JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

An Optimized Machine Learning Framework for Detecting Alzheimer's Disease By MRI

Dr. T. S. Suganya
Assistant Professor, Dept of CA
SRMIST
Ramapuram Campus,India
tssuganya07@gmail.com

Dr. K. Geetha
Professor, Dept of CSE
Excel Engineering college
Kumarapalayam, Namakkal, India
geetharajsri@gmail.com

Dr. C. Rajan
Professor, Dept of IT
K.S.R College of Technology
Namakkal,India
rajancsg@gmail.com

ABSTRACT

Machine learning has extensive application in diverse medical fields.With advancements in medical technologies, access has been given to data for the identification of diseases in theirearly stages. Alzheimer's Disease (AD) is a chronic illnessthat will cause degeneration of the brain cells and ultimately will lead to memoryloss. AD causedcognitive mental problems like forgetfulness and confusion, as well as other symptoms such aspsychologicaland behavioralproblems, are further recommended to undergo test procedures usingneuroimagingtechniques. This work's objective is to utilize the machinelearning algorithms for processing the data acquired via neuroimaging technologies for early-stage AD detection. The framework extracts featuresusingcurvelet transform from MRI brain image. This work will also present the Decision Tree, the Adaptive Boosting (AdaBoost), and the Extreme Gradient Boosting (XGBoost) classifiers. In machine learning, Population-Based Incremental Learning (PBIL) is an optimization algorithm, in spite of being simpler than a conventional genetic algorithm, the PBIL algorithm is able to achieve much better results in several cases.PBIL is used to optimize the AdaBoost and XGBoost classifiers to improve AD classification. The experimental outcomes will demonstrate the proposed approach's superior performance over that of other existing approaches.

Keywords: *Alzheimer's disease (AD), Machine Learning, Curvelet Transform, Decision Tree, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), and Population-Based Incremental Learning (PBIL).*

1 INTRODUCTION

Alois Alzheimer, a German physicist and neuro-pathologist, was the first person to identify as well as to discuss about theAlzheimer'sDisease (AD). As per the World Alzheimer Report, around 50 millionpeopleacross the globe were affected by dementia in the year 2018,andabout two-third of that population were suffering from AD. There will be about 152 million AD patients in the year 2050, and this disease's cost has been forecasted to be USD 2 trillion in the year 2030. At present, the AD treatment is predominantly involved with the usage of either Alzheimer's disease-modifying or delaying drugs instead of drugs which are able to either reverse or permanently stop the disease's progression. Hence, it is essential for early-stage AD prediction so as to make it feasible to delay the effects of the disease [1].

AD is characterized by the loss of neurons as well as synapses within the cerebral cortex,and also specific subcortical regions, that will result in gross atrophy of the affected regions, which is inclusive of degeneration in the temporal lobe andparietal lobe, and also portions of the frontal cortex as well as the cingulate gyrus. Earlier studies have shown the correlation between several impairments in the AD as well as the atrophy in many regions such as the amygdala,the temporal lobe, and the hippocampus. These characteristics are used for delineating the AD patients from the normal patients.

The key focus of the researchers is to monitor the change in a patient's health, the disease's clinical progression as well as reaction to the therapy. However, they find it most cumbersome to identify relevant bio-markers which are good representations of the AD as well as the Mild Cognitive Impairment (MCI). The researchers' objective is inclusive of diagnosingearly-stage AD as well as identifyingthe individuals who are at most risk for AD development. Magnetic Resonance Imaging (MRI) is employed by physicians to diagnose AD. The multi-class classification of AD, MCI as well as Normal Control (NC) will employ as biomarkers the individual or combined structural MRIbiomarkerslike the hippocampus's shape as well as texture, corticalmeasurements, and volume measurements [2].

The MRI's key role in the AD analysis is to assess the volume alteration in the characteristic positions so as to provide up to 87% of analytical accuracy of up to 87%. Quite often, the appraisal is carried out on the mesial temporal lobe atrophyand the temporoparietalcortical atrophy. Direct or indirect estimation is done for the mesial temporal lobe atrophy. While the direct estimation is based on measuring the volume loss of hippocampal orparahippocampal tissue, the indirect estimation is dependent ontheparahippocampalfissures' magnification. Normally, analysis of these estimations is done along with the medial temporalatrophy score, that has been proved to be predictive of the progression from MCI to dementia[3].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

Over the last few decades, techniquesof machine learning have garnered much interestandalso have been widely utilized in medical/neuroimaging applications. There are two distinct types of machine learning techniques: the supervised techniques, and the unsupervised techniques.Supervised learning will separate the groups of data on the basis of a trainingset. These techniques are based on the premise that a supervisor will instruct the classifier to relate the unlabeleddata with a class label by means of the training set. That is, the supervised learning will offer a mapping between an input variable set as well as an output variable set which can be later employed to predict the unseen data's outputs. The mapping's basis will be on the features that have been acquired fromthe data [4].

With techniques of machine learning,it is possible to analyzehighdimensionaldataas well as automated classification for learning the complicated structural changes of the complex patterns in the brain images. Numerous approaches such as the deep learning networks as well as the techniques of machine learning have been employed for the development of early AD classificationon an individual basis[5]. There is acquisition of predefined features such as voxeland regional measures for combining multiple algorithms together with classifiers like the SupportVector Machines (SVM), decision trees. Even though several existing approaches are being employed for AD classification,they do suffer from drawbacks such as over-fitting as well as data imbalance problems. Moreover, these approaches are vulnerable to getting easilytrappedin the local optima. For resolution of the ADclassification'sproblem of local optima, this work has offered the proposals of the PBIL with AdaBoost and XGBoost techniques for AD detection as well as diagnosis from the MRI framework to boost theclassificationperformance.The remaining part of the paper is organized into four sections. Section twopresents some of works available in the literature. Section threedetails the various techniques used in the investigation. Section fourpresents the experimental results and section five concludes the work.

2 RELATED WORKS

The application of machine learning research to the MRI techniques has helped in contributing to a more rapid AD diagnosis as well as to prediction of this disease's evolution. Also, AD screening data as well as machine learning classifiers could be utilized for the prediction of individual dementia in older adults. With prediction of the AD subject status, a patient's MRI demographic information as well as pre-existing conditions are able to aid in boosting the classifier performance. Battineni et al., [6] had proposed a supervised learning classifier-based framework for categorization of the dementia subjects as either non-AD or AD on the basis of the longitudinal brain MRI features. There was integration of six distinct supervised classifiers the AD subject classification.Furthermore, the simulated outcomes had showed that the gradient boosting algorithm had far exceeded the performance of other models with an accuracy of 97.58%.

Garg et al., [7] had utilized the mean energies of both real as well as imaginary components of the complex wavelet coefficients which were individually taken as the features for the classification of AD and normal control. The experimental outcomes with seven distinct classifiers had demonstrated that,themean energy which was acquired were separately taken as features when compared with the mean energy of the real component. The proposed approach was employed with the Open Access Series of Imaging Studies (OASIS) dataset and the K-Nearest Neighbor (KNN) classifier,the accomplished prediction accuracy, sensitivity, and specificity were 97.3%, 98.2%, and 96.4%, respectively.Moreover, when the same dataset was employed along with the Cubic Support Vector Machine (CSVM) classifier,the accomplished prediction accuracy, sensitivity, and specificity were 97.3%, 96%, and 99%, respectively. In addition, when compared with other existing complex algorithms,theproposed approach was found to be much simpler and also offered better results.

Kaplan et al., [8] had presented feed-forward Local Phase Quantization Network (LPQNet),that utilized brain images automatically for detection of AD. The proposed LPQNetwas composed of the following distinct phases: (i) the multi-level feature generation phase which was based on the LPQ as well as the average pooling, (ii) the feature selection phase which used the Neighborhood Component Analysis (NCA), and (iii) the classification phase. The LPQNet's key purpose was to arrive at high accuracy with minimal complexity of the computation. Moreover, the LPQNetwould generate the features on six different levels. Hence, out of 1536 features which were generated from an image, only the most important 256 features out of all the 1536 features were picked. Afterwards, classification of these 256 features was done on the standard classifiers so as to indicate the classification abilities of both the generated as well as the picked features by the LPQNet.

Karadayi-Ataş et al., [9] had presented a Variable Neighborhood Search (VNS) based framework which had employed MRI data for the diagnosis of early conversion from MCI to AD. The proposed framework's construction was made up of three key phases: the datasetpreparation, the feature selection, and the classification. The VNS algorithm would pick the most predictive MRI features for the classification. Afterwards, the classification fo the chosen features was done with the use of a Linear Support Vector Machine (LSVM). For this study,all the data was acquired from the ADNI database,which in turn was composed of 860 subjects, 8 distinct monthly periods as well as 286 features for every

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

period. It was evident from the experimental results that, in comparison with earlier research studies, the proposed framework had superior performance with regards to accuracy, sensitivity as well as specificity values. Moreover, this work's results had demonstrated the proposed framework's tremendous potential for early prediction as well as detection of the MCI to the AD conversion.

Knox et al., [10] had offered the proposal for a parallel machine learning framework which would detect the AD via the T1-weighted MRI scans that were localized to the hippocampus, and segmented between the right as well as the left hippocampi. Initially, the feature extraction was carried out by two individually trained, unsupervised learning-based AutoEncoders, in which the right as well as the left hippocampi were fed into their corresponding AutoEncoders. Later, a pair of classifiers had performed the classification on the encoded data from the AutoEncoders, to which each classifier pair were gathered together by means of a soft voting ensemble procedure. Results of80%/81% were the accomplished sensitivity/specificity, respectively while 80% was the balanced accuracy score.

Urooj et al., [11] had put forward a novel hybrid approach for the early AD detection with utilization of the Polar Harmonic Transforms (PHT). The feature selection procedure would eliminate the irrelevant features through assessment of the in-class as well as the among-class variances. Of late, the Wavelet Neural Networks (WNNs) had garnered much interest in tasks of classification despite being encumbered with parameter setting. The authors had submitted the proposal for a WNN with the technique of self-adaptation so as to control the following Differential Evolution (DE) parameters. The results of experimentations on the ADNI database had indicated that the proposed approachhad yielded the best overall classification results between AD and MCI.

Accurate classification of the AD as well as its prodromal stage of MCI serves critical roles in the AD diagnosis's computer-assisted intervention. For resolution of feature selectionproblem, Cui et al., [12] had proposed an Adaptive LASSO Logistic Regression model. The proposed algorithm was composed of the following two distinct phases. The first phasehad utilization of the Particle Swarm Optimization (PSO) algorithm for global search so as toeliminate the redundant features as well as to mitigate the subsequent stage's computational time. On the other hand, the second phasehad employed the adaptive LASSO as a local search for picking the optimal features for the classification of AD. The authors had assessed the proposed algorithm's performance using ADNI database and had found that the proposed algorithm had accomplished 96.27%classification accuracy for AD vs. HC, 84.81%classification accuracy for MCI vs. HC, and 76.13% classification accuracy for, cMCI vs. sMCI.

Ganotra et al., [13] had developed an approach for detection of the AD-contributing regions of the brain regions with utilization of the SVM classifiers as well as the newly-devisedSelf Regulating PSO (SRPSO) algorithm. The results had demonstrated that, unlike the classifiers which were constructed with either GM or WM features only, the classifier which was constructed with both GM as well as WM features had achieved a superior accuracy of 89.26%. Furthermore, due consideration of the clinical features as well as the volumetric features had greatly enhanced the accuracy to 94.63%,that was far superior to that of the performance reported by the recent literary works. For identification of the brain regions which were vital for the AD vs. CN classification problem, the authors had employed the SRPSO, which had accomplished better classification performance for the extraction of GM as well as WM features.

The traditional deep clustering models will employ backpropagation to update the model's assigned weights, and also will randomly initialize the involved parameters. For avoidance of the over-fitting problem, Dhanusha et al., [14] had adapted the chaotic chicken swarm optimization,that exhibited the behavior of the chicken farms when searching for their food sources,for fine-tuning the deep clustering's involved parameters. The deep clustering's local optima would be conquered via the application of chaotic theory for the detection of AD. The outcomes whichwere acquired by the proposed chaotic chicken swarm optimization-based deep adaptive clustering had yielded more accurate clustering compared to that of other traditional models.

3 METHODOLOGY

In this section, the feature extraction using curvelet transform, decision tree, Adaboost and XGBoost classifiers and PBIL algorithm are discussed.

3.1 Feature Extraction using Curvelet Transform

In basic terms, curvelet transforms [15]have a tendency to expanditsridgelet transforms for that of amultiple scale analysis as well as an image f (x, y), which has been expressed in Eq. (1) as continuous ridgelet from the coefficients:

$$R_f(a,b,\theta) = \iint \psi_{a,b,\theta}(x,y)\, f(x,y)\, dxdy \qquad (1)$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

For the above equation, anwill be the scale parameter such that a>0, b$\in$R will be the translation parameter with $\theta \in [0, 2\pi]$ as the orientation parameter. Upon completion of the exact reconstruction, the definition of a ridgeletwill be given inEq. (2) as: possible from all of these coefficients. Furthermore, a ridgelet's definition can be as the belowEq. (2):

$$\psi_{a,b,\theta}(x, y) = a^{\frac{1}{2}}\psi \leq \left(\frac{x\cos\theta + y\sin\theta - b}{a} \right) \quad (2)$$

Here, θ will indicatearidgelet'sorientation as aconstant line, and thetransverse to that of such ridges that are wavelets. The high anisotropyis acquired from these ridgeletswhich will capture better edges over a traditional sinusoidal wavelet. A curvelet transform which is based on the wrapping of these Fourier samples will take a 2-Dimage as its input in a Cartesian array f [m, n] such that $0 \leq m < M, 0 \leq n < N,$ and will acquire the curvelet coefficient which has been indexed by a scale j having an orientation l together with two distinct parameters of such spatial location(k_1, k_2). For formation of these curvelet texture descriptors,its is necessary to apply statistical operations and Eq. (3) will define thesediscrete coefficients of the curvelets as follows:

$$C^D(j, l, k_1, k_2) = \sum_{0 \leq m \leq M} f[m, n]\phi^D_{j,l,k_1,k_2}[m, n] \quad (3)$$

Wherein, every $\phi^D_{j,l,k_1,k_2}[m, n]$ will indicate the digital curveletwaveformswhich will deploy scaling law of effective parabolicon their sub-bands in this particular frequency. Moreover, the curvelets will begin to demonstrate an oscillating behavior edges. Fundamentally, these wrapping-based transforms are in case of a multi-scaletransformwhich will employ the pyramid structure with various orientations on both the scales. Its product will be the inverse Fourier which has been transformed so as to acquire the curvelet coefficients.

3.2 Decision Tree Classifier

A decision tree is a classifier which is represented as the instance space's recursive partition. The decision tree will constitute nodes that will form a rooted tree, i.e., it will be a directed tree having a node known as the "root"without any incoming edges. While a node with outgoing edges is referred to as an interna,all the other nodes are referred to as leaves. Within a decision tree, every internal node will partition the instance space into two or more sub-spaces in accordance with a specific discrete function of the input attributes values. Every leaf will be allocated to a single class that will indicate the most appropriate target value. Classification of the instances is done via their navigation from the tree's root all the way down to a leaf as per the tests'outcome along the path [16].

In general, the decision tree is employed for information acquisition so as to make decisions. The decision tree will commence with a root node upon which the users are responsible for taking all the actions. From this root node, the users will partition every node in a recursive manner as per the decision tree learning algorithm. The final result will be a decision tree wherein every branch will be indicative of a probable scenario of decision as well as its outcome [17].

The below points will demonstrate the Decision Tree Algorithm's suitability:

1. Representation of the instance as attribute-value pairs.

2. The target function will have discrete output values. It is easily able to handle an instance that is allocated to a Boolean decision like 'true' and 'false'.

3. The training data may consist of errors which ismanagedwith pruning methods.

Decision Tree learning is an attractive approach for Inductive learning due to the following causes:

1. Decision tree is an appropriate generalization for the unobservedinstance only if the description of the instances isdone with regards to features which are correlated with the target class.

2. The technique are computationally economical in proportion to the number of observed training instances.

3. The resultant decision tree is able to offer a representation of the concept which is most appealing to the human beings as it will render the classification procedure to be self-evident.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

3.3 Population Based Incremental Learning (PBIL) Algorithm

Baluja had introduced the PBIL as one of thesimplest Estimation of Distribution Algorithms (EDAs). The PBIL will employ a probability vector for representing the population, in which every element will indicate the probability of acquiring a value of 1 in a specificgene position. The Hebbianlearning rule is employed for updating the probability vector towards the selected best solutions [21].

In the PBIL, representation of the individual population is given by a probability vector in Eq. (11) as below:

$$p_1(x) = (p_1(x_1), p_1(x_2), ..., p_1(x_n)) \quad (11)$$

Here, $p_1(x_i)$ will indicate the probability of acquiring avalueof 1 in the l^{th} generation'sith gene position.

At the algorithm's commencement, there will be initializationofthe probabilityvector $p_0(x)$ by setting all the elements to a value of 0.5,that is, the generation of random solutions upon sampling from this probability vector will occur as there is equal probability for generating either 0 or 1. With the search's progression, theprobability vector's values will gradually lean towards those values which represent the high evaluation solutions.

The PBIL's evolution procedure will be: At every generation, M solutions will get formed based on the probabilities that have been specific in the currentprobability vector, $p_1(x)$. Afterwards, these solutions will undergo evaluation, and there will be the selection of N best solutions (N≤M) which will be denoted as $x^l_{1:M} ....., x^l_{i:M} ....., x^l_{N:M}$. With these chosen best solutions, an update of theprobability vector is done by means of a Hebbian-inspired rule in the below Eq. (12):

$$p_{l+1}(x) = (1-a)\, p_l(x) + a \frac{1}{N} \sum_{K=1}^{N} x^l_{k:M} \quad (12)$$

Here, a ∈ (0,1] will indicate the learning rate.

Upon updating the probability vector, there will be generation of a new solution set through sampling from the newprobability vector $p_{l+1}(x)$, and there will be repetition of the cycle till fulfilment of a specific condition of termination, for instance, convergence of theprobability vector to either 0.0 or 1.0 for every bit position.

The pseudo-code of PBIL Algorithm:

Initialize probability vector $p_0(x)$
While no convergence do
Begin
$U \sin g\ p_l(x)$ *obtain* M *solutions*: $x^l_1, ..., x^l_M$
Evaluate and rank $x^l_1, ..., x^l_M$
Select the $N(N \le M)$ *best soultions*: $x^l_{1:M}, ..., x^l_{N:M}$
Update the probaility vector $p_{l+1}(x) = (p_{l+1}(x_1) ..... p_{l+1}(x_n))$:
For $i = 1....n$ *do*
$p_{l+1}(x_i) = (1-a)\, p_l(x) + a \frac{1}{N} \sum_{K=1}^{N} x^l_{i,k:M}$
End

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

3.4 Proposed PBIL Adaptive Boosting (Adaboost) Classifier

The AdaBoost algorithm will combine diverse classifiers so as to offer a boost in the accuracy. This algorithm will operate in an iterative manner, and also will aggregate numerous weak classifers to offer a strong as well as efficient classifier. The weight will be set across every classifier, and for every iteration of the training samples, accurate predictions will be executed for unusual classifier behavior.The AdaBoost classifier has teh ability to mitigate the training error during the training procedure in every iteration [18].

Consider a training set $\{x_1, y_1\}, \{x_2, y_2\}.....\{x_i, y_i\}, ..., \{x_n, y_n\}$, in which every $x_i$will be a sample of space X while every $y_i$willalabelwhich corresponds to the output result set Y of x_i. Y{- 1, +1}. The AdaBoost will call a given weak learningalgorithm in a series of loops t = 1, 2… T. Below are the algorithm'sspecificsteps [19]:

(1) For n learning samples, the Eq. (4) will be:

$$\{x_1, y_1\}, \{x_2, y_2\}.....\{x_i, y_i\}, ..., \{x_n, y_n\} \quad (4)$$

$x_i \in X, y_i \in Y = \{-1, +1\}$. Let n of the n learningsamplesconsist of q negative samples as well as p positivesamples.

(2) For y_i = 1, - 1, the initialization weights will be $D_{1,i} = 1/2p \; and \; 1/2q$, respectively.

Output the final assumption as the below Eq. (5):

$$H(x) = \operatorname{sgn}\left(\sum_{t=1}^{T} \alpha_t h_t(x) \right) = \begin{cases} 1, \; \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{n=1}^{T} \alpha_t \\ -1 \end{cases} \quad (5)$$

Definition of the aforementioned algorithm's variables will be as follows: Let $h_1$indicate a weak hypothesis, H indicate the final hypothesis, vote for the weight at of T weakhypothesesh$_t$; α_t indicate the weight of h_t; ε_t indicate the training errorofh$_t$; and $D_t$indicatethe probability distribution of h_t. The weak learner's job is to seek a suitable appropriate hypothesis, $h_t : X \to \{-1, +1\}$, for the probability distribution, D_t. Eq. (6) will gauge the weak hypotheses'fitness by means of their own error:

$$\varepsilon_t = P_{r_i \sim D_i}[h_t(x_t) \neq y_i] \quad (6)$$

The error, ε_t ,will be related to the distribution, D_t. For practicalapplications, the weak learner's evaluationisdone from the weightD_t on the training sample. If the weak learning hypothesish$_t$ is true, the AdaBoost will pick a parameter β_t, that isrelated to α_t while α_t will indicate the weight of h_t. When $\varepsilon_t < 1/2, \frac{\varepsilon_t}{1-\varepsilon_t} < 1, \beta_t < 1$, the weight of the correct classification's sample will become smaller, that will imply that there will be an increase in weight of the misclassification'ssample,andassumptions ofasmaller ε_t, a smaller β_t as well as abigger α_t, namely, a bigger h_t weight will be taken into account.

In the proposed PBIL AdaBoost, the hyperparameters of the Adaboost is optimized. In this work, the number of trees used in the ensemble and the learning rate of the Adaboost is optimized. The initial population is encoded with number of trees and learning rate, on iteration, the optimal values are obtained.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

3.5 Proposed PBIL Extreme Gradient Boosting (XGBoost) Classifier

XGBoost [20],thegradient boosting-based machine learning approach will employ an ensemble model through summation of the prediction values from multiple decision trees that grow during the iterations. Suppose that $\{(f_m, y_m), m = 1,...,n\}$ is the datasetwhich is offered for the training, in which $f_m$will represent the m^{th} brain scan's feature vector, and $y_m$will represent the label. Suppose that $\hat{y}_m^{(j)}$ is the predicted label at the j^{th}iteration.Then, there will be addition of a new tree structure $q_j$for minimization of the below Eq. (7),i.e., objective function $L^{(j)}$:

$$L^{(j)} = \sum_{m=1}^{n} l\left(y_m, \hat{y}_m^{(j-1)} + q_j(f_m) \right) + \Omega(q_j) \tag{7}$$

Here, *l* will indicate a loss function that measures the difference between the predicted label and the true label, $\Omega = \gamma J + 1/2\lambda \| W \|^2$,will be the regularization term which will penalize the model's complexity, wherein**w**will indicate the leaf weights in the tree,*T*, while *J* will indicate the number of leaves in the tree,*T*. With application of theXGBoost, acquisition of the feature importance is done by means of a score which will indicate how valuable each feature is upon construction of an enhanced decision tree. The feature's relative importance will become higher with more utilization of a feature for the decision. Eq. (8) will express the gain which is attained after the tree split as follows:

$$G = \frac{g_L^2}{h_L + \lambda} + \frac{g_R^2}{h_R + \lambda} - \frac{g^2}{h + \lambda} \tag{8}$$

For the above equation, *g* will indicate the first order gradient on the loss function *l* (.) with respect to the predicted label while *h* will indicate the second order gradient on the loss function *l* (.) with respect to the predicted label, the subscript *L* will indicate the left nodes after the tree split, and the subscript *R* will indicate the right nodes after the tree split. For a single decision tree, the below Eq. (9) will evaluate the relative influence of the k^{th} feature in the tree *T* as:

$$U_k(T) = \sum_{i=1}^{J-1} G_i 1(v_i = k) \tag{9}$$

Wherein, the summation will be over all the non-terminal nodes *i*, $i = 1, \ldots, J-1$, in the tree *T*,$v_i$will indicate the feature which is employed for splitting in association with the node *i*, and $G_i$will indicate the corresponding gain after the split. Afterwards, the feature importance will get averaged across all the decision trees T_p,$p = 1, \ldots, M$, within the model in the following Eq. (10):

$$U_k = \frac{1}{M} \sum_{p=1}^{M} U_k(T_p) \tag{10}$$

Every feature inf will be ranked in accordance with its relative importance as denoted by Eq. (10), followed by a threshold to pick a specific number of features for the consequent classification. It must be noted that the training data will form the basis for selection as well as recording of the important features. The testing will involve the direct utilization of the feature indices which have been acquired in the training procedure. In the end, the final AD classification is done by feeding the chosen features to a two-layer neural network.

In the proposed PBIL XGBoost, the hyperparameters of the XGBoost is optimized. In this work, the number of trees, depth of the tree and the learning rate of the XGBoost is optimized. The initial population is encoded with number of trees, depth of tree and learning rate, on iteration, the optimal values are obtained.

4 RESULTS AND DISCUSSION

In this section, the 1100 images of AD /MCI / CUis used. The decision tree, Adaboost, XGBoost, PBIL Adaboost and PBIL XGBoost methods are discussed. Table 1 shows the summary of results. The classification accuracy, sensitivity and specificity as shown in figures 2 to 4.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

Table 1 Summary of Results

Techniques	Decision Tree	AdaBoost	XGBoost	PBIL Adaboost	PBIL XGBoost
Classification accuracy	89.27	91.09	91.73	93.73	94.73
AD- Sensitivity	0.8208	0.8292	0.8708	0.9	0.9167
MCI - Sensitivity	0.8774	0.9129	0.8903	0.9258	0.9355
CU-Sensitivity	0.9327	0.9455	0.9527	0.96	0.9673
AD- Specificity	0.9504	0.964	0.9615	0.9702	0.9739
MCI -Specificity	0.9467	0.9561	0.9619	0.9688	0.9792
CU- Specificity	0.9269	0.9323	0.9417	0.9618	0.9623

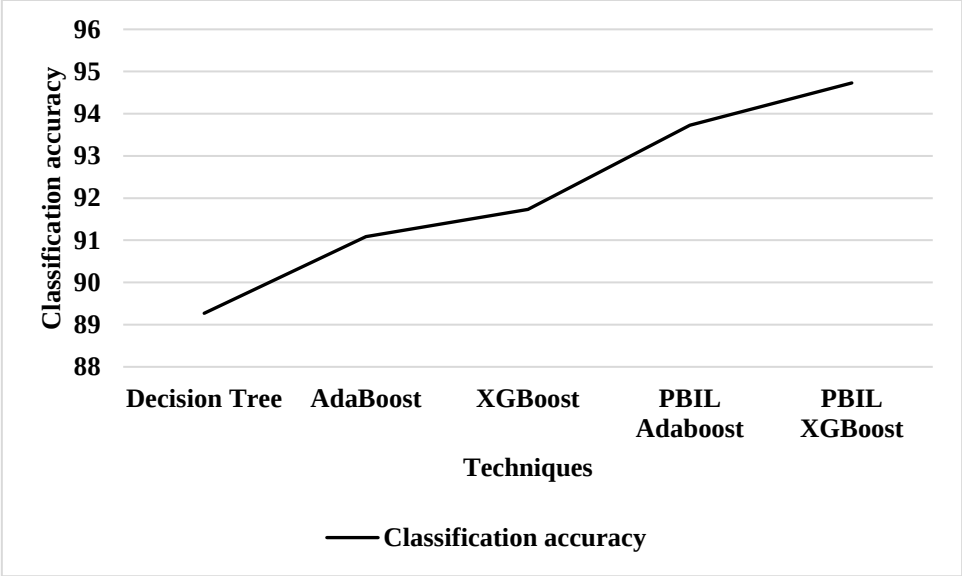

Figure 2 Classification Accuracy for PBIL XGBoost

From the figure 2, it can be observed that the PBIL XGBoost has higher classification accuracy by 5.93% for decision tree, by 3.92% for Adaboost, by 3.22% for XGBoost and by 1.06% for PBIL Adaboost respectively.

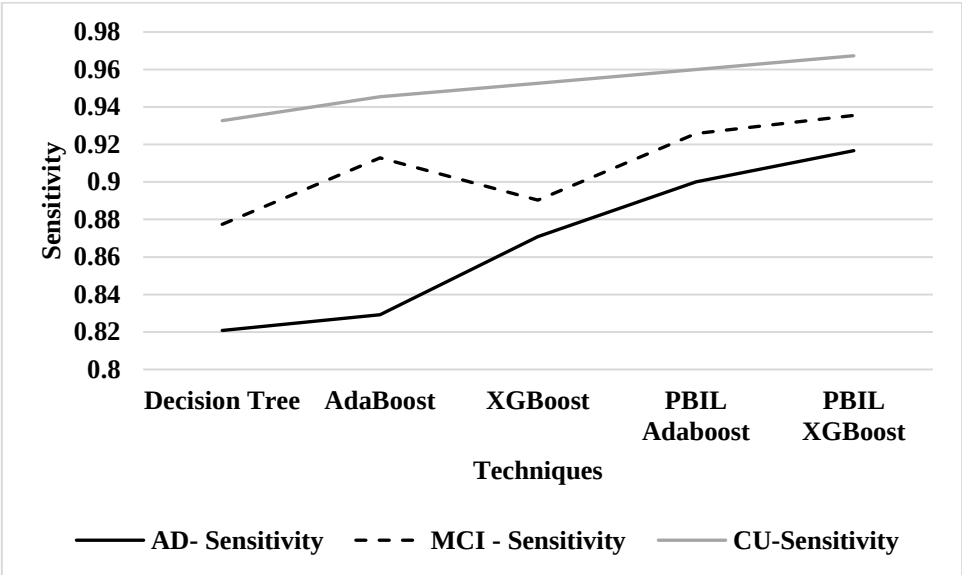

Figure 3 Sensitivity for PBIL XGBoost

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 4137-4146
https://publishoa.com
ISSN: 1309-3452

From the figure 3, it can be observed that the PBIL XGBoost has higher sensitivity for AD by 11.04% for decision tree, by 10.02% for Adaboost, by 5.13% for XGBoost and by 1.84% for PBIL Adaboost respectively. The PBIL XGBoost has higher sensitivity for MCI by 6.41% for decision tree, by 2.44% for Adaboost, by 4.95% for XGBoost and by 1.04% for PBIL Adaboost respectively. The PBIL XGBoost has higher sensitivity for CU by 3.64% for decision tree, by 2.28% for Adaboost, by 1.52% for XGBoost and by 0.75% for PBIL Adaboost respectively.

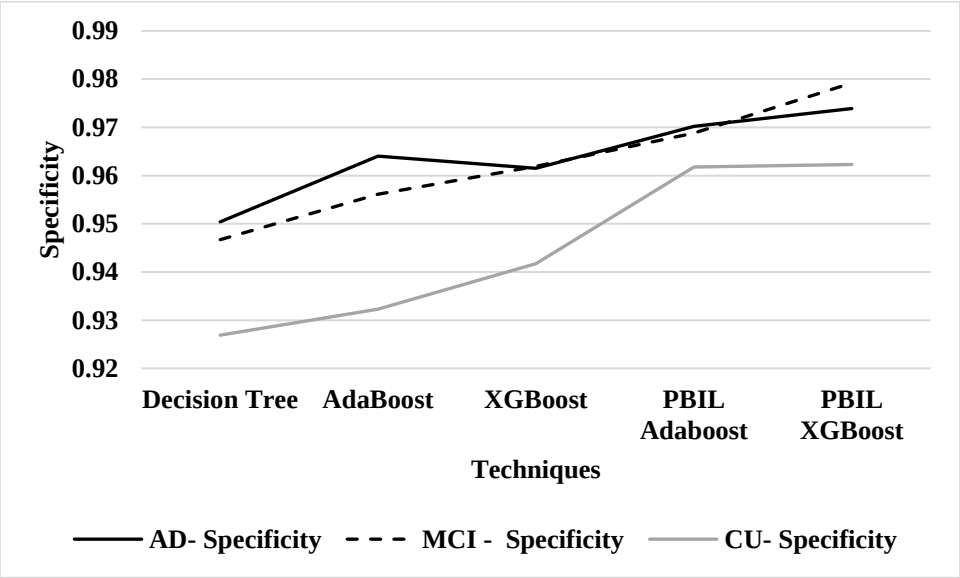

Figure 4 Specificity for PBIL XGBoost

From the figure 4, it can be observed that the PBIL XGBoost has higher specificity for AD by 0.54% for decision tree, by 1.02% for Adaboost, by 1.28% for XGBoost and by 0.38% for PBIL Adaboost respectively. The PBIL XGBoost has higher specificity for MCI by 3.37% for decision tree, by 2.38% for Adaboost, by 1.78% for XGBoost and by 1.06% for PBIL Adaboost respectively. The PBIL XGBoost has higher specificity for CU by 3.74% for decision tree, by 3.17% for Adaboost, by 2.16% for XGBoost and by 0.05% for PBIL Adaboost respectively.

5 CONCLUSION

It is becoming increasingly evident that the MRI is a critical technique for the AD diagnosis as well as for the prediction this neurodegenerative disorder's onset. This work had presented a high-accuracy sophisticated machine learning model for early-stage AD diagnosis. Feature extraction with the curvelet transform is a multi-scale as well as multi-directional transform which has needle-shaped basis functions. AdaBoost is capable of assembling classifiers, and at the time of the training procedure, there will be an increase in the incorrectly-classified weights of the samples while there will be a decrease in the correctly-classified weights of the samples. XGBoost is an optimized distributed gradient boosting library which has been designed to be portable, flexible as well as highly efficient. The PBIL algorithm is a simple stochastic optimization technique which can be rapidly employed on a broad range of problems. The PBIL's chief area of application is for problems which are either too multimodal or discontinuous for gradient or simplex approaches, but also do not necessitate a full evolutionary algorithm solution. Results show that the PBIL XGBoost has higher classification accuracy by 5.93% for decision tree, by 3.92% for Adaboost, by 3.22% for XGBoost and by 1.06% for PBIL Adaboost respectively.

REFERENCES

1. Thapa, S., Singh, P., Jain, D. K., Bharill, N., Gupta, A., & Prasad, M. (2020, July). Data-driven approach based on feature selection technique for early diagnosis of Alzheimer's disease. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
2. Kruthika, K. R., Maheshappa, H. D., & Alzheimer's Disease Neuroimaging Initiative. (2019). Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. Informatics in Medicine Unlocked, 14, 34-42.
3. Acharya, U. R., Fernandes, S. L., WeiKoh, J. E., Ciaccio, E. J., Fabell, M. K. M., Tanik, U. J., ... & Yeong, C. H. (2019). Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques. Journal of Medical Systems, 43(9), 1-14.
4. Mirzaei, G., Adeli, A., &Adeli, H. (2016). Imaging and machine learning techniques for diagnosis of Alzheimer's disease. Reviews in the Neurosciences, 27(8), 857-870.

5. Kaka, J. R., & Satya Prasad, K. (2022). Differential Evolution and Multiclass Support Vector Machine for Alzheimer's Classification. Security and Communication Networks, 2022.
6. Battineni, G., Hossain, M. A., Chintalapudi, N., Traini, E., Dhulipalla, V. R., Ramasamy, M., &Amenta, F. (2021). Improved Alzheimer's Disease Detection by MRI Using Multimodal Machine Learning Algorithms. Diagnostics, 11(11), 2103.
7. Garg, N., & Chaudhary, M. (2021, November). Implementation of Dual Tree Complex Wavelet Transform with Mean Energy Features to detect Alzheimer's Disease. In 2021 7th International Conference on Signal Processing and Communication (ICSC) (pp. 188-193). IEEE.
8. Kaplan, E., Dogan, S., Tuncer, T., Baygin, M., &Altunisik, E. (2021). Feed-forward LPQNet based Automatic Alzheimer's Disease Detection Model. Computers in Biology and Medicine, 137, 104828.
9. Karadayi-Atas, P., Sevkli, A. Z., & Tufan, K. (2021). A VNS based framework for early diagnosis of the Alzheimer's disease converted from mild cognitive impairment. Optimization Letters, 1-20.
10. Knox, S. A., Chen, T., Su, P., & Antoniou, G. (2021, September). A Parallel Machine Learning Framework for Detecting Alzheimer's Disease. In International Conference on Brain Informatics (pp. 423-432). Springer, Cham.
11. Urooj, S., Singh, S. P., Malibari, A., Alrowais, F., &Kalathil, S. (2021). Early Detection of Alzheimer's Disease Using Polar Harmonic Transforms and Optimized Wavelet Neural Network. Applied Sciences, 11(4), 1574.
12. Cui, X., Xiao, R., Liu, X., Qiao, H., Zheng, X., Zhang, Y., & Du, J. (2021). Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis. Chemometrics and Intelligent Laboratory Systems, 215, 104316.
13. Ganotra, R., Dora, S., & Gupta, S. (2021). Identifying brain regions contributing to Alzheimer's disease using self regulating particle swarm optimization. International Journal of Imaging Systems and Technology, 31(1), 106-117.
14. Dhanusha, C., Kumar, A. V., Musirin, I. B., & Abdullah, H. M. A. (2022). Chaotic Chicken Swarm Optimization-Based Deep Adaptive Clustering for Alzheimer Disease Detection. In Pervasive Computing and Social Networking (pp. 709-719). Springer, Singapore.
15. Saravanakumar, S., &Thangaraj, P. (2019). A voxel based morphometry approach for identifying Alzheimer from MRI images. Cluster Computing, 22(6), 14081-14089.
16. Bari Antor, M., Jamil, A. H. M., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., ... &Masud, M. (2021). A comparative analysis of machine learning algorithms to predict alzheimer's disease. Journal of Healthcare Engineering, 2021.
17. Singh, S., & Gupta, P. (2014). Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. International Journal of Advanced Information Science and Technology (IJAIST), 27(27), 97-103.
18. Karthiga, M., Sountharrajan, S., Nandhini, S. S., & Kumar, B. S. (2020, May). Machine Learning Based Diagnosis of Alzheimer's Disease. In International Conference on Image Processing and Capsule Networks (pp. 607-619). Springer, Cham.
19. Fan, Z., Xu, F., Li, C., & Yao, L. (2020). Application of KPCA and AdaBoost algorithm in classification of functional magnetic resonance imaging of Alzheimer's disease. Neural Computing and Applications, 1-10.
20. Ge, C., Qu, Q., Gu, I. Y. H., &Jakola, A. S. (2019). Multi-stream multi-scale deep convolutional networks for Alzheimer's disease detection using MR images. Neurocomputing, 350, 60-69.
21. Zhang, Q., Wu, T., & Liu, B. (2007, August). A population-based incremental learning algorithm with elitist strategy. In Third International Conference on Natural Computation (ICNC 2007) (Vol. 3, pp. 583-587). IEEE.