A Novel Feature Selection based Classification Model for Disease Severity Prediction on Alzheimer's Database

I. Murali Krishna, Research Scholar, JNTU, Kakinada, Andhra Pradesh, India. E-mail:imuralijntuk@gmail.com

Dr. Challa Narsimham, Principal, Vignan's Institute of Information Technology, Vishakhapatnam, Andhra Pradesh, India.

E-mail:Narasimham_c@yahoo.com

Dr.A.S.N. Chakravarthy, Professor, Department of Computer Science and Engineering, University College of Engineering, JNTU, Kakinada, Andhra Pradesh, India. E-mail:drasnchakravarthy@yahoo.com

Abstract--- In the current era, research on automatic image classification on high dimensional medical disease databases is growing rapidly. Since most of the Alzheimer's disease databases have heterogeneous features with different levels of severity patterns. Detection and classification of high risk patterns has many potential benefits for decision making. Traditional image classification models such as Naïve Bayesian, Neural Networks, SVM, Regression models,. etc are used to classify the image using the annotated ROI and image texture features. As the size of the Alzheimer's disease patterns and its categories are increasing, traditional data classification models are failed to process the disease patterns due to class imbalance, inconsistent, and sparsity issues, which may affect the disease prediction rate and error rate. Unlike the existing solutions, which require a prior knowledge of classification parameters for various types of image features, which is not possible to obtain in practice. Also, as the size of the training images increases, it becomes difficult to find the relevant features using the image features and ROI values. In this proposed model, a novel filtered based automatic Alzheimer's disease classification model is proposed to improve the disease prediction rate and to minimize the error rate of the classification model. Experimental results show that the proposed model has high prediction rate compared to the traditional models in terms of true positive rate and error rate are concerned.

I. Introduction

Abnormal behavior and loss of memory could indicate a brain disorder that is neurodegenerative, known as Alzheimer's disease. In other words, Older people suffer from some form of dementia which could lead to neuronal loss known as Alzheimer's disease (AD). Neuropsychological examination and psychometric assessment mainly determine the clinical evaluation. The confusing factor of cognitive research may, however, hide the early signs of Alzheimer's disease. Successful diagnosis of the disease can be achieved through using structural MRI to detect the brain's anatomical changes that are brought about by the disease. The advantage of using automation for AD detection lies in the improved accuracy and the increased speed of the process of treatment. Structural neuroimaging can provide good markers for the detection of diseases such as AD because of their sensitivity to degeneration. AD detectors fall into three typical types. The types which base on cortical thickness are the most used ones. Querbes and colleagues produced an 85% accuracy of distinction of normal aging healthy controls from patients with Alzheimer's disease. The second class of popular structural measurements is those who consider the volumes of certain structures, for example, the hippocampus. An example of a technique which reduces dimensions efficiently is Principal component analysis. It removes the repetitive elements from the data and does not alter a large part of the useful information, hence resulting in data compression. The technique converts the original feature data to an ordered, and uncorrelated variable set called the Principal Components (PCs) so that the first few PCs contain most of the original variables' variations. PCA can greatly reduce the brain image data's dimension. Support vector machine has, however, produced better results in the categorization of clinical diseases. It is a supervised learning model that allows high dimensional data to be trained and classified [1][2]. Some studies prove the SVM's high accuracy of AD classification using high-dimensional data set.

Machine learning models are used to classify different medical datasets such as microarray data, clinical data, and proteomic data as input. Most of the traditional approaches consider the features as independent and linear. Most of the biological systems are non-linear and its parameters are interdependent, thus machine learning has become better choice. Both machine learning and conventional methods suffer from high dimensionality problem. This problem can be resolved by either decreasing the number of variables or increasing the number of training

Received: 5 Feb 2018/Accepted: 15 Mar 2018

datasets[3]. Machine learning algorithms can be categorized into three broad types, they are:- Supervised machine learning, Unsupervised machine learning and Reinforcement machine learning. Supervised learning consists of a prescient provider which provides the labeled training dataset as input to the algorithm and produces output after mapping. But on the contrary, for unsupervised machine learning only training datasets are given as input without labels. Some of the examples of unsupervised learning are:- Self Organizing feature Maps, hierarchical clustering, k-means clustering, and so on. Some of the traditional methods basically implement the baseline image features for prediction of significant baseline scores. Other kinds of approaches require both baseline image features and scores for prediction of accurate future scores. The partial set of scores generated out of missing data can be properly utilized in this proposed technique [6].

Generally, the SVM classification scheme is based upon the characteristics of statistical learning mechanism. The classification of SVM classifier supports structural risk minimization to carry out the whole process of classification smoothly and effectively. Additionally, the SVM technique is quite efficient in case of small training data. It has many other applications such as, disease assessment, detection of exudates in digital data, Alzheimer's disease prediction, glioma recognition, and so on. Initially, two different statistical evaluation processes are carried out in order to verify whether variances are similar across healthy and diseased persons[4]. A MIRIAD database is the source of the Alzheimer and the normal images used in this work. The edge detection's threshold values, both local and global, are calculated using Tsallis entropy and Shannon entropy methods. Morphological operations are further applied in the processing of the acquired edges map. The edge map generates a map which helps in extraction of the brain tissues. An improvement to traditional linear SVM model is regression based SVM (RSVM) which is used to obtain complex relationship among disease features and the future clinical scores. RSVM is categorized under a special type of ensemble learning scheme which is mostly implemented during the process of brain tissue classification. The non-linear decision boundary is built on the complex patterns (for e.g., relationship among brain disease features and clinical scores).Additionally for the evaluation process of dysphonia patterns, two-sample Kolmogorov–Smirnov (K–S) test is used in order to distinguish patient from healthy person. Apart from these, primary statistical differences in characteristics of dysphonia measurements of both patients and healthy individuals are evaluated. Three set of genes are implicated in the pathophysiology of early onset AD (EOAD). But, each and every ADG can't be detected due to the complex nature of Alzheimer's disease.

II. Related Works

Traditional dimension reduction techniques like PCA and feature selection models are appropriate for datasets of limited size and high dimensionality. In this paper, they integrated traditional classifier with fuzzy k-nearest neighbor and feature selection through random projection technique in order to carry out classification of big datasets. In PCA approach, the projection matrix is evaluated only once through a least square optimality. But in random projections, the projection matrix is evaluated more than once. Generally accuracy of the ensemble disease prediction approaches depends on base classifiers and data dimensionality[5]. In this work, an advanced classifier ensemble framework is developed along with a group of predefined base classifiers. The main objective of this approach is to classify the test disease patterns based on the majority voting of multiple base classifiers. Here, a large amount of high dimensionality data are required to process real time classification applications such as medical disease prediction and disease prediction applications.

Traditional approaches such as LS-SVM, PNN and GRNN are implemented in order to discriminate the voice signals of disease affected persons from normal healthy persons[7-9]. The disease progression can be predicted more accurately with the help of clinical scores. Most of the traditional classification models initiate a linear function among features and the clinical scores. Hence, most of the traditional models such as linear SVM, RVM, Lasso regression, etc. are linear in nature for Alzheimer's or Parkinson disease prediction. They proved that at least three years before the symptoms of Alzheimer's disease could be noticed, the patients of Alzheimer's disease showed a divergence in the rates of whole brain atrophy as compared to normal subjects. The fast progression of the disease is noted by the whole brain's atrophy. The brain structures and their atrophy can be studied using an MRI. Different MRI subjects and anatomical regions provide information of different intensity. The main drawback in examining the progression of neurodegenerative diseases is that the study of volumes using MRI is complicated and broad.

Simpson, et.al, developed an advanced hybrid intelligent system for appropriate identification of Alzheimer's disease [8]. Feature pre-processing and feature reduction are considered as compulsory phase of each and every pattern recognition approach. The selected interesting features out of dataset have the responsibility to decrease the complexity of learning phase. It also enhances the generalization capability of classifiers. Spectral decompositions, local filters, tissue classifications, curvatures, gradients, and image intensity are some of the popularly used features

Received: 5 Feb 2018/Accepted: 15 Mar 2018

extraction method in disease prediction. There are several statistical classification algorithms, and one of them is AdaBoost. It identifies the classifiers which may be prone to error (weak classifiers) from a strong classifiers and uses their error rates to assign weights to them. For each instance in AdaBoost, a weight is assigned to predict the disease class label in the next iteration. On the other hand, SVMs try to achieve a hyper-surface for all features with the objective of maximizing the margins and minimizing the training errors. In the training data, the difference between the closest value in feature space and the hyper-surface defines the margin. The ability of AdaBoost is to automatically select the features to make the Ada-SVM and AdaBoost better at classification using a hierarchical decision tree framework [12].Extreme Learning Machine can be defined as a single-hidden layer feed-forward neural network (SLFN) with learning model . The traditional optimization approaches like gradient descent based back-propagation evaluate weights and biases. This technique is responsible for decreasing the training time effectively through random assignment of weights and biases. The extended ELM method results better efficiency and performance as compared to all traditional approaches. ELM has wide range of applications in different domains like face recognition, human action recognition, landmark recognition and protein sequence classification, medical disease prediction [1, 3,4]. But, ELM has two major issues, those are:- 1) This model has over fitting problem and the performance can't be predicted for unknown disease categories. 2) This model is not applicable to binary classification and uncertain datasets. They showed a two-dimensional method for atrophy quantification in the early diagnosis of Alzheimer's disease as well as progression tracking of the disease in clinical practices. In the initial step, MR brain is analyzed using the quantification method. The similarity of intensities between non-brain and brain tissues, coupled with their weak boundaries complicates the process of brain extraction. There are many tools for the whole brain extraction such as the Minneapolis consensus Strip, Statistical Parametric Mapping, Brain Surface Extractor and Brain Extraction tool. The stated tools were compared by Boesen tool. There have also been attempts to use semi-automated techniques in which morphological operations are preceded by edge detection.

Brain segmentation is done to divide brain images according to the anatomical structures into distinct homogenous regions. There are numerous neurological disorders such as Parkinson-related syndrome, and Alzheimer disease whose diagnosis requires the segmentation or classification of the Magnetic Resonance Image of the brain image. It, however, poses a challenge to leave the segmentation of the brain image from an MRI scan to be fully automatic, hence the development of supervised and unsupervised classification algorithms to help out in this task. The problem with supervised techniques is that they require a lot of training data, and they take more time. *Escudero, et.al,* introduced a new longitudinal clinical score prediction technique in case of Alzheimer's disease [9]. Machine learning approaches are very much efficient for the prediction of clinical scores. Linearity assumption and missing data exclusion are major disadvantages of traditional approaches. They proposed a non-linear supervised sparse regression based random forest (RF) framework in order to predict numbers of different longitudinal clinical scores. Additionally, a soft-split approach is presented for assignment of probabilistic paths in order to evaluate appropriate prediction. Furthermore, the subjects having missing scores are eliminated completely. These missing scores are generally estimated by soft-split sparse regression based RF.

Pachauri, et.al, proposed a new machine learning based classification technique [10]. This method emphasizes on simple drawing movements in Alzheimer's disease. It is capable of differentiating healthy state from diseased state by simply drawing straight lines. There are certain other approaches which involve writing words, drawing Archimedes spiral, circles. White matter hyper-intensities normally require expert raters to do manual segmentation/classification on Fluid-attenuated Magnetic Resonance images. It is difficult to perform consistent and accurate segmentation of white matter hyper intensities for a couple of reasons. Their patterns and texture are heterogeneous, and the borders between the intensities are not clear. The main problem to determine the border between the non-WMH and WMH tissue make it better to use the intra-rater and inter-rater agreements. Detection of WMHs uses various MRI contrasts such as FLAIR which shows the hyper-intensity of WMHs, proton density, T2-w, and T1-w which is mostly useful for co-registration.

Morra, et.al, proposed a new gene identification approach for prediction and detection of Alzheimer's and Parkinson diseases [11]. Complete research on AD-related genes (ADGs) is not completed till now. The National Center for Biotechnology Information provides an AD dataset of 22283 genes. In between the above huge numbers of genes, 71 genes are known as AD-related genes. But there also exist some other genes among those 22212 genes those are not included under the category of AD-related genes till date. The major objective of this research work is to detect all those extra AD-related genes through implementing an efficient machine learning approach. In order to enhance the overall accuracy of AD-related genes, a new gene detection approach is introduced which uses multiple classifier integration technique. Initially, a feature selection technique is implemented in order to choose every individual relevant attributes. In the paper [10] anatomical structure and pre-defined disease patterns are used to detect the Alzheimer's disease effectively. Such anatomical structures include the entorhinal cortex and the

Received: 5 Feb 2018/Accepted: 15 Mar 2018

hippocampus. They implemented a novel scoring method known as Non-local Image Patch Estimator (SNIPE), which uses a patch-based non-local framework and measures the similarity between the training population's patches and the patch surrounding the voxel. This method efficiently captures the changes in anatomy by segmenting and grading the structures simultaneously.

The training set in this case comprised of subjects randomly chosen from an ADNI database 50 AD patients and 50 cognitively normal subjects. The non-local means filter depends on two aspects: patch-based comparison and the image's self-similarity.

First, an estimator's accuracy can be increased by increasing the number of samples to reduce the training error. The non-local means filter uses the similarity of the voxels over the whole image. Second, a voxel's surrounding neighbor can be used to detect the next surrounding neighbor, and hence confirm that the samples belong to a common population. *A. Ozcift and A. Gulten* developed a classifier ensemble construction with method with the help of rotation forest [16].

It plays an important role to enhance medical diagnosis performance of machine learning approaches. The machine learning applications require different classifiers along with improved accuracy levels. This model can be divided into two important and significant phases:- 1) In the initial phase, a relevant feature Selection technique is implemented in order to detect the most interesting and useful features. 2) In the subsequent phase, a high accuracy classifier is required to get the highest classification performance. Classifier ensemble techniques are very vital in order to enhance the classifier performances.

Li *et.al [12]*combined the concepts of machine learning with neural networks in order to predict Alzheimer's and Parkinson's disease through the most affecting risks . They included attribute evaluation strategy for their said work. They noticed that, classification accuracy was greatly affected by the interdependent risks. They considered classification accuracies, genes, age etc. as the influencing risks for Alzheimer's disease.

III. Proposed Model

Classification based disease prediction models performs the classification of Alzheimer disease patients which indicates more enlargement in the ventricular section, and loss in brain volume as subjects through semiautomatic segmentation on high-quality, high-resolution MR images.

The cause of brain atrophy is the loss of either gray or white matter or a combination of both. The rate of reduction of white matter is overshadowed by the cortical gray matter's reduction due to the neuronal loss of the Alzheimer disease. Pathologic and imaging evidence have proven that the Alzheimer disease affects white matter, though in a secondary way rather than primarily.

The hyper-intensities of white matter signals which is indicated by the MRIs of the AD patients has been described by several investigators and confirmed by others.

Existing feature selection models can be classified into two groups, one is the feature subset selection and another one is feature rank prediction. In feature rank prediction models, each feature is assessed according to the evaluated ranking measures and then relevant features for a given data set are selected. In the feature subset selection, different statistical measures are evaluated on the training data to select the subset of features and instances for the classification model.

Features are extracted generally by neighborhood operations applied on an image. Features can be of different structures like points, edges, regions etc.

Some feature types can be easy to calculate but generally may not be good enough for future use. Different kinds of features work well for different types of scenarios but features that are mostly used are feature points. Feature points are good because they work well in most of the computer vision techniques.

Feature detection can be computationally expensive and even using those features for any other purpose can be computationally costly.

Edges are curves or lines or set of points that show the intensity transitions in an image. Edges are easy to calculate but are difficult to track if the image content is changed slightly. For features like edges we use detection techniques which are based on gradients, Laplacian zero crossing etc. Image derivatives are generally used to locate edges in an image. Gradient is a first derivative of an image and is used in edge detection. The gradient of an image gives us the direction and magnitude of maximum intensity change at every pixel of an image.

Received: 5 Feb 2018/Accepted: 15 Mar 2018

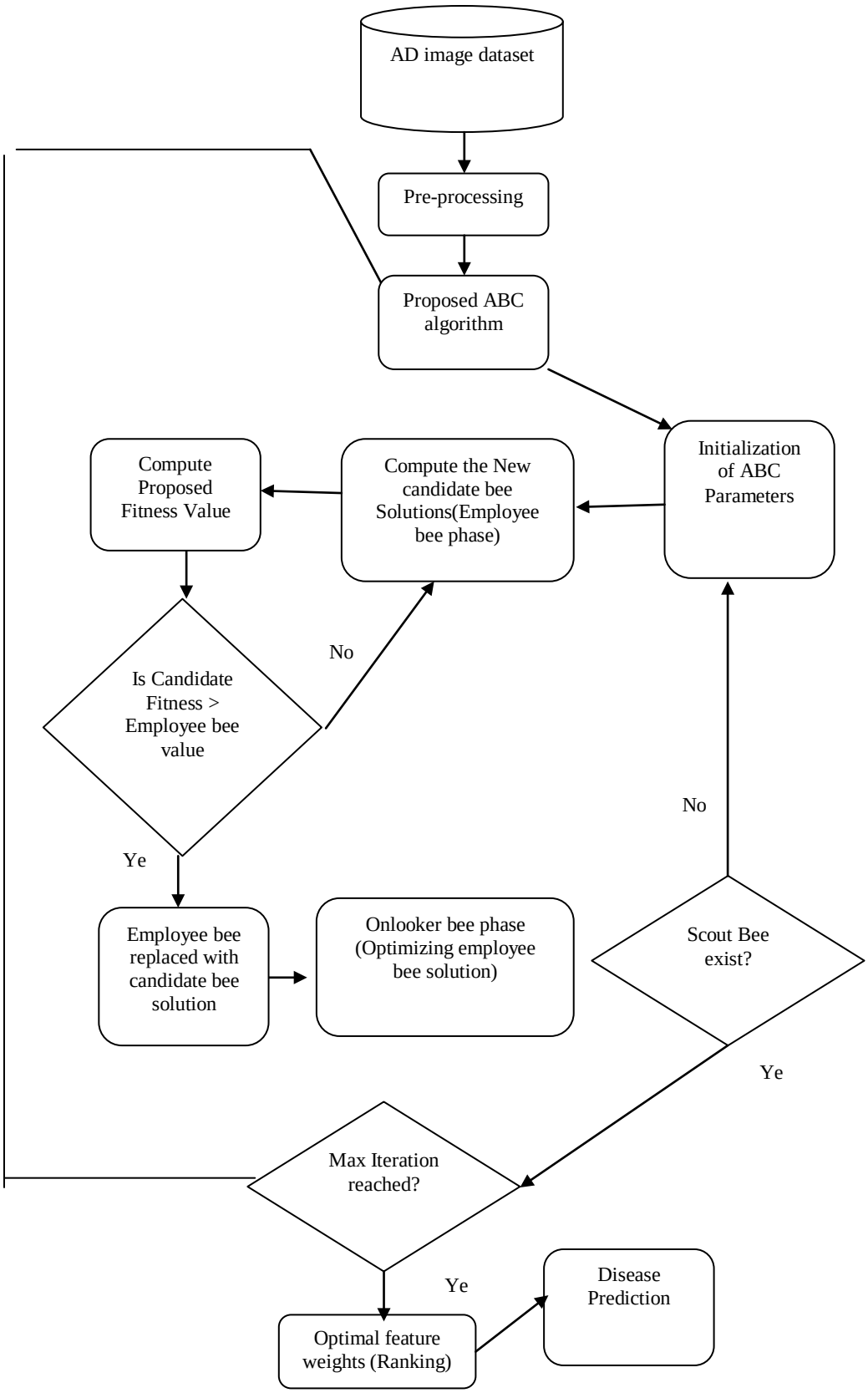


Figure 1: Proposed ABC optimization FS

Received: 5 Feb 2018/Accepted: 15 Mar 2018

Proposed Improved ABC based Ensemble Classification Algorithm

Step 1: Data pre-processing on Training high dimensional data.

Load dataset $AD^1, AD^2 .... AD^n$

For each feature F(i) in the $AD^1, AD^2 .... AD^n$

do

For each record value I(j) in the F(i)

do

if(isNum(I(j)) && I(j)==null)

then

$$I(j) = \max\{|I(i) - \mu_{F(I(j))}|) / (Max_{F(I(j))} - Min_{F(I(j))})\} \quad \text{----(1)}$$

end if

if(isNominal(A_i) && A_i(I)==null)

then

$$I(j) = MaxClass\,Prob(A(i), class(m)); \quad \text{---(2)}$$

Here, mth class of the missing value is used to find the maximum class prior probability in place of missing value

end if

End for

Step 2: Proposed ABC algorithm

Initialization Phase

Initialization of ABC parameters Phi $\in [0,1]$, rho $\in [0,1]$,, alpha, beta, #iterations, initialize all bees solutions to false.

Initial optimal solutions are derived using the following equation.

$$\phi_i = \max\{\mu.\beta_j^k(1-\beta_j^k), \frac{1}{\sqrt{2\pi\sigma_D}} e^{-\frac{(X-\mu_D)^2}{\sigma_X^2}}\}$$

$K = 1, 2...iterations$

$\beta_j^k \in (0,1)$

$x(p,q) \leftarrow x_{min}^q + \phi(x_q^{max} - x_q^{min}), p = 12...N, q = 12...D$

where x(p,q) is pth employed bee with qth dimension.

x_q^{max} is the upper bounds of qth dimension

x_q^{min} is the lower bounds of qth dimension

ϕ is the optimized random number from 0 to 1.

N total employed bees

D is the dimensionality

Step 3: Employed bee phase

In this phase , new candidate solutions are generated for each employed bee.Initially, each value of employed bee is initialized to the new candidate solution(c(i)=x(i)). Candidate solution of each employed bee is computed using the following equation.

$$v(p,q) \leftarrow x(p,q) + \eta(x(p,q) - x(r,q)), p = 12...N, q = 12...D, p \neq r$$

where x(p,q) is pth employed bee with qth dimension.

$\eta = \phi.[x_q^{max} - x_q^{min}]$

x_q^{max} is the upper bounds of qth dimension

x_q^{min} is the lower bounds of qth dimension

ϕ is the optimized random number from 0 to 1.

N total employed bees

D is the dimensionality

Received: 5 Feb 2018/Accepted: 15 Mar 2018

Step 4: Fitness value of the candidate solutions can be computed using

$$\text{fit}_i = \frac{1}{1+\text{fitFunc}_i}; \text{if}\,(\text{fitFunc}_i >= 0)$$
$$=1+\text{abs}(\text{fitFunc}_i); \text{otherwise}$$

Step 5: Onlooker Bee Phase:

In the proposed ABC algorithm, each onlooker bee selects an employed bee in order to improve its feasible solution. This selection is performed using the fitness values of employed bees by roulette wheel as

$$\text{Prob}_i = \frac{\text{fit}_i}{\sum_{i=1}^{N} \text{fit}_i}$$

Where Prob_i is the ith employee bee probability.

Step 6: Scout Bee Phase

The employed bee, which cannot improve self-solution until the abandonment counter reaches to the limit, becomes scout bee.

Apply ensemble classification model (FFNN, Improved Random forest Tree, NAÏVE BAYES) with accuracy acc_i on the selected features in the ith iteration

Proposed Model: Multi-Layered Ensemble Decision Tree

Ensemble classification is defined as the training of multiple base classifiers to detect the disease severity in the test data. Class imbalance and data uncertainty are the growing research direction in the disease severity prediction that aims to discover better classification rate. A feature subset selection extracts a subset of features from the large set of features using selection measures as shown in Figure 2.

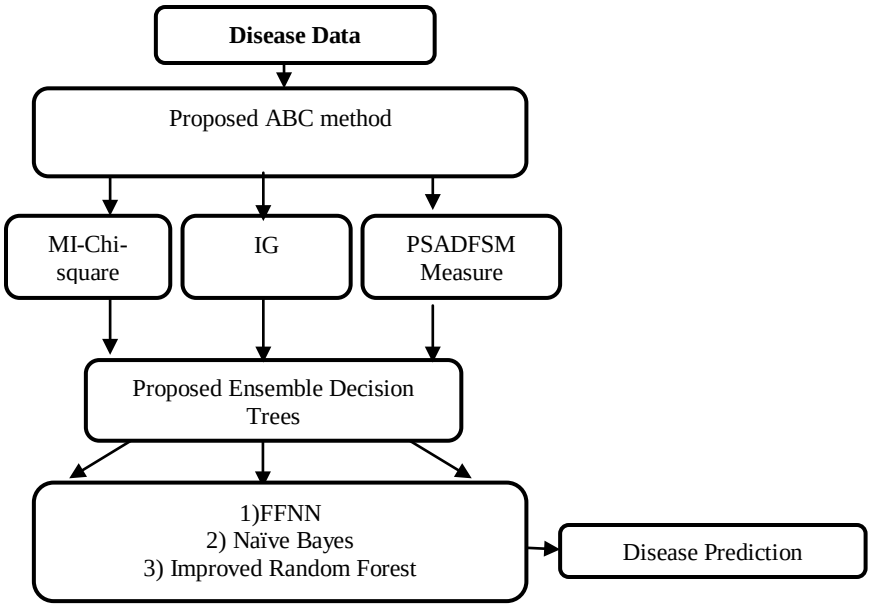


Figure 2: Proposed Model for AD Prediction

In the proposed model, proposed ABC feature selection model with different AD feature selection measures MI-Chi-Square FS, and Probabilistic Similarity Based Feature selection to evaluate the ensemble learning model as shown in Fig 1 and 2.

Proposed Probabilistic Similarity Based AD Feature Selection Measure (PSADFSM)

Proposed attribute selection measure is a hybrid integrated measure of three computational measures: 1) Probabilistic Similarity measure 2) Chi-square measure.

Received: 5 Feb 2018/Accepted: 15 Mar 2018

Probabilistic AD Similarity Measure

Cubic Similarity (CS) measure is used to find the similarity between the disease instances among the class distributions. It is basically derived from KL-divergence and gain ratio. In our feature selection measure, we have proposed a novel probabilistic similarity measure using cubic distance measure. The computation formula used to measure the feature selection is given in equation (1) as

$$PADSM = \sqrt[3]{(\sum_{i=1}^{|D_i|}\sum_{j=1}^{|D_j|}(\sqrt[3]{D_i/|D_i|}-\sqrt[3]{D_j/|D_j|})^2)} * \max\{GainRatio(D_i), GainRatio(D_j)\} \quad \text{-----(1)}$$

Where , D_i is the positive disease class data

D_j is the normal class data

MI-Chi-Square Measure

MI based Chi-square can evaluate the AD data by computing the chi-square statistic with respect to the disease class distribution. This is non-parametric statistical approach used to find the difference between the observed disease distributions to the actual non-disease distribution.

$$MIChi-square(D_i, D_j) = Max\{P(D_i) * \log(D_i / D), (\sum\sum (D_i - D_o)^2 / D_o\}$$

$$YatesCorrection(MIChi-square) = \sum\sum (|D_i - D_o| - 0.5)^2 / D_o \quad \text{------(2)}$$

Multi-Feature Selection based Ensemble Decision Tree Construction

Input: Selected Feature Set SFList[];
Output: Disease patterns
Procedure:
Read Feature Dataset SFList
For each Feature SF[i] in SFList
Do
For each instance $I(A_i)$ in A_i do
Do
Divide the data instances of $SF(D_i)$ into 'k' independent sets.
Select classifier $C_{i/i=1...m}$
Load training features and instances

a) Construct N subset of trained data and N subset of test data sampling with replacement.
b) In the tree growing phase, each and every node select k features at random from N, compute for best split computation as

$$IRF\mathbf{FSM} = -\frac{\sqrt[3]{IG(D_i, D_j) * |D| * PADSM(D_i, D_j)}}{YatesCorr(MI-Chisquare)} \quad \text{----(3)}$$

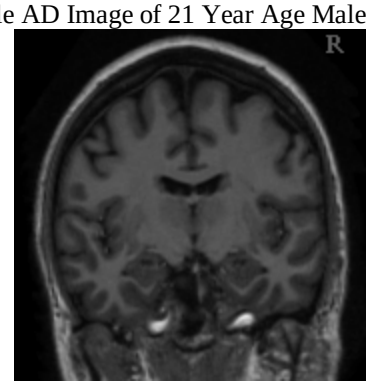
c) Sort the k individual trees according to AD and non-AD in each category.
d) Select the majority voting available in each iteration using ensemble learning.
End while
Calculate misclassified rate and statistical f-measure, accuracy and true positive rates;
Done
Done

IV. Experimental Results

Experimental results for AD detection is performed on OASIS and ADNI databases which is an open access Imaging brain Magnetic Resonance images freely available to the scientific and research community [13]. The dataset comprises a collection of MR images along with computed data in CSV format from 416 patients between 18 to 96 years, where all the subjects are both men and women. 100 patients over 60 years have been clinically diagnosed with very mild to moderate Alzheimer 's disease (AD), rated using the Clinical Dementia Rating (CDR): very mild AD correspond to a CDR value of 0.5, while moderate AD has a CDR value of 2.0. OASIS data used in this study were retrieved from the Alzheimer's disease neuroimaging initiative (ADNI) database (http://adni.loni.usc.edu/)[13].

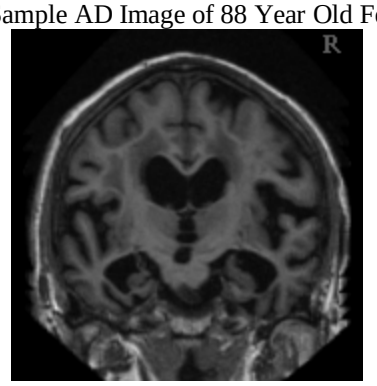
Received: 5 Feb 2018/Accepted: 15 Mar 2018

Sample AD Image of 21 Year Age Male



CDR=0 (Normal Case)

Sample AD Image of 88 Year Old Female



CDR=0.672 (Mild Case)

Sample AD Dataset

OAS1_0027_MR1	F	R	43					1194	0.834	1.47
OAS1_0028_MR1	F	R	86	2	4	27	1	1449	0.738	1.211
OAS1_0029_MR1	M	R	21					1653	0.858	1.062
OAS1_0030_MR1	F	R	65	2	3	29	0	1392	0.764	1.261
OAS1_0031_MR1	M	R	88	1	4	26	1	1419	0.674	1.236
OAS1_0032_MR1	M	R	89	4	1	28	0	1631	0.682	1.076
OAS1_0033_MR1	F	R	80	4	2	29	0	1323	0.735	1.326
OAS1_0034_MR1	M	R	51	5	1	29	0	1538	0.831	1.141
OAS1_0035_MR1	F	R	84	3	2	28	1	1402	0.695	1.252
OAS1_0037_MR1	M	R	27					1313	0.842	1.336
OAS1_0038_MR1	F	R	23					1443	0.839	1.216
OAS1_0039_MR1	M	R	70	4	3	29	0.5	1463	0.772	1.2
OAS1_0040_MR1	F	R	38					1244	0.824	1.411
OAS1_0041_MR1	F	R	62	2		28	0.5	1350	0.758	1.3
OAS1_0042_MR1	M	R	80	4	2	29	0.5	1854	0.709	0.947
OAS1_0043_MR1	M	R	21					1511	0.846	1.162
OAS1_0044_MR1	F	R	47	4	2	30	0	1346	0.829	1.304
OAS1_0045_MR1	M	R	29					1590	0.829	1.104
OAS1_0046_MR1	M	R	64	2		22	0.5	1351	0.787	1.299
OAS1_0047_MR1	F	R	57					1408	0.784	1.247
OAS1_0049_MR1	F	R	20					1329	0.887	1.321
OAS1_0050_MR1	F	R	48					1358	0.841	1.293
OAS1_0051_MR1	F	R	24					1567	0.835	1.12
OAS1_0052_MR1	F	R	78	1	5	23	1	1462	0.697	1.2
OAS1_0053_MR1	F	R	83	1	4	21	1	1384	0.699	1.268

```
Correctly Classified Instances         742               99.4638 %
Incorrectly Classified Instances         4                0.5362 %
Mean absolute error                      0.0065
Total Number of Instances              746

               TP Rate   FP Rate   Precision   Class
                 0.975     0.000     1.000     M
                 1.000     0.025     0.982     F
```

The above result describes the performance of the proposed ABC based ensemble prediction model on the AD dataset.

Table 1: Comparison of AD Prediction Rate for Male and Female Patients Compared to Traditional Models

AD data	PSO+ SVM	PSO+ NN	ABC+ Naiveabayes	ABC+ Ensemble	Proposed Model
Male	0.868	0.796	0.876	0.923	0.9835
Female	0.846	0.815	0.932	0.9535	0.9953

Received: 5 Feb 2018/Accepted: 15 Mar 2018

Table 1 describes the performance of the proposed AD detection rate for male and female category to the traditional models. From the table , it is clearly observed that the proposed model has high computational accuracy for disease prediction for male and female AD data compared to the traditional models.

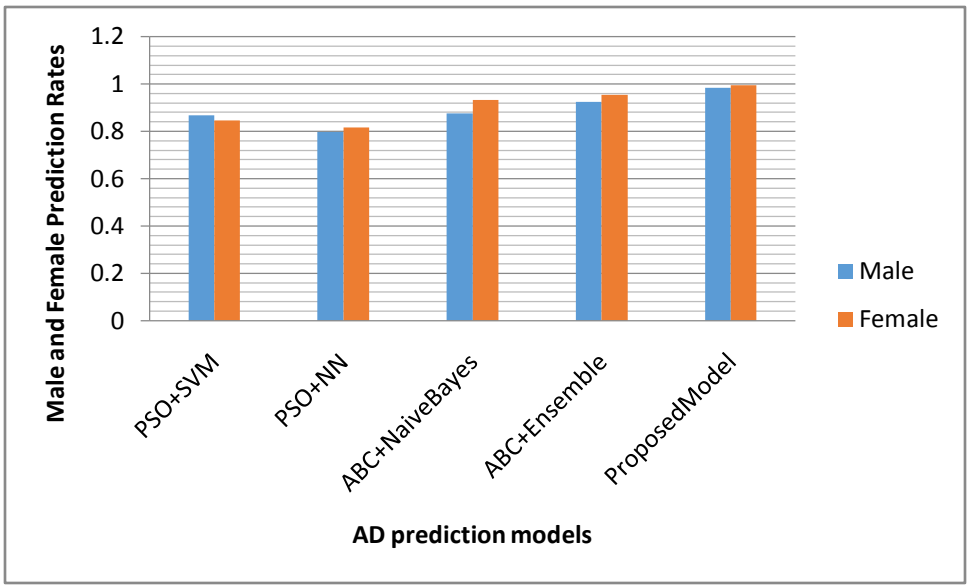


Figure 3: Performance of the proposed model to the existing models in terms of accuracy rate.

Figure3 describes the performance of the proposed AD detection rate for male and female category to the traditional models. From the table, it is clearly observed that the proposed model has high computational accuracy for disease prediction for male and female AD data compared to the traditional models.

Table 2: Comparison of the Proposed Model to the Traditional Models in Terms of Error Rate

AD data	PSO+SVM	PSO+NN	ABC+Ensemble	ABC+NaiveBayes	ProposedModel
Male	0.18	0.25	0.124	0.104	0.083
Female	0.142	0.193	0.1894	0.1194	0.0053

Table 2 describes the performance of the proposed model to the existing models for error rate on AD dataset. From the table, it is clearly observed that the proposed model has low error rate compared to the traditional models.

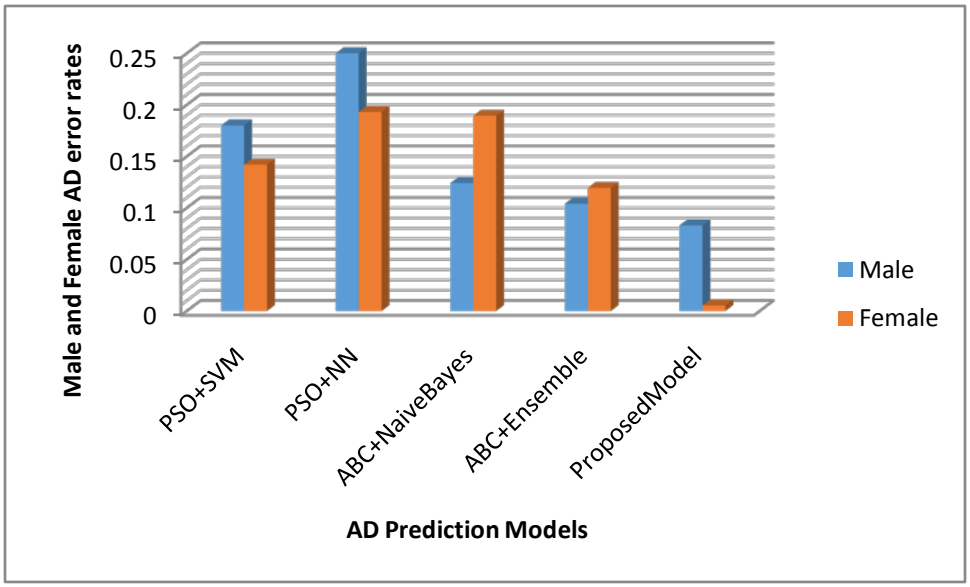


Figure 4 describes the performance of the proposed model to the existing models for error rate on AD dataset. From the table, it is clearly observed that the proposed model has low error rate compared to the traditional models.

Received: 5 Feb 2018/Accepted: 15 Mar 2018

Table 3: Comparison of the Proposed Model to the Existing Models in Terms of Runtime(ms)

Runtime (ms)	PSO+SVM	PSO+NN	ABC+NaiveBayes	ABC+Ensemble	ProposedModel
Male	8344	9437	8693	7963	7084
Female	7937	8935	7937	7619	7329

Table 3, illustrates the performance of the proposed model to the existing models in terms of AD detection runtime is concerned. From the table, it is observed that the proposed model has less runtime compared to the traditional models for AD male and female detection.

V. Conclusion

In this paper, a novel feature selection based ensemble learning model is implemented on Alzheimer's disease dataset. Tradition classification models are not applicable to Alzheimer's dataset due to uncertainty, sparsity and inconsistency. Proposed model optimizes the traditional ensemble learning model with high AD prediction rate and error rate. Experimental results proved that the proposed model improves the prediction rate of Alzheimer's disease for male and female using training data. Also, proposed model has high computational efficiency compared to the traditional feature selection based classification models on the Alzheimer's dataset. In future, this work will be extended to segmentation based Alzheimer's disease prediction on large dataset.

References

[1] Raut, A. and Dalal, V. A machine learning based approach for detection of alzheimer's disease using analysis of hippocampus region from MRI scan. *IEEE International Conference on Computing Methodologies and Communication (ICCMC)*, 2017, 236-242.

[2] Ye, J., Wu, T., Li, J. and Chen, K. Machine learning approaches for the neuroimaging study of Alzheimer's disease. *Computer* **44** (4) (2011) 99-101.

[3] Nie, L., Zhang, L., Meng, L., Song, X., Chang, X. and Li, X. Modeling disease progression via multisource multitask learners: A case study with Alzheimer's disease. *IEEE transactions on neural networks and learning systems* **28** (7) (2017) 1508-1519.

[4] Armañanzas, R., Iglesias, M., Morales, D.A. and Alonso-Nanclares, L. Voxel-Based Diagnosis of Alzheimer's Disease Using Classifier Ensembles. *IEEE journal of biomedical and health informatics* **21** (3) (2017) 778-784.

[5] Liu, M., Zhang, D. and Shen, D. Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. *IEEE transactions on medical imaging* **35** (6) (2016) 1463-1474.

[6] Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R. and Fulham, M.J. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Transactions on Biomedical Engineering* **62** (4) (2015) 1132-1140.

[7] Liu, F., Zhou, L., Shen, C. and Yin, J. Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. *IEEE journal of biomedical and health informatics* **18** (3) (2014) 984-990.

[8] Simpson, I.J., Woolrich, M.W., Andersson, J.R., Groves, A.R. and Schnabel, J. Ensemble learning incorporating uncertain registration. *IEEE transactions on medical imaging* **32** (4) (2013) 748-756.

[9] Escudero, J., Ifeachor, E., Zajicek, J.P., Green, C., Shearer, J., Pearson, S. and Alzheimer's Disease Neuroimaging Initiative. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. *IEEE transactions on biomedical engineering* **60** (1) (2013) 164-168.

[10] Pachauri, D., Hinrichs, C., Chung, M.K., Johnson, S.C. and Singh, V. Topology-based kernels with application to inference problems in Alzheimer's disease. *IEEE transactions on medical imaging* **30** (10) (2011) 1760-1770.

[11] Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Toga, A.W. and Thompson, P.M. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampus segmentation. *IEEE transactions on medical imaging* **29** (1) (2010) 30-43.

[12] Li, F., Tran, L., Thung, K.H., Ji, S., Shen, D. and Li, J. A robust deep model for improved classification of AD/MCI patients. *IEEE journal of biomedical and health informatics* **19** (5) (2015) 1610-1616.

[13] http://adni.loni.usc.edu/

Received: 5 Feb 2018/Accepted: 15 Mar 2018