can fully stop the disease's progression. Therefore, it has become imperative to create intelligent systems and techniques for its early detection. For this purpose, we propose an extreme learning machine algorithm that outperforms contemporary binary classifiers in terms of

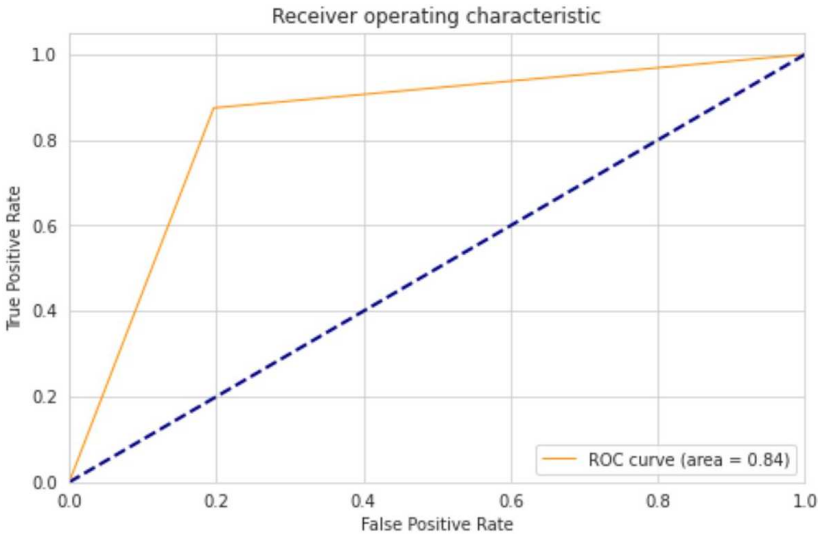

FIG. 6: ROC characteristic curve of ELM

accuracy, sensitivity, and specificity. Extreme learning machines also outperform other algorithms in terms of computational efficiency and easy deployment on edge devices. ELMs also prove to overcome the disadvantages of traditional neural networks that adopt gradient-based training methods and converge at the local minima. The proposed model is designed to be implemented in healthcare systems that can alert clinicians and doctors of the patient's onset of AD and help in providing early diagnosis and treatment.

REFERENCES

Aldhyani, T.H.H., Alshebami, A.S., and Alzahrani, M.Y., Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms, *J. Healthcare Eng.*, vol. **2020**, p. 4984967, 2020.

Amoroso, N., Rocca, M., Bellotti, R., Fanizzi, A., Monaco, A., and Tangaro, S., Alzheimer's Disease Diagnosis Based on the Hippocampal Unified Multi-Atlas Network (HUMAN) Algorithm, *Biomed. Eng. Online*, vol. **17**, no. 6, 2018. DOI: 10.1186/s12938-018-0439-y

Bartos, A., Gregus, D., Ibrahim, I., and Tintěra, J., Brain Volumes and Their Ratios in Alzheimer's Disease on Magnetic Resonance Imaging Segmented Using Freesurfer 6.0, *Psych. Res. Neuroimaging*, vol. **287**, pp. 70–74, 2019.

Basheera, S. and Sai Ram, M.S., Convolution Neural Network–Based Alzheimer's Disease Classification Using Hybrid Enhanced Independent Component Analysis Based Segmented Gray Matter of T2 Weighted Magnetic Resonance Imaging with Clinical Valuation, *Alzheimer's Dementia: Translat. Res. Clin. Intervent.*, vol. **5**, pp. 974–986, 2019.

Beg, M.F., Raamana, P.R., Barbieri, S., and Wang, L., Comparison of Four Shape Features for Detecting Hippocampal Shape Changes in Early Alzheimer's, *Stat. Methods Med. Res.*, vol. **22**, no. 4, pp. 439–462, 2013.

Cao, L., Li, L., Zheng, J., Fan, X., Yin, F., Shen, H., and Zhang, J., Multi-Task Neural Networks for Joint Hippocampus Segmentation and Clinical Score Regression, *Multimedia Tools Appl.*, vol. **77**, pp. 29669–29686, 2018.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., and Colliot, O., Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data from ADNI, *Hippocampus*, vol. **19**, no. 6, pp. 579–587, 2009.

De Strooper, B. and Karran, E., The Cellular Phase of Alzheimer's Disease, *Cell*, vol. **164**, no. 4, pp. 603–615, 2016. DOI: 10.1016/j.cell.2015.12.056

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., and Collins, D.L., Unbiased Average Age-Appropriate Atlases for Pediatric Studies, *Neuroimage*, vol. **54**, no. 1, pp. 313–327, 2011.

Galvin, J.E., Prevention of Alzheimer's Disease: Lessons Learned and Applied, *J. Am. Geriatr. Soc.*, vol. **65**, no. 10, pp. 2128–2133, 2017. DOI: 10.1111/jgs.14997

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., et al., Multidimensional Classification of Hippocampal Shape Features Discriminates Alzheimer's Disease and Mild Cognitive Impairment from Normal Aging, *Neuroimage*, vol. **47**, no. 4, pp. 1476–1486, 2009.

Gorji, H.T. and Kaabouch, N., A Deep Learning Approach for Diagnosis of Mild Cognitive Impairment Based on MRI Images, *Brain Sci.*, vol. **9**, no. 9, pp. 217–231, 2019.

Ho, A.J., Raji, C.A., Saharan, P., DeGiorgio, A., Madsen, S.K., Hibar, D.P., Stein, J.L., et al., Hippocampal Volume is Related to Body Mass Index in Alzheimer's Disease, *Neuroreport*, vol. **22**, no. 1, pp. 10–14, 2011.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S., Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images, *Neuroimage*, vol. **17**, no. 2, pp. 825–841, 2002. DOI: 10.1006/nimg.2002.1132

Karim, A., Jenny, B.P., and Karim, A., Classification of sMRI for Alzheimer's Disease Diagnosis with CNN: Single Siamese Networks with 2D+? Approach and Fusion on ADNI, *Proc. of the 2017 ACM on Int. Conf. on Multimedia Retrieval*, Bucharest, Romania, pp. 494–498, 2017.

Korolev, S., Safiullin, A., Belyaev, M., and Dodonova, Y., Residual and Plain Convolutional Neural Networks for 3D Brain MRI Classification, *2017 IEEE 14th Int. Symp. on Biomedical Imaging (ISBI 2017)*, Melbourne, Australia, pp. 835–838, 2017.

Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schuff, N., Fox, N.C., Ourselin, S., and Alzheimer's Disease Neuroimaging Initiative, Automated Cross-Sectional and Longitudinal Hippocampal Volume Measurement in Mild Cognitive Impairment and Alzheimer's Disease, *Neuroimage*, vol. **51**, no. 4, pp. 1345–1359, 2010.

Li, F., Liu, M., and Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Diagnosis Based on Multiple Cluster Dense Convolutional Networks, *Comput. Med. Imaging Graphics*, vol. **70**, pp. 101–110, 2018.

Li, X., Xia, H., Zhou, Z., and Tong, L., 3D Textures Analysis of Hippocampus Based on MR Images in Patients with Alzheimer Disease and Mild Cognitive Impairment, *J. Beijing Univ. Technol.*, pp. 942–948, 2012.

Lindberg, O., Walterfang, M., Looi, J.C., Malykhin, N., Ostberg, P., Zandbelt, B., Styner, M., Paniagua, B., Velakoulis, D., and Orndahl, E., Hippocampal Shape Analysis in Alzheimer's Disease and Frontotemporal Lobar Degeneration Subtypes, *J. Alzheimers Dis.*, vol. **30**, no. 2, pp. 355–365, 2012.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., and Sánchez, C.I., A Survey on Deep Learning in Medical Image Analysis, *Med. Image Anal.*, pp. 60–88, 2017.

Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., and Fulham, M.J., Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease, *IEEE Trans. Biomed. Eng.*, vol. **62**, no. 4, pp. 1132–1140, 2015.

Liu, M., Li, F., Yan, H., Wang, K., and Ma, Y., A Multi-Model Deep Convolutional Neural Network for Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease, *Neuroimage*, 2020.

Marcus, D., Buckner, R., Csernansky, J., and Morris, J., Principal Investigators, OASIS-2: Longitudinal, P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382, 2007.

Ortiz, A., Munilla, J., Górriz, J.M., and Ramírez, J., Ensembles of Deep Learning Architectures for the Early Diagnosis of Alzheimer's Disease, *Int. J. Neural Syst.*, vol. **26**, no. 7, 2016.

Payan, A. and Montana, G., Predicting Alzheimer's Disease: A Neuroimaging Study with 3D Convolutional Neural Networks, *Proc. Int. Conf. Pattern Recog. Appl. Methods*, 2015.

Platero, C. and Tobar, M.C., A Fast Approach for Hippocampal Segmentation from T1-MRI for Predicting Progression in Alzheimer's Disease from Elderly Controls, *J. Neurosci. Methods*, vol. **270**, pp. 61–75, 2016.

Qiu, S., Chang, G.H., Panagia, M., Gopal, D.M., Au, R., and Kolachalama, V.B., Fusion of Deep Learning Models of MRI Scans, Mini–Mental State Examination, and Logical Memory Test Enhances Diagnosis of Mild Cognitive Impairment, *Alzheimer's Dementia: Diagnosis Assess. Dis. Monitor.*, vol. **10**, pp. 737–749, 2018.

Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., and Davatzikos, C., A Review on Neuroimaging-Based Classification Studies and Associated Feature Extraction Methods for Alzheimer's Disease and Its Prodromal Stages, *Neuroimage*, vol. **155**, pp. 530–548, 2017. DOI: 10.1016/j.neuroimage.2017.03.057

Shen, K.K., Fripp, J., Mériaudeau, F., Chételat, G., Salvado, O., Bourgeat, P., and Alzheimer's Disease Neuroimaging Initiative, Detecting Global and Local Hippocampal Shape Changes in Alzheimer's Disease Using Statistical Shape Models, *Neuroimage*, vol. **59**, no. 3, pp. 2155–2166, 2012.

Suk, H.I., Lee, S.W., Shen, D., and Alzheimer's Disease Neuroimaging Initiative, Latent Feature Representation with Stacked Autoencoder for AD/MCI Diagnosis, *Brain Struct. Funct.*, vol. **220**, no. 2, pp. 841–859, 2015.

Suma, K.V., Raghavan, D., and Ganesh, P., Deep Learning for Alzheimer's Disease Detection Using Multimodal MRI-PET Fusion, *2022 4th Int. Conf. on Circuits, Control, Communication and Computing (I4C)*, Bangalore, India, pp. 287–292, 2022a. DOI: 10.1109/I4C57141.2022.10057623

Suma, K.V., Selvi, S., Nanda, P., Shetty, M., Vikas, M., and Awasthi, K., Deep Learning Approach to Nailfold Capillaroscopy Based Diabetes Mellitus Detection, *Int. J. Online Biomed. Eng.*, vol. **18**, no. 6, pp. 95–109, 2022b.

Veitch, D.P., Weiner, M.W., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jr., Jagust, W., Morris, J.C., Petersen, R.C., Saykin, A.J., Shaw, L.M., Toga, A.W., and Trojanowski, J.Q., Alzheimer's Disease Neuroimaging Initiative. Understanding Disease Progression and Improving Alzheimer's Disease Clinical Trials: Recent Highlights from the Alzheimer's Disease Neuroimaging Initiative, *Alzheimer's Dement.*, vol. **15**, no. 1, pp. 106–152, 2019. DOI: 10.1016/j.jalz.2018.08.005

Vidoni, E.D., The Whole Brain Atlas, *J. Neurol. Phys. Therapy*, vol. **36**, no. 2, p. 108, 2012. DOI: 10.1097/NPT.0b013e3182563795

Wang, J., Lu, S., Wang, S.H., and Zhang, Y.D., A Review on Extreme Learning Machine, *Multimedia Tools Appl.*, vol. **81**, pp. 41611–41660, 2022. DOI: 10.1007/s11042-021-11007-7

Zarandi, M.H.F., Zarinbal, M., and Izadi, M., Systematic Image Processing for Diagnosing Brain Tumors: A Type-II Fuzzy Expert System Approach, *Appl. Soft Comput.*, vol. **11**, no. 1, pp. 285–294, 2011.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D., Multimodal Classification of Alzheimer's Disease and Mild Cognitive Impairment, *Neuroimage*, vol. **55**, no. 3, pp. 856–867, 2011.

Zhang, Y., Brady, M., and Smith, S., Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, *IEEE Trans. Med. Imaging*, vol. **20**, no. 1, pp. 45–57, 2001. DOI: 10.1109/42.906424