International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 11989-12003

An Integrated Hybrid Feature Selection based Ensemble Learning Model for Parkinson and Alzheimer's Disease Prediction

Dr. B. Raja Srinivasa Reddy
Professor, Department of Computer Science and Engineering,
Ramachandra College of Engineering, Eluru, Andhra Pradesh, India.

Bala Brahmeswara Kadaru
Assistant Professor, Department of Computer Science and Engineering,
Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India.
Orcid Id: 0000-0001-8327-379X

Abstract
Parkinson's disease and Alzeimer's disease are the most critical diseases which affect the 2% of the people older than 55 years . A number of machine learning models have been implemented for diagnosis and prediction of various neurodegenerative diseases using brain imaging modalities. In the recent time, machine learning approaches are integrated with pattern a recognition technique which has an important role in the process of automatic diagnosis. Extensive amount of research works were done since years for prediction of Parkinson's disease in the early stage. Till date there is no significant approach which will provide optimized performance for disease prediction. Alzheimer's disease is another neurological disease which generally leads to dementia in most cases. Machine learning techniques are considered as the most efficient techniques for the prediction and detection of both Parkinson and Alzheimer's diseases. In this paper, a novel particle swarm optimization technique for feature subset selection was implemented on diseases datasets for ensemble classification model. In this proposed framework , feed forward neural network, hybrid decision tree technique, naïve Bayesian technique are used as majority voting base classifiers for ensemble disease classification and prediction process. Experimental results proved that proposed model has high computational accuracy and true positive rate compared to traditional feature selection measures and ensemble models.

Keywords: FFNN, PSO, Enemble learning, Parkinson, Disease classification.

INTRODUCTION

Now-a-days, Parkinson's disease and Alzheimer's disease are the most common neurodegenerative disease which is found in many geographical regions. It is very difficult and challenging task to diagnose various kinds of neurodegenerative brain issues with the help of medical imaging techniques. Numbers of factors have significant influences in this process of data acquisition. In this disease, the brain neurons are gradually deteriorated along with the growth of disease. It also hampers the creation of vital chemical messenger. Some of the early symptoms include loss of neurotransmitters in the brain. In other words it can also be stated that, there are significant losses of dopaminergic neurons in the substantia nigra (SN).

According to a survey, it is recorded that most people do lose 80% of dopamine prior to any noticeable symptom. According to an international survey, seven to ten million of total population of world are affected with this dangerous and incurable disease. Therefore, it is very much essential to detect Parkinson's disease and its severity at a very early stage. A disease-based pattern construction is constructed which has significant importance in the feature extraction process. It can also be considered that, this number will increase exponentially in the coming years. Aging is a vital factor which generally influences the growth of this disease. In other words it can be stated that, the disease become worse in case of elderly people. But with rapid growth of medical technology, different advanced techniques have been developed and significant improvements are noticed in this field. There are various non-invasive techniques which are usually implemented in order to determine the severity of Parkinson's disease.

Unified Parkinson's Disease Rating Scale (UPDRS) is a common metric which is generally used to major the severity of disease. It includes clinician-scored motor computation process along with auto-evaluation of certain day to day activities such as:- speech, swallowing, handwriting, dressing, etc. Completely affected and severe diseased state is represented by the number 176 and normal state is represented by the number 0. All the intermediary states are included within this range of 0 to 176. The overall process of this assessment is very much lengthy and it is carried out mostly by experienced and professionals. Frequent re-assessment process needs fine-tuning of dosage of drugs or the parameters of the electrical pulse train in deep brain stimulations.

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. 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.

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. There are some frameworks developed that include both the longitudinal image and score information. In these cases, the subjects those have missing data are discarded completely. Thus, this technique decreases the numbers of samples along with the prediction power of the above suggested technique. The partial set of scores generated out of missing data can be properly utilized in the above proposed technique.

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, Parkinson's disease assessment, detection of exudates in retina 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.

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). Gene related Alzheimer's disease is found to be 70% out of all other types of Alzheimer's disease. These genes are known as AD-related genes (ADGs). Different ADGs are predicted and evaluated in the clinical trials. But, each and every ADG can't be detected due to the complex nature of Alzheimer's disease.

In this mining approach, interesting genes out of huge numbers of gene data are selected for prediction. Gene microarrays play a vital role in order to address the complexities of Alzheimer's disease. It also includes parallel activities of various numbers of cellular pathways. There are certain advanced methods those include evaluation of multiple gene expressions at the same time in a particular experiment. It also enables some machine learning approaches to extract interesting biological information out of huge datasets. These approaches are used to analyze the high throughput microarray gene databases. Apart from this, this approach is useful in case of problems which involve less than three classes. Some of the traditional randomized SVM classification technique is modified and extended for very small sets of disease based genes.

Traditional PCA, EM, and fuzzy rule-based approaches are used to classify and cluster the disease datasets with limited dimensional features. Additionally, PCA technique is applied for dimensionality reduction. Initially, the whole data are pre-processed. In the subsequent stage, EM clustering process is carried out to cluster data. The PCA technique is implemented primarily for reduction of dimensionality and separation of the potential noise from data. CART technique is applied for identification of various decision rules from data. All the prediction models are built with the help of fuzzy rule-based technique in every individual cluster. The basic steps used in the fuzzy rule based techniques are:- input fuzzification, production of membership function, extraction of fuzzy rules and defuzzification of outcomes. The degree of inputs is evaluated through Gaussian membership functions. It has an objective to identify all degree of inputs which is present in a particular fuzzy set. During the defuzzification process, Centroid of Area (COA) has the responsibility to return the center of area under the curve.

Proposed work emphasizes on the development of a hybrid intelligence system which will be very much efficient for the diagnosis of Parkinson's or Alzheimer's disease. The main objective of this model is to decrease the overall computation time and the accuracy of diagnosis for uncertain and large datasets. In the proposed work, a non-linear iterative probabilistic PSO method was implemented to reduce the dimensionality problem and hybrid neural network algorithm for ensemble data classification model. A strong correlation was achieved among all the features which may affect the system's accuracy to a great extent.

RELATED WORKS

G. S. Babu et.al, implemented a new technique in order to predict Parkinson's disease with the help of gene expression [1]. In this piece of research work, they presented a gene expression-based advanced technique for the disease prediction. They implemented projection based learning model for meta-cognitive radial basis function network (PBL-

McRBFN). Cognitive component works together with meta-cognitive component in order to decide what-to-learn, when-to-learn and how-to-learn. Independent Component Analysis (ICA) plays an important role in the evaluation process of PBL-McRBFN. It also decreases the feature sets from total genes and the chosen genes have two separate significance levels. The resulted performance of the proposed technique is analyzed and compared with all pre-existing approaches with the help of one-way repeated ANOVA test. Additionally, standard vocal and gait PD data sets are mostly used in the suggested technique.

A. Agarwal, et.al, developed a new strategy for accurate prediction of Parkinson's disease with the help of Extreme Machine Learning approach [2]. Speech impairments analysis is considered as the most effective strategy which is mostly used for the early detection of Parkinson's disease. In this research paper, a static technique is implemented that includes the features of Extreme Machine Learning approach. The main objective of this model is to discriminate the Parkinson's disease patients from normal healthy people. In the evaluation phase, accuracy is achieved up to 90.76% and 0.81 MCC for the training dataset. If the same technique is evaluated with independent dataset of diseased persons, then the accuracy level reduces up to 81.55%. The main limitation of this model is, it incorporate various datasets along with additional and extended features.

Y. Chen et.al*,* presented an advanced prediction and diagnosis mechanism through Parsimonious Fuzzy Neural Networks [3]. This technique can be implemented in practical application which makes it better than that of other traditional fuzzy neural networks. In the evaluation phase, the proposed technique is evaluated on diagnosis and prediction of Parkinson's disease. This approach provides most effective solution for both classifications as well as regression problems.

M. Novotný et.al, proposed a new method for automatic evaluation of articulatory disorders in case of Parkinson's disease [4]. Articulatory deficits is identified and considered as the most useful and widely accepted approach for automatic analysis of speech performance. Accuracy of 80% can be achieved through this technique for a 5ms threshold of absolute difference. This threshold difference is computed among manually labeled references and automatically detected positions. The accuracy can be enhanced up to 88%. In future, this technique can be used as a basic building block for acoustic approaches.

Z. A. Bakar, et.al, developed a new classification scheme which depends upon multilayer perceptrons neural network and ANOVA as feature extraction techniques [5]. An effective diagnosis of Parkinson's disease may include numbers of different test results. In this presented piece of research work, two separate training approaches also known as Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) of Multilayer Perceptrons are implemented together. The LM approach shows 90% accuracy without implementation of feature selection technique, whereas it shows 85% accuracy after the implementation of feature selection technique. Here, MLP is required to improve the true positive rate for the classification of Parkinson's dataset.

A. Bayestehtashk, et.al*,* proposed a new technique which is capable of complete automated assessment of the severity of Parkinson's disease from speech samples [6]. This proposed process is basically split into numbers of sub-tasks, those are:- sustained phonation, diadochokinetic task and reading task. All of these tasks are required to be completed within a fixed amount of time that is, 4mins. In the evaluation phase, total 1582 features are extracted for every individual subject through openSMILE. OpenSMILE is also known as a most common feature extraction tool which is used in many applications. The complete process of feature extraction is refined in order to obtain pitch-related cues, including jitter and shimmer.

B. R. Brewer, et.al, presented a regression approach for quantitative assessment of motor signs [7]. Clinical trials are carried out to detect Parkinson's disease through analyzing various motor symptoms and changes in those systems. The clinical scales are not capable enough for measuring all aspects of motor control. In this model, various commercial sensors are implemented in order to construct a protocol known as Advanced Sensing for Assessment of Parkinson's disease (ASAP). This protocol also evaluates the grip force and this evaluation process is restricted to only three conditions of increasing cognitive load. An advanced regression approach is implemented in order to predict a person's score on the Unified Parkinson Disease Rating Scale.

K. N. Challa, et.al, proposed an enhanced technique for prediction of Parkinson's disease with the help of machine learning approaches [8]. Healthcare professionals are required to perform series of different neurological studies and examination in order to detect the disease accurately. After that, they can successfully predict whether that person is affected by the disease or not. The chance of mis-diagnosis is quite high because there is no standard test to detect Parkinson's disease and its severity. In such scenarios, machine learning approaches are more useful and effective. All of these prediction models involved different efficient machine learning approaches such as logistic regression, bayesian network and multilayer perceptron.

P. Drot´ar, et.al developed a decision support framework for detection of Parkinson's disease using handwriting markers[9]. These handwriting measures are relevant to PD- score. PD-score can be defined as a binary score which is generally used to identify whether a particular sample belongs to a person having disease. The resulted accuracy is detected as 8% along with equivalent values for specificity and sensitivity. The approach focus on each and every disease affected subjects and they are examined in their ON motor state. It can be mentioned that, handwriting is a bio-marker standard and universal bio-marker which is used in many classification

process in order to classify Parkinson's disease data. Though the presented technique is better than that of other existing approaches, but still it has some major limitations which are needed to be resolved in future.

M. Hariharan, et.al, developed an advanced hybrid intelligent system for appropriate identification of Parkinson's disease [10]. 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. In this research paper, a combined technique is presented which is used to enhance the accuracy of detection of PWP. Model-based clustering technique is a special kind of feature weighting approach which is basically implemented in order to enhance the robustness and discriminative capability of original dysphonia features. Here, there are two projection based feature reduction approaches which include 5 features from PCA and 1 feature from LDA. The two step-wise feature subset selection technique is also responsible for decreasing the overall numbers of features. The main objective of the above presented approach is to detect the most appropriate and best subsets of features out of 22 numbers of weighted features. There are three numbers of supervised classifiers such as, LS-SVM, PNN and GRNN which are implemented in order to evaluate the complete effectiveness of the weighted features. Series of different experiments are carried out with the help of various datasets and the resulted outcomes show that, best features are selected by SBS method.

L. Huang, et.al, introduced a new longitudinal clinical score prediction technique in case of Alzheimer's disease [11]. Machine learning approaches are very much efficient for prediction of clinical scores. Linearity assumption and missing data exclusion are major disadvantages of traditional approaches. In this paper, non-linear supervised sparse regression based random forest (RF) framework is developed 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.

C. Kotsavasiloglou, et.al, proposed a new machine learning based classification technique [12]. This method emphasizes on simple drawing movements in Parkinson'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. Both the hands and its movements are monitored by the proposed technique. The simple drawing tasks make it possible to differentiate healthy state from disease affected state. The impairment of the coordination of antagonist muscular systems can be analysed in a better way in case of simple and easy drawing tasks. There exists a single problem in the above suggested technique and that is, this approach is only efficient to evaluate the impairment in motor coordination in case of upper limbs only. This problem must be resolved in future in order to propose a complete solution. Among the advantages, velocity micro variations[13] can be noticed at a few milliseconds level. In case of every digitized line of subject, they computed four major metrics.

Y. Miao, et.al, proposed a new gene identification approach for prediction and detection of Alzheimer's and Parkinson diseases [14]. 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. Furthermore, a two-stage cascading classifier is introduced in order to detect ADGs. According to the first step, the whole classification process depends upon the concept of relevance vector machine. In the subsequent phase, the outcomes of three major classification schemes (SVM, RF and ELM) are integrated with each other by the method of voting. This model achieves accuracy, sensitivity and specificity as 78.77%, 83.10% and 74.67%, respectively.

M. Nilashi, et.al developed a hybrid intelligent system for prediction and identification of Parkinson's disease [15]. They implemented a hybrid approach which merges two significant techniques Total-UPDRS and Motor-UPDRS. ISVR technique is implemented in order to estimate the Total-UPDRS and Motor-UPDRS. Real-world PD dataset is extracted and considered from UCI. Additionally, the problems related to data processing time must be resolved properly.

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.

R. Prashanth, et.al, presented a method for high-accuracy detection of early Parkinson's disease [17]. Both multimodal features and machine learning techniques are combined and implemented here in order to classify the disease data. In this paper, an emphasis is given on the preclinical markers of non-motor features of RBD and olfactory loss, CSF measurements, and dopaminergic imaging features.

S. L. Smith, et.al, implemented new computational technique for the diagnosis and detection of Parkinson's disease [18]. The presented research technique explains the procedure for application of evolutionary algorithms (EAs). This method has the complete responsibility to analyse motor function in humans with Parkinson's disease and in animal models of Parkinson's disease. Usually commercial sensors are used to gather human data through numbers of non-invasive procedures. EAs are implemented in order to take part in the classification process. Additionally, efficient disease diagnosis and disease monitoring are two other important aspects of this presented approach. The outcomes also explain the classification mechanism of fruit flies with and without genetic mutations that cause Parkinson's disease. Certain measurements of the proboscis extension reflex are used during the whole process. This approach is implemented both for human as well as animal studies of Parkinson's disease. The presented research work is responsible for reviewing the application of EAs in the assessment of movements. This technique plays a very important role in discriminating patients from healthy people. It not only detects the disease, but also identifies its severity.

E. E. Tripoliti, et.al*,* presented a new method for automated diagnosis of diseases based on classification [19]. It is dynamic determination of number of trees in random forests approach. Proper and accurate diagnosis of diseases is very hard and challenging biomedical issues. Furthermore, the above suggested technique generates an ensemble not only accurate, but also diverse. It satisfies two very important characteristics. The complete process depends upon an online fitting procedure and all the computation is carried out by using eight numbers of biomedical datasets and five numbers of random forests approaches. The outcomes of the evaluation phase results 90% accuracy rate. Through implementing the above approach, the ensemble classifier is capable of identifying optimal size of classifier.

DISEASE DATASETS

Sample Alzeimer's Dataset

@relation 'Alz Train (Final)'

@attribute ACE_CD143_Angiotensin_Converti numeric

@attribute ACTH_Adrenocorticotropic_Hormon numeric

@attribute AXL numeric

@attribute Adiponectin numeric

@attribute Alpha_1_Antichymotrypsin numeric

@attribute Alpha_1_Antitrypsin numeric

@attribute Alpha_1_Microglobulin numeric

@attribute Alpha_2_Macroglobulin numeric

@attribute Angiopoietin_2_ANG_2 numeric

@attribute Angiotensinogen numeric

@attribute Apolipoprotein_A_IV numeric

@attribute Apolipoprotein_A1 numeric

@attribute Apolipoprotein_A2 numeric

@attribute Apolipoprotein_B numeric

@attribute Apolipoprotein_CI numeric

@attribute Apolipoprotein_CIII numeric

@attribute Apolipoprotein_D numeric

@attribute Apolipoprotein_E numeric

@attribute Apolipoprotein_H numeric

@attribute B_Lymphocyte_Chemoattractant_BL numeric

@attribute BMP_6 numeric

@attribute Beta_2_Microglobulin numeric

@attribute Betacellulin numeric

@attribute C_Reactive_Protein numeric

@attribute CD40 numeric

@attribute CD5L numeric

@attribute Calbindin numeric

@attribute Calcitonin numeric

@attribute CgA numeric

@attribute Clusterin_Apo_J numeric

@attribute Complement_3 numeric

@attribute Complement_Factor_H numeric

@attribute Connective_Tissue_Growth_Factor numeric

@attribute Cortisol numeric

@attribute Creatine_Kinase_MB numeric

@attribute Cystatin_C numeric

@attribute EGF_R numeric

@attribute EN_RAGE numeric

@attribute ENA_78 numeric

@attribute Eotaxin_3 numeric

@attribute FAS numeric

@attribute FSH_Follicle_Stimulation_Hormon numeric

@attribute Fas_Ligand numeric
@attribute Fatty_Acid_Binding_Protein numeric
@attribute Ferritin numeric
@attribute Fetuin_A numeric
@attribute Fibrinogen numeric
@attribute GRO_alpha numeric
@attribute Gamma_Interferon_induced_Monokin numeric
@attribute Glutathione_S_Transferase_alpha numeric
@attribute HB_EGF numeric
@attribute HCC_4 numeric
@attribute Hepatocyte_Growth_Factor_HGF numeric
@attribute I_309 numeric
@attribute ICAM_1 numeric
@attribute IGF_BP_2 numeric
@attribute IL_11 numeric
@attribute IL_13 numeric
@attribute IL_16 numeric
@attribute IL_17E numeric
@attribute IL_1alpha numeric
@attribute IL_3 numeric
@attribute IL_4 numeric
@attribute IL_5 numeric
@attribute IL_6 numeric
@attribute IL_6_Receptor numeric
@attribute IL_7 numeric
@attribute IL_8 numeric
@attribute IP_10_Inducible_Protein_10 numeric
@attribute IgA numeric
@attribute Insulin numeric
@attribute Kidney_Injury_Molecule_1_KIM_1 numeric
@attribute LOX_1 numeric
@attribute Leptin numeric
@attribute Lipoprotein_a numeric
@attribute MCP_1 numeric
@attribute MCP_2 numeric
@attribute MIF numeric
@attribute MIP_1alpha numeric
@attribute MIP_1beta numeric
@attribute MMP_2 numeric
@attribute MMP_3 numeric
@attribute MMP10 numeric
@attribute MMP7 numeric
@attribute Myoglobin numeric
@attribute NT_proBNP numeric
@attribute NrCAM numeric
@attribute Osteopontin numeric
@attribute PAI_1 numeric
@attribute PAPP_A numeric
@attribute PLGF numeric
@attribute PYY numeric
@attribute Pancreatic_polypeptide numeric
@attribute Prolactin numeric
@attribute Prostatic_Acid_Phosphatase numeric
@attribute Protein_S numeric
@attribute Pulmonary_and_Activation_Regulat numeric
@attribute RANTES numeric
@attribute Resistin numeric
@attribute S100b numeric
@attribute SGOT numeric
@attribute SHBG numeric
@attribute SOD numeric
@attribute Serum_Amyloid_P numeric
@attribute Sortilin numeric
@attribute Stem_Cell_Factor numeric
@attribute TGF_alpha numeric
@attribute TIMP_1 numeric
@attribute TNF_RII numeric
@attribute TRAIL_R3 numeric
@attribute TTR_prealbumin numeric
@attribute Tamm_Horsfall_Protein_THP numeric
@attribute Thrombomodulin numeric
@attribute Thrombopoietin numeric
@attribute Thymus_Expressed_Chemokine_TECK numeric
@attribute Thyroid_Stimulating_Hormone numeric
@attribute Thyroxine_Binding_Globulin numeric
@attribute Tissue_Factor numeric

@attribute Transferrin numeric

@attribute Trefoil_Factor_3_TFF3 numeric

@attribute VCAM_1 numeric

@attribute VEGF numeric

@attribute Vitronectin numeric

@attribute von_Willebrand_Factor numeric

@attribute male numeric

@attribute E4 numeric

@attribute E3 numeric

@attribute E2 numeric

@attribute Class {Control,Impaired}

@data

2.0031,-1.386294,1.098387,-5.360193,1.740466,-12.631361,-2.577022,-72.65029,1.064711,2.510547,-1.427116,-7.402052,-0.261365,-4.624044,-1.272966,-2.312635,2.079442,3.754521,-0.157349,2.296982,-2.200744,0.693147,34,-4.074542,-0.796415,0.09531,33.213634,1.386294,397.653601,3.555348,-10.363053,3.573725,0.530628,10,-1.710172,9.041922,-0.135454,-3.688879,-1.349543,53,-0.083382,-0.651672,3.101492,2.520871,3.329165,1.280934,-7.035589,1.38183,2.949822,1.064127,6.559746,-3.036554,0.587787,3.433987,-0.190779,5.609472,5.121987,1.282549,4.192081,5.731246,-6.571283,-3.244194,2.484907,1.098612,0.26937,0.642796,4.805045,1.711325,6.242223,-6.812445,-0.625825,-1.204295,1.704748,-1.529063,-4.268698,6.740519,1.980509,-1.237874,4.968453,3.258097,4.478566,-2.207275,-3.270169,-3.773503,-1.89712,4.553877,5.003946,5.356586,1.003502,-2.902226,4.442651,3.218876,0.578781,0,-1.620527,-1.784998,-0.84397,-6.214608,-16.475315,1.561856,-0.941609,-1.89712,5.609472,-5.599422,4.908629,4.174387,8.649098,15.204651,-0.061875,-0.1829,2.944439,-3.09581,-1.340566,-0.102633,4.149327,-3.863233,-1.427116,2.04122,3.332205,-3.381395,3.258097,22.034564,-0.040822,-3.146555,0.987624,6.297754,4.348108,12.019678,0,0,1,0,Control

1.52066,-1.714798,-0.145276,-5.809143,1.193922,-13.642963,-2.882404,-136.529178,0.832909,1.976365,-1.660731,-7.684284,-0.653926,-3.976069,-1.714798,-2.748872,1.335001,2.753056,-0.344839,1.673121,-2.062421,0.336472,49,-8.04719,-1.24152,0.09531,22.166092,2.116256,347.863875,2.772589,-16.108237,4.474569,0.641854,10,-1.383559,8.954157,-0.732987,-4.755993,-1.390672,62,-0.634878,-1.563,1.36001,0.906301,3.176872,1.410987,-7.195437,1.412679,2.762231,0.889015,7.745463,-3.649659,0.09531,2.397895,-0.462017,5.181784,4.66591,1.274133,2.616102,4.149327,-8.180721,-4.645992,1.824549,-0.248461,0.185686,0.096686,1.005622,1.691393,5.049856,-6.319969,-1.446619,-1.191191,1.163151,-1.662268,-5.843045,6.767343,0.400596,-2.302585,4.049508,2.397895,2.866631,-2.302585,-2.733368,-4.030227,-1.386294,4.248495,4.744932,5.01728,0.438372,-2.935541,4.51086,2.890372,-0.891598,-0.139262,-1.636682,-2.259135,-1.660731,-6.645391,-16.475315,1.435728,0.336472,-2.207275,5.723585,-5.381699,3.810182,3.433987,10.858497,12.282857,-0.415515,-0.924034,2.944439,-3.166721,-1.534276,-0.922967,2.791992,-4.990833,-1.89712,1.435085,2.890372,-3.729701,2.639057,17.476191,-0.223144,-3.540459,0.986667,6.270988,4.400247,12.302271,1,1,1,0,Control

0.946207,-1.89712,0.529822,-6.119298,0.832909,-12.813142,-3.270169,-149.604408,0.788457,2.106653,-2.040221,-7.751725,-0.941609,-7.28883,-1.660731,-2.375156,1.526056,4.2548,-0.344839,1.673121,-1.845213,-0.040822,41,-7.5811,-1.096541,-0.248461,27.12044,1.098612,336.953157,3.178054,-16.54531,2.40792,0.875469,4.9,-1.739232,8.740337,-0.699799,-5.067206,-1.349543,64,-0.71335,-1.298876,3.101492,1.59754,3.440588,1.131402,-7.621105,1.338425,2.739315,1.142197,6.262563,-3.506558,0.405465,2.995732,-0.63604,5.438079,2.031412,1.286356,3.476091,6.705891,-6.907755,-3.296837,1.871802,0.832909,0.096224,0.431156,4.009916,1.698489,5.451038,-6.645391,-0.839808,-1.143534,1.223775,-1.705141,-6.319969,6.359574,2.103023,-2.120264,3.551208,2.564949,3.265601,-2.120264,-4.135167,-5.968191,-1.771957,4.330733,4.521789,5.31812,0,-2.590291,3.367296,2.833213,-0.597837,-0.527633,-1.659412,-2.259135,-1.560648,-6.812445,-20.660678,1.301297,-0.693147,-3.123566,5.509388,-6.032287,4.037285,3.401197,8.649098,12.422205,-0.776529,-0.629974,2.890372,-3.133732,-1.675252,-0.338675,3.342694,-3.963316,-2.120264,1.648659,2.70805,-3.649659,2.639057,17.911392,-0.371064,-4.017384,0.985465,5.740789,3.747667,14.257867,1,0,1,1,Control

Parkinson's Dataset

The dataset used in this experimental study is a voice based records originally derived from Oxford University by Max Little. The disease consists of 195 records collected from 31 people whom 23 are suffering from Parkinson's disease. We extended the original data to 10000 records using synthetic program. The features used to classify the Parkinson's disease are:

%Attribute Information:

%Matrix column entries (attributes):
%name - ASCII subject name and recording number
%MDVP:Fo(Hz) - Average vocal fundamental frequency
%MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
%MDVP:Flo(Hz) - Minimum vocal fundamental frequency
%MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several
%measures of variation in fundamental frequency
%MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
%NHR,HNR - Two measures of ratio of noise to tonal components in the voice
%status - Health status of the subject (one) - Parkinson's, (zero) - healthy
%RPDE,D2 - Two nonlinear dynamical complexity measures
%DFA - Signal fractal scaling exponent
%spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

@relation Parkinson

@attribute name string
@attribute MDVP:Fo(Hz) real
@attribute MDVP:Fhi(Hz) real
@attribute MDVP:Flo(Hz) real
@attribute MDVP:Jitter(Pct) real
@attribute MDVP:Jitter(Abs) real
@attribute MDVP:RAP real
@attribute MDVP:PPQ real
@attribute Jitter:DDP real
@attribute MDVP:Shimmer real
@attribute MDVP:Shimmer(dB) real
@attribute Shimmer:APQ3 real
@attribute Shimmer:APQ5 real
@attribute MDVP:APQ real
@attribute Shimmer:DDA real
@attribute NHR real
@attribute HNR real
@attribute RPDE real
@attribute DFA real
@attribute spread1 real
@attribute spread2 real
@attribute D2 real
@attribute PPE real
@attribute status {0,1}

@data
phon_R01_S01_1,119.992,157.302,74.997,0.00784,0.00007,0.0037,0.00554,0.01109,0.04374,0.426,0.02182,0.0313,0.02971,0.06545,0.02211,21.033,0.414783,0.815285,-4.813031,0.266482,2.301442,0.284654,1
phon_R01_S01_2,122.4,148.65,113.819,0.00968,0.00008,0.00465,0.00696,0.01394,0.06134,0.626,0.03134,0.04518,0.04368,0.09403,0.01929,19.085,0.458359,0.819521,-4.075192,0.33559,2.486855,0.368674,1
phon_R01_S01_3,116.682,131.111,111.555,0.0105,0.00009,0.00544,0.00781,0.01633,0.05233,0.482,0.02757,0.03858,0.0359,0.0827,0.01309,20.651,0.429895,0.825288,-4.443179,0.311173,2.342259,0.332634,1
phon_R01_S01_4,116.676,137.871,111.366,0.00997,0.00009,0.00502,0.00698,0.01505,0.05492,0.517,0.02924,0.04005,0.03772,0.08771,0.01353,20.644,0.434969,0.819235,-4.117501,0.334147,2.405554,0.368975,1
phon_R01_S01_5,116.014,141.781,110.655,0.01284,0.00011,0.00655,0.00908,0.01966,0.06425,0.584,0.0349,0.04825,0.04465,0.1047,0.01767,19.649,0.417356,0.823484,-3.747787,0.234513,2.33218,0.410335,1
phon_R01_S01_6,120.552,131.162,113.787,0.00968,0.00008,0.00463,0.0075,0.01388,0.04701,0.456,0.02328,0.03526,0.03243,0.06985,0.01222,21.378,0.415564,0.825069,-4.242867,0.299111,2.18756,0.357775,1
phon_R01_S02_1,120.267,137.244,114.82,0.00333,0.00003,0.00155,0.00202,0.00466,0.01608,0.14,0.00779,0.00937,0.01351,0.02337,0.00607,24.886,0.59604,0.764112,-5.634322,0.257682,1.854785,0.211756,1
phon_R01_S02_2,107.332,113.84,104.315,0.0029,0.00003,0.00144,0.00182,0.00431,0.01567,0.134,0.00829,0.00946,0.01256,0.02487,0.00344,26.892,0.63742,0.763262,-6.167603,0.183721,2.064693,0.163755,1
phon_R01_S02_3,95.73,132.068,91.754,0.00551,0.00006,0.00293,0.00332,0.0088,0.02093,0.191,0.01073,0.01277,0.01717,0.03218,0.0107,21.812,0.615551,0.773587,-5.498678,0.327769,2.322511,0.231571,1
phon_R01_S02_4,95.056,120.103,91.226,0.00532,0.00006,0.00268,0.00332,0.00803,0.02838,0.255,0.01441,0.01725,0.02444,0.04324,0.01022,21.862,0.547037,0.798463,-5.011879,0.325996,2.432792,0.271362,1
phon_R01_S02_5,88.333,112.24,84.072,0.00505,0.00006,0.00254,0.0033,0.00763,0.02143,0.197,0.01079,0.01342,0.01892,0.03237,0.01166,21.118,0.611137,0.776156,-5.24977,0.391002,2.407313,0.24974,1

PROPOSED MODEL

The main objective of the proposed model is to determine the Parkinson and Alzheimer's disease prediction using the novel machine learning technique. This approach plays a vital role to detect and predict diseases at a very early stage. Additionally, an extraction procedure is carried out in order to extract different potential speech features. This results an effective prediction of clinical ratings out of speech features. All the

classical pitch detection approaches never include time-frequency resolution for capturing very fine tremors. Hence in order to overcome the limitations of all traditional pitch estimation approaches, the proposed scheme was optimized.

Different linear and non-linear prediciton technques are used to find the non-linear and pitch/amplitude perturbation for discrimination between diseased stated and normal state.

All of these speech samples are elicited, recorded and analyzed automatically. In fact, this technique is very much efficient in terms of true positive rate as compared to the other traditional techniques. Such types of assessment process are responsible for monitoring changes with respect to time. These approaches usually implement for the speech analysis of Parkinson disease. All the data gathered from an individual clinic may be biased because of the subjective nature of clinical assessments.

Basically, high dimensional disease datasets with missing values minimizes the true positive rate and accuracy in large datasets. Among these features, most of them are unsuitable for classification. After successful extraction of features, an optimal subset of the features is chosen for the process of classification. The classifier has the responsibility to split training data set into two separate datasets such as:- test dataset and validation dataset. Traditional neural network approach such as Feed forward neural network is extended to predict the high dimensional disease patterns with high true positive rate using ensemble learning model.

In this proposed model a novel particle swarm optimization (PSO) method was developed to improve the feature subset selection for classification problem.

Proposed PSO is a multi-objective technique which finds the local and global optimum measures by iteratively searching in a high dimensional feature subspace. In this filtering method, each attribute is tested for missing values.

Proposed PSO is used to select optimal feature values to improve the overall classification true positive rate on high dimensional disease datasets.

Proposed Improved PSO based Ensemble Classification Algorithm:

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

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

For each attribute A(i) in the $HD^1, HD^2 .... HD^n$

do

For each instance value I(j) in the A(i)

Do

if(isNumerical(I(j)) && I(j)==empty)

then

$$I(j) = \frac{\sum_{j=1/i\neq j}^{n} (|\operatorname{Max}\{A(I(j))\} - \operatorname{Mode}_{A(I(j))}|)}{(\operatorname{Max}_{A(I(j))} - \operatorname{Min}_{A(I(j))})} * \text{Scaling\_factor} \quad (1)$$

$\text{Scaling\_factor} \in (0,1)$

end if

if(isCategoricalA$_i$) && A$_i$(I)==empty)

then

$$P(j) = \operatorname{Prior}\operatorname{Prob}(A(i), \text{class}(m));$$

$$I(j) = \frac{\sum_{j=1/i\neq j}^{n} (|\operatorname{Max}\{P(j)\} - \operatorname{Min}\{P(j)\}|)}{\text{Scaling\_factor}} \quad (2)$$

$\text{Scaling\_factor} \in (0,1)$

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

end if

End for

Step 2: //Optimal feature selection for feed-forward neural network and Decision tree techniques

Initializing particles with feature space, number of iterations, velocity, number of particles etc.
Compute hybrid velocity and position for each particle in 'd' dimensions using the following equations

$$V_p(d(i+1),i) = \xi_p.[\omega_p(d(i),i).V_p(d(i),i).\phi_{m1}(pPBst_i - X_p(d(i),i)) + \phi_{m2}(gPBst_i - X_p(d(i),i))]$$

$$X_p(d+1,i) = X_p(d(i+1),i) + V_p(d(i+1),i)$$

ξ_p is the particle convergence parameter formulated as

$$\xi_p = \frac{1}{|2\pi - (\theta_{m1}+\theta_{m2}) - \sqrt{(\theta_{m1}+\theta_{m2})^2 - 4(\theta_{m1}+\theta_{m2})}|}$$

where $\theta_{m1}, \theta_{m2} \in$ Orthogonal gaussian functions (3)

In this optimized model, particle inertia weights are calculated using the following equation as

$$\omega_p(d(i+1),i) = \omega_{p\_max} - (\text{Iter}_{current} / \text{Iter}_{max}).(\omega_{p\_max} - \omega_{p\_min})$$

ω_{p_max} : maximum inertia

ω_{p_min} : minimum inertia

Iter_{max} : maximum iteration

$\text{Iter}_{current}$: current iteration (4)

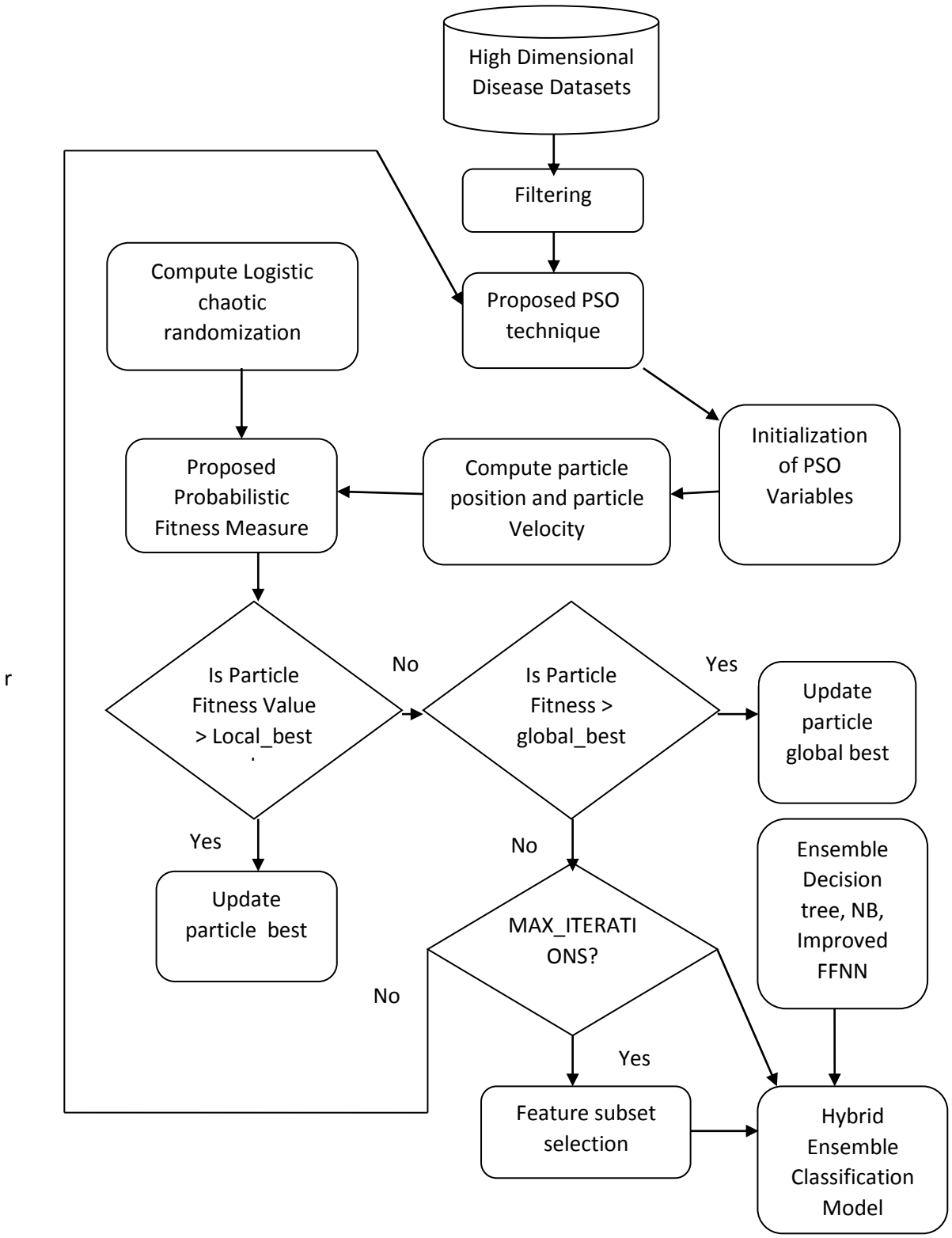

Figure 1: Proposed Model

Step 3: Computation of probabilistic correlated fitness value for orthogonal Gaussian function.

In the hybrid particle swarm optimization model, fitness value and initial logistic chaotic parameters are computed using the following equation as

$$\mathrm{Rand}_i = \max\{\psi_j^r(1-\psi_j^r), \frac{1}{\psi_X\sqrt{2\pi}} e^{-\frac{(X-\mu_X)}{\sqrt{\sigma_X}}}\}$$

$$r = 1, 2 \ldots n$$

$$\psi_j^k \in \mathrm{logisticchaostic}(0,1) \quad (5)$$

Proposed fitness method is given as

$$\mathrm{Fit}_i = k_1.TP_i + k_2.(1 - \frac{\sum_{i=1}^{|F|} fs_i}{N_A})$$

where $k_1, k_2 \in R_i$

fs_i is the feature selected labels 1 or 0 .

'1' indicates selected attribute ,

'0' indicates non-selected attribute.

$|N_A|$ indicates total attributes size. (6)

TP_i : Truepositive rate of the selected attributes in the current iteration.

Step 4: Apply ensemble classification model with highest true positive rate acc_i on the selected features in the ith iteration. Here, ensemble decision trees, naïve bayes, feed forward neural network are used as base classifiers for high dimensional classification.

Selected attributes list SAList

For each attribute SA[i] in SAList

Do

For each instance $I(A_i)$ in A_i do

Do

Select the base classifiers from the list $C_{i/i=1\ldots m}$

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 using the feature selection measures from the paper [21]

$$\mathrm{RFPSOFSM} = -\frac{\sqrt[3]{\mathrm{PDRelief}(D_i, D_j) * |D| * \mathrm{ProbPSO}(D_i, D_j)}}{\mathrm{YatesCorr}(\mathrm{MI-Chisquare})} \quad (7)$$

(c) Sort the k individual trees according to Alzheimer's and non- Alzheimer's.

d). Select the majority voting available in each tree using ensemble learning.

End while

Calculate misclassified rate and statistical f-measure ,accuracy and true positive rates;

Done

Done

Step 5: To each particle compute fitness value and true positive rate of the attribute in the previous step.

Step 6: Update the particle position, velocity, local best, global best, fitness and logistic chaos randomization functions.

Step 7: This procedure is repeated until minimum error rate or high true positive rate.

Experimental Results

Experimental results are carried out on Parkinson's UCI dataset and Alzheimer's dataset [http://course1.winona.edu /bdeppa/Stat%20425/Data/Alz%20Train%20(Final).csv]. Here training dataset is mapped to original Parkinson's dataset for data filtering and ensemble classification model. We have extended the traditional UCI Parkinson's training data ranging from 10000 to 1000000 instances.

Generated Ensemble Prediction Results for Alzheimer's disease:

[Hepatocyte_Growth_Factor_HGF=0.09531, E2=0]: 34 ==> [male=0]: 23

[Cortisol=12, E3=1]: 34 ==> [Class=Control]: 23

[IL_13=1.278484]: 34 ==> [male=0]: 23

[E3=1, E2=0, Class=Control]: 158 ==> [E4=0]: 107

[TTR_prealbumin=2.944439]: 31 ==> [E3=1, E2=0, Class=Control]: 21

[TTR_prealbumin=2.944439]: 31 ==> [E4=0, E3=1, Class=Control]: 21

[PAPP_A=-2.971157]: 31 ==> [male=0, E3=1, E2=0]: 21

[Cortisol=12, E2=0]: 31 ==> [E3=1, Class=Control]: 21

[TTR_prealbumin=2.944439]: 31 ==> [E2=0, Class=Control]: 21

[TTR_prealbumin=2.944439]: 31 ==> [E4=0, Class=Control]: 21

[PYY=2.995732, E3=1]: 31 ==> [E4=0]: 21

[PYY=2.995732, E3=1]: 31 ==> [male=0]: 21

[PAPP_A=-2.971157]: 31 ==> [male=0, E2=0]: 21

[IL_13=1.278484, E3=1]: 31 ==> [Class=Control]: 21

[IL_13=1.278484, E3=1]: 31 ==> [E4=0]: 21

[IL_13=1.278484, E2=0]: 31 ==> [male=0]: 21

[IL_13=1.278484, E3=1]: 31 ==> [male=0]: 21

[Cortisol=12, E2=0]: 31 ==> [Class=Control]: 21

[Cortisol=11]: 31 ==> [E3=1, Class=Control]: 21

[RANTES=-6.571283]: 31 ==> [Class=Control]: 21

[Cortisol=11]: 31 ==> [Class=Control]: 21

[male=0, E4=1, E3=1]: 56 ==> [E2=0, Class=Control]: 38

[male=0, E4=1, E3=1, E2=0]: 56 ==> [Class=Control]: 38

[male=0, E4=1, E3=1]: 56 ==> [Class=Control]: 38

[E2=0, Class=Control]: 162 ==> [male=0, E3=1]: 110

[TTR_prealbumin=2.772589, E3=1, E2=0]: 35 ==> [Class=Control]: 24

[TTR_prealbumin=2.772589, E3=1, E2=0]: 35 ==> [male=0]: 24

[IL_13=1.282549, E2=0]: 35 ==> [E3=1, Class=Control]: 24

[IL_13=1.282549, E2=0]: 35 ==> [Class=Control]: 24

[MCP_2=1.530376]: 35 ==> [E4=0]: 24

[MCP_2=1.530376]: 35 ==> [male=0]: 24

[MCP_2=1.530376, E3=1]: 32 ==> [male=0, Class=Control]: 22

[IL_13=1.286356, E2=0]: 32 ==> [E3=1, Class=Control]: 22

[IL_13=1.286356, E3=1, E2=0]: 32 ==> [Class=Control]: 22

[TTR_prealbumin=2.772589, Class=Control]: 32 ==> [E4=0]: 22

[IL_13=1.286356, E2=0]: 32 ==> [Class=Control]: 22

[ENA_78=-1.367775]: 32 ==> [E4=0, E3=1]: 22

[ENA_78=-1.367775]: 32 ==> [male=0, E3=1]: 22

[Beta_2_Microglobulin=0.182322]: 32 ==> [male=0, E3=1]: 22

[Beta_2_Microglobulin=0.09531, E3=1]: 32 ==> [E4=0]: 22

[B_Lymphocyte_Chemoattractant_BL=2.371361]: 32 ==> [E3=1, Class=Control]: 22

[PYY=2.995732]: 32 ==> [male=0]: 22

[ENA_78=-1.367775]: 32 ==> [E4=0]: 22

[Beta_2_Microglobulin=0.182322]: 32 ==> [male=0]: 22

[TTR_prealbumin=2.890372, E3=1]: 42 ==> [Class=Control]: 29

[E4=1, E2=0, Class=Control]: 55 ==> [male=0, E3=1]: 38

[male=1, E4=0]: 65 ==> [Class=Control]: 45

[E3=1, Class=Control]: 182 ==> [male=0]: 126

[E2=0]: 228 ==> [E3=1, Class=Control]: 158

[E3=1, E2=0, Class=Control]: 158 ==> [male=0]: 110

[Beta_2_Microglobulin=0.09531]: 33 ==> [E3=1, E2=0, Class=Control]: 23

[Beta_2_Microglobulin=0.09531]: 33 ==> [E2=0, Class=Control]: 23

[PYY=2.833213]: 33 ==> [Class=Control]: 23

[E2=0, Class=Control]: 162 ==> [male=0]: 113

[male=0, E3=1, Class=Control]: 126 ==> [E4=0]: 88

[TTR_prealbumin=2.944439, E3=1]: 30 ==> [E2=0, Class=Control]: 21

[TTR_prealbumin=2.944439, E3=1]: 30 ==> [E4=0, Class=Control]: 21

[MCP_2=1.530376, Class=Control]: 30 ==> [E4=0, E3=1]: 21

[ENA_78=-1.367775, E3=1]: 30 ==> [E4=0, Class=Control]: 21

Table 1: Classification accuracy of Proposed Model to the existing Models on Parkinson Dataset

Models	TruePositive	Accuracy	Error Rate
PCA Ensemble	0.794	0.7834	0.284
PSO Ensemble	0.8324	0.8624	0.245
Decision tree based Ensemble	0.8574	0.8973	0.1986
Multi-feature Ensemble Decision tree	0.932	0.967	0.0952
Proposed Model	0.9536	0.9723	0.046

Table 1, describes the performance analysis of the proposed model with the traditional classification models in terms of true positive rate, accuracy and error rate on Parkinson's dataset. From the table , it is observed that proposed model has high computational accuracy in terms of accuracy , true positive rate and error rate are concerned.

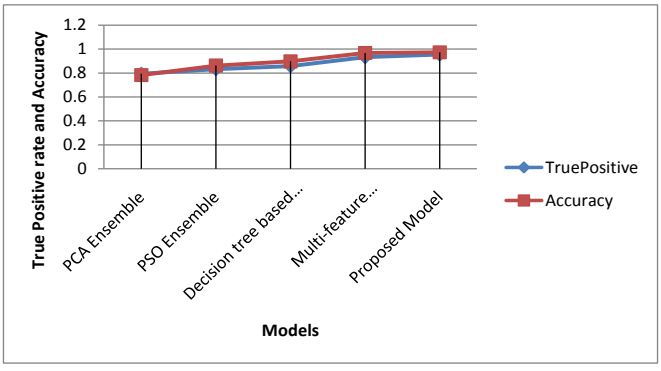

Figure 2: Graphical comparison of the classification accuracy and true positive rate on Parkinson Dataset**.**

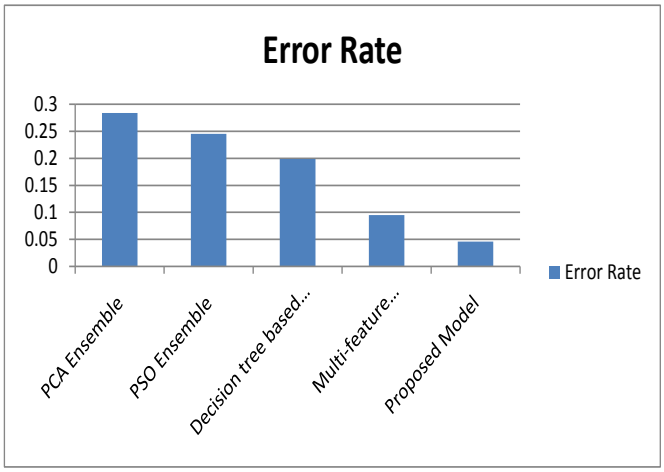

Figure 3: Graphical comparison of the classification error rate on Parkinson Dataset**.**

Table 2: Classification accuracy of Proposed Model to the existing Models on Alzheimer's Dataset

Models	TruePositive	Accuracy	Error Rate
PCA Ensemble	0.823	0.853	0.199
PSO Ensemble	0.893	0.9024	0.196
Decision tree based Ensemble	0.902	0.9132	0.173
Multi-feature Ensemble Decision tree	0.9472	0.9537	0.0964
Proposed Model	0.9686	0.9714	0.0479

Table 2, describes the performance analysis of the proposed model with the traditional classification models in terms of true positive rate, accuracy and error rate on Alzheimer's dataset. From the table , it is observed that proposed model has high computational accuracy in terms of accuracy , true positive rate and error rate are concerned.

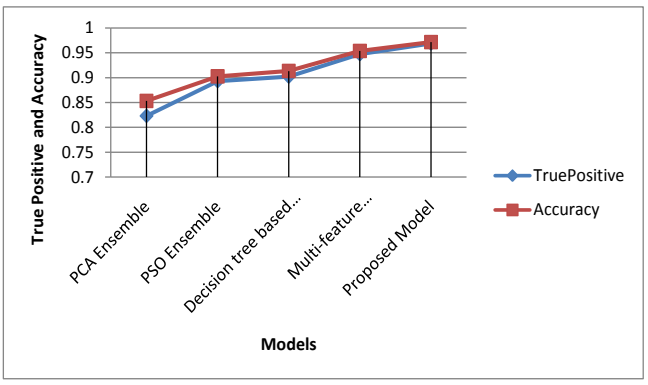

Figure 4: Graphical comparison of the classification accuracy and true positive rate on Alzheimer's Dataset**.**

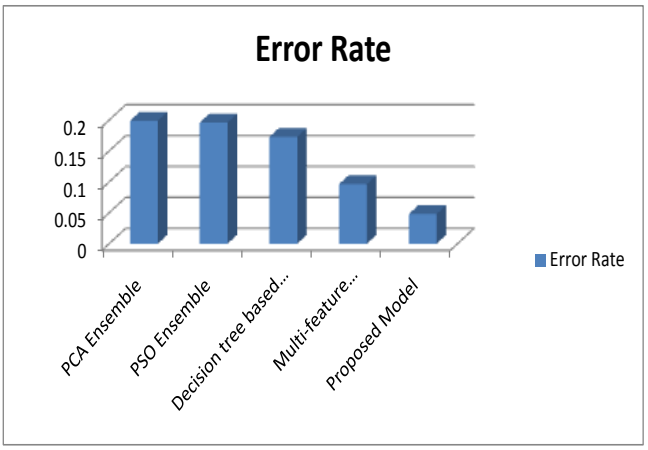

Figure 5: Graphical comparison of the classification error rate on Alzheimer's Dataset**.**

CONCLUSION

Parkinson's disease and Alzheimer's disease are considered as the most frequently found neurodegenerative diseases for disease prediction using machine learning models. Machine learning techniques are considered as the most efficient techniques for the prediction and detection of both Parkinson and Alzheimer's diseases. In this paper, a novel particle swarm optimization technique for feature subset selection was implemented on diseases datasets for ensemble classification model. In this proposed framework , feed forward neural network, hybrid decision tree technique, naïve Bayesian technique are used as majority voting base classifiers for ensemble disease classification and prediction process. Experimental results proved that proposed model has high computational accuracy and true positive rate compared to traditional feature selection measures and ensemble models.In this paper, we have performed experimental study of proposed model to the existing classification learning algorithms for disease detection and prediction. In future, this model will be extended to Hadoop based ensemble decision tree model with multiple cluster nodes.

REFERENCES

[1] G. S. Babu and S. Suresh, "Parkinson's disease prediction using gene expression – A projection based learning meta-cognitive neural classifier approach", "Expert Systems with Applications", pp.1519–1529, 2013.

[2] A. Agarwal, S. Chandrayan and S. S. Sah, "Prediction of Parkinson's Disease using Speech Signal with Extreme Learning Machine", "International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)", pp.3776-3779, 2016.

[3] Y. Chen and M. Joo, “Biomedical Diagnosis and Prediction using Parsimonious Fuzzy Neural Networks”, pp.1477-1482, 2012.

[4] M. Novotný, J. Rusz, R. Čmejla, and E. Růžička, “Automatic Evaluation of Articulatory Disorders in Parkinson’s Disease”, “IEEE/ACM transactions on audio, speech, and language processing, vol. 22, no. 9, september 2014 “pp.1366-1378, 2014.

[5] Z. A. Bakar, D. I. Ispawi, N. F. Ibrahim and N. Md. Tahir, “Classification of Parkinson’s disease Based on Multilayer Perceptrons (MLPs) Neural Network and ANOVA as a Feature Extraction”, “2012 IEEE 8th International Colloquium on Signal Processing and its Applications”, pp.63-67, 2012.

[6] A. Bayestehtashk, M. Asgari, I. Shafran and J. Mc Names, “Fully automated assessment of the severity of Parkinson’s diseasefrom speech”, “Computer Speech and Language”, pp.1-14, 2013.

[7] B. R. Brewer, S. Pradhan, G. Carvell, and A. Delitto, “Application of Modified Regression Techniques to a Quantitative Assessment for the Motor Signs of Parkinson’s Disease”, “IEEE transactions on neural systems and rehabilitation engineering, vol. 17, no. 6, december 2009”, pp.568-575, 2009.

[8] K. N. Challa, V. S. Pagolu, G. Panda and B. Majhi, “An Improved Approach for Prediction of Parkinson’s Disease using Machine Learning Techniques”, “International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016 “pp.1446-1451, 2016.

[9] P. Drot´ar, J. Mekyska, I. Rektorov´a, L. Masarov´a, Z. Sm´ekal, and M. Faundez-Zanuy, “Decision support framework for Parkinson’s disease based on novel handwriting markers”, “IEEE Transactions on Neural Systems and Rehabilitation Engineering”, pp. 1-8, 2012.

[10] M. Hariharan, K. Polat and R. Sindhu, “A new hybrid intelligent system for accuratedetection of Parkinson’s disease”, “Computer Methods and Programs in bio-Medicine “, pp.1-10, 2014.

[11] L. Huang, Y. Jin, Y. Gao, K. Thung and D. Shen, “Longitudinal Clinical Score Prediction in Alzheimer’s Disease with Soft-Split Sparse Regression Based Random Forest”, “Neurobiology of Aging”, pp.1-37, 2016.

[12] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis and M. Arnaoutoglou, “Machine learning-based classification of simple drawing movementsin Parkinson’s disease”, “Biomedical signal Processing and Control”, pp. 174-180, 2017.

[13] S. Lahmiri, “Parkinson’s disease detection based on dysphonia Measurements”,pp. 98-105, 2017.

[14] Y. Miao, H. Jiang, H. Liu and Y. Yao, “An Alzheimers disease related genes identiÞcation method based on multiple classiÞer integration”, “Computer Methods and Programs in Biomedicine”, pp. 1-15, 2017.

[15] M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi and M. Farahmand, “A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques”, “Biocybernetics and biomedical engineering”, pp.1-15, 2016.

[16] A. Ozcift and A. Gulten, “Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms”, “Computer Methods and Programs in Biomedicine”, pp.443-451, 2011.

[17] R. Prashanth, S. D. Roy, P. K. Mandal and S. Ghosh, “High-Accuracy Detection of Early Parkinson’s Disease throughMultimodal Features and Machine Learning”, “International Journal of Medical Informatics 90 (2016)”, pp.13–21, 2016.

[18] S. L. Smith, M. A. Lones, M. Bedder, J. E. Alty, J. Cosgrove, R. J. Maguire, M. E. Pownall, D. Ivanoiu, C. Lyle, A. Cording and C. J. Elliott, “Computational approaches for understanding the diagnosis and treatment of Parkinson’s disease”, “Computational Models and Methods in Systems Biology and Medicine”, pp. 226-233, 2015.

[19] E. E. Tripoliti, D. I. Fotiadis and G. Manis, “Automated Diagnosis of Diseases Based on Classification: Dynamic Determination of the Number of Trees in Random Forests Algorithm”, “IEEE transactions on information technology in biomedicine, vol. 16, no. 4, july 2012”, pp. 615-622, 2012.

[20] W. Zeng, F. Liu, Q. Wang, Y. Wang, L. Ma and Y. Zhang, “Parkinson’s disease classification using gait analysis via deterministic learning”, “Neuroscience Letters 633 (2016)”, pp.268–278, 2016.

[21] Bala Brahmeswara Kadaru,Raja Srinivas Reddy.B, ”A Novel Ensemble decision tree classifier using hybrid feadture selection measures for Parkinson’s disease prediction”, International Journal of Data Science.(InPress).