Expert System for Diagnosing Parkinson Disease Using Two Stage Feature Selection Algorithms

J. Jayashree[1], J. Vijayashree[2], and N.Ch.S.N. Iyengar[3*]

[1,2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
[3]Department of Information Technology, SNIST, Hyderabad, Telangana, India
E-mail: [1]jayashree.j@vit.ac.in, [2]vijayashree.j@vit.ac.in, and [3]srimannarayanach@sreenidhi.edu.in

Abstract— Parkinson's syndrome is a common issue with mass calculation in public health. Machine-based technology issued to distinguish between those with healthy dementia and those with Parkinson's disease. In this paper a two stage feature selection is applied for selecting the best features. In the first stage, correlation feature selection is applied and in the second stage, the selected features are then given to Particle Swarm Optimization and Ant Colony Optimization technique for feature selection. Then the selected features are used by KNN, RF, NB, SVM and MLP classifiers.

Keywords— Parkinson's disease, optimization, features, correlation.

I. INTRODUCTION

Parkinson's syndrome is a nervous system irregularity that causes movement. It's a disorder that affects the brain that regulates the activity of the body. This can get going so gradually that really you don't even even realize it. But after some time, it begins with a bit of hand shakiness that may have an effect on how you're talking, walking, dreaming, and sleeping. This can even develop when you're younger, but sometimes it doesn't happen. Parkinson's disease isn't fixed, so it can be managed to can relieve the symptoms of Parkinson's disease.

Some 6.3 million people worldwide are estimated to suffer from PD. The WHO gives a "estimated crude prevalence" of 160 per 100,000 (the total number of existing cases per year, old and new) and an estimated incidence of 16-19 per 100,000 (number of the new cases per annum). Incidence of PD is worldwide variable. That distribution, however, might not be quite as simple as a geographic (or) ethnic factor. In North America & Europe the PD is known to be more prevalent than in Asia and West Africa. Research was conducted to know the causes of PD and the use of medicinal herbs for its cure, treatment and prevention.

A. Parkinson Disease Stages

It is a chronic disorder of Parkinson's disease. This means the effects of the disease typically worsen over time. Most doctors use the scale of Hoehn and Yahr to explain the stages. The scale breaks down conditions into five stages and lets medical providers know how the illness signs and effects are developing.

B. Statistics

Parkinson's disease is the second most frequently diagnosed age-linked nervous system condition. An report indicates about 10 million individuals suffering from this illness worldwide.

During the fourth decade of life, for those who are 80 and older, the incidence of the disorder from 41 persons per 100000 to more than 1900 persons for every 100000.

Speaking of, the prevalence of this condition, or occurrence of the newly diagnosed cases, decreases with the age, while it can reach 80-plus in humans. About 4% of these individuals are diagnosed at 50 years of age.

Because of the large prescription costs associated with condition, the condition takes the toll on patient quality of the life by making socialization too difficult& adding a massive financial burden.

Table 1: Parkinsons Disease Stages

STAGES	SYMPTOMS SCALE	REGULAR ACTIVITIES	DESCRIPTION	IMAGES
Stage-one	Mild	Do not Interfere	Just tremor and other symptoms of motion occur in one side of the body. There will be shifts in sitting, stance and the facial expressions.	

Stage-two	Getting Worse	Difficult and Lengthy	Tremor, rigidity & other symptoms ofmotion affect both parts of body. The Walking issues and bad coordination can be obvious.	
Stage-three	Middle-Stage	Impairing the Activities like eating & dressing.	The lack of control & the slow motion. Falls are highly common.	
Stage-four	Severe and Limiting	Person wants support with day-to-day tasks and cannot survive alone	It is possible to stand without support but walking will require walker.	
Stage-five	Most Advance and Degrading Stage	Person requires wheelchair. Both things warrant round the clock health treatment.	Steadiness of legs can make it difficult to walk or to stand. People can suffer hallucinations (false beliefs) and delusions.	

Table 2: Statistics of PD

Prevalence Rate of PD				
INTERNATIONAL			**NATIONAL**	
US	Tens of the thousands of Americans diagnosis every year		**Kashmir**	14.1 per 100000
UK	One in 500 people living with the disease.		**Mumbai**	328 per 100000
Canada	>100,000		**Bangalore**	27.1 per 100000
Italy	407/ 100000		**Kolkata**	45.82 per 100000
Finland	>15000		**Bengal**	16.1/100,000 from rural

II. RELATED WORK

1. The early identification of any illness is an important factor. It allows the patient to be well ahead of care. In this paper the authors tried to develop a method that will aid doctors in the medical detection. The research in this paper posts various methods of equivocal manner and methods of pattern identification to make use of a set of vocal signals to help detect this disorder. This research aims to test the soundness of applying such classifiers in the Parkinson Dataset. This dataset is composed of 11 specific value range attributes.

2. Over the last few years Artificial Neural Network's have been highly popular in every kind of predictive issues. The prediction of tested individuals 'PD, multi feed-forward Artificial Neural Network's with different arrangements are being used based on the extracted characteristics from 26 voice samples for each individual.

3. Data mining provides tremendous potential for medicinal development in relation to disease prognosis, diagnosis and remedy. PD is a highly dreaded disease affecting elderly citizens tending to show unusual therapy reaction at a further advance phase of the condition. Early diagnosis of this

illness will definitely aid in arresting the growing disorder, offering hope to various unwell minds.

4. In the paper, the authors explored the alliance among engine and non-engine effects for the first time. Here, the authors have superintended the categorization and sub-set election characteristics for estimating the 2 indices of severity, HY and CISI-PD, in terms of symptoms that are not engine-only.

5. There's been a lot of research recently on the correlation among PD and articulation loss. Broad varieties of articulation signal refinement algorithms have recently evolved keeping in mind the intention of diagnosing the seriousness of PD manifestation using articulation signals.

6. The Standardized Parkinson's Disease Rating Scale (UPDRS) is the clinical recommendation score calculating the symptom frequency of a general PD. It is actually being discouraged by an internal clinical assessment of the patient's ability to manage adequately with different tasks.

7. This research shows a successful address for developing a precise guessing model for telemonitoring Parkinson's was proposed using the ELM. This model can pinpoint PD patients with a precision of 82%. דדד

8. The definition and classification of pathological voice is still a very difficult field for the work on articulation processing. The acoustic characteristics of articulation are used to distinguish between ordinary sounds and abnormal sound. This research discusses & contrasts different methods of categorization to detect the capacity of auditory limitations to differentiate between normal sounds and affected sounds.

9. Therapy of Parkinson's manifestations hangs on the neurologist's encounters, UPDRS, and measures based on the Hoehn & Yahr scales to estimate the PD stage.

10. PD is one of the most common diseases of the nervous system affecting hundred million people worldwide. The basic objective of PD diagnosis is to gain a single distinguishing detection in order to device the finest care for all patients.

III. PROPOSED WORK

In this work feature selection is done in two steps. In the first step correlation feature selection method is applied for removing the redundant and irrelevant features from the feature set. In the second step, Particle Swarm Optimization along with Ant Colony Optimization technique is used for selection the best features. The selected feature is then given to various classifies such as KNN, NB, RF, SVM and MLP for classifying the
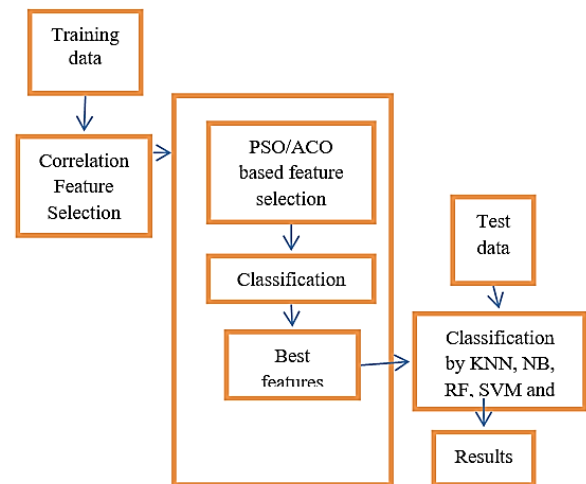
Parkinson's disease. The proposed work is shown in fig1.

Figure 1: Proposed Architecture

Data Set Information:

data used for this analysis were obtained from 188 PD patients (107 men & 81 women) aged between 33 and 87 (65±10) at the Department of Neurology at the Faculty of Medicine at the University of Istanbul in Cerrahpa lie. The target category consists of 64 stable citizens (23 men and 41 women), with ages ranging from 41 to 82 (61±8). The microphone is set to 44.1 KHz and during data collection process and the repeated vowel a phoning was obtained from each subject for three repetitions after the physician TMs test.

Attribute Information:

Numerous speech signal processing algorithms including Mel Frequency Cepstral Coefficients Time Frequency Features, Wavelet Transform Based Features and Vocal Fold features have been added to Parkinson's Disease (PD) patients' speech records to obtain clinically relevant information for PD assessment.

Random Forest:

Yet one thing to remember is that constructing the forest is not the same as setting up the decision using the information gain or index gain strategy. Random forests (or) random decision forests which are an system of ensemblelearning for classification, regression & the other activities that functions by constructing a mixture of decision trees during training time & outputting class which would be the class mode (classification) (or) regression of individual trees and the Random decision-making forests are perfect for decision trees' ability to overfit to their training set.

Naive Bayes (NB):

Naive Bayes is a simple but an effective classification technique which is based on the Bayes Theorem. It

assumes independence among predictors, i.e., the attributes or features should be not correlated to one another or should not, in anyway, be related to each other. Even if there is dependency, still all these features or attributes independently contribute to the probability and that is why it is called Naïve.

Support Vector Machine (SVM):

The aim of (SVM)support vector machine algorithm is to find a hyper-plane in an N-dimensional space which separately classifies the distinct data points.

SVM's are supervised learning models in machine learning which are associated learning algorithms which evaluate data which are used for the classification& regression analysis. Based on a collection of training instances, each of which is classified as belonging to either one or both of two categories, a Svm classification algorithm creates model which assigns the training

examples to one or both category, rendering it a non – probability binary linear classification (though methods like Platt scaling occur to use Support vector machine in classification). The Support vector machine model is a description of the instances as spatial points, mapped in such a way as to split the examples of the different groups by as large a simple distance as possible.

K-Nearest Neighbour (KNN):

K-Nearest Neighbour technique is one of the most elementary but very effective classification techniques. It makes no assumptions about the data and is generally be used for classification tasks when there is very less or no prior knowledge about the data distribution. This algorithm involves finding the k nearest data points in the training set to the data point for which a target value is unavailable and assigning the average value of the found data points to it.

IV. RESULTS AND DISCUSSION

The result of the proposed work should be evaluated using the following metrics:

$$Sensitivity = \frac{True\ Positive}{(True\ Positive + Falsee\ Negative)} * 100$$

$$Accuracy = \frac{Number\ of\ instances\ classifed\ correctly}{total\ number\ of\ instances} * 100$$

$$Specificity = \frac{True\ Positive}{(True\ Positive + False\ Positive)} * 100$$

Table 3: Performance of various classifiers

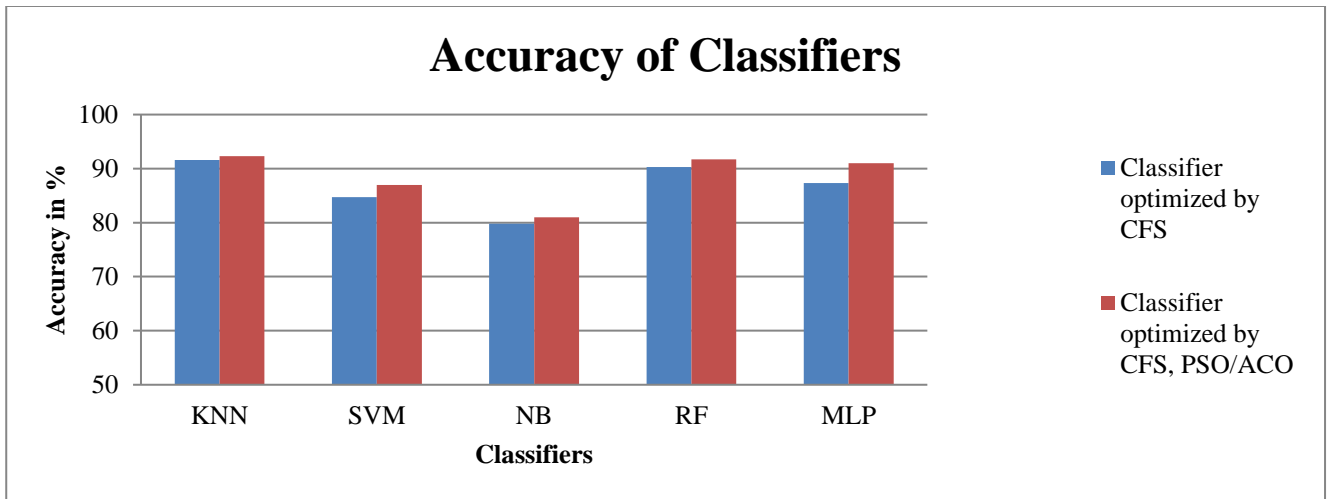| Evaluation Criteria	Classifiers	Accuracy	Sensitivity	Specificity
Classifier optimized by correlation feature selection	KNN	91.6	90.4	88.1
	SVM	84.7	81.7	79.8
	NB	79.8	76.5	75.9
	RF	90.3	88.9	86.3
	MLP	87.3	84.6	82.8
Classifier optimized by Correlation based feature selection and PSO / ACO	KNN	92.3	91.3	89.7
	SVM	87	84.7	83.7
	NB	81	78.5	76.2
	RF	91.7	90.8	88.6
	MLP	91	89.6	85.7

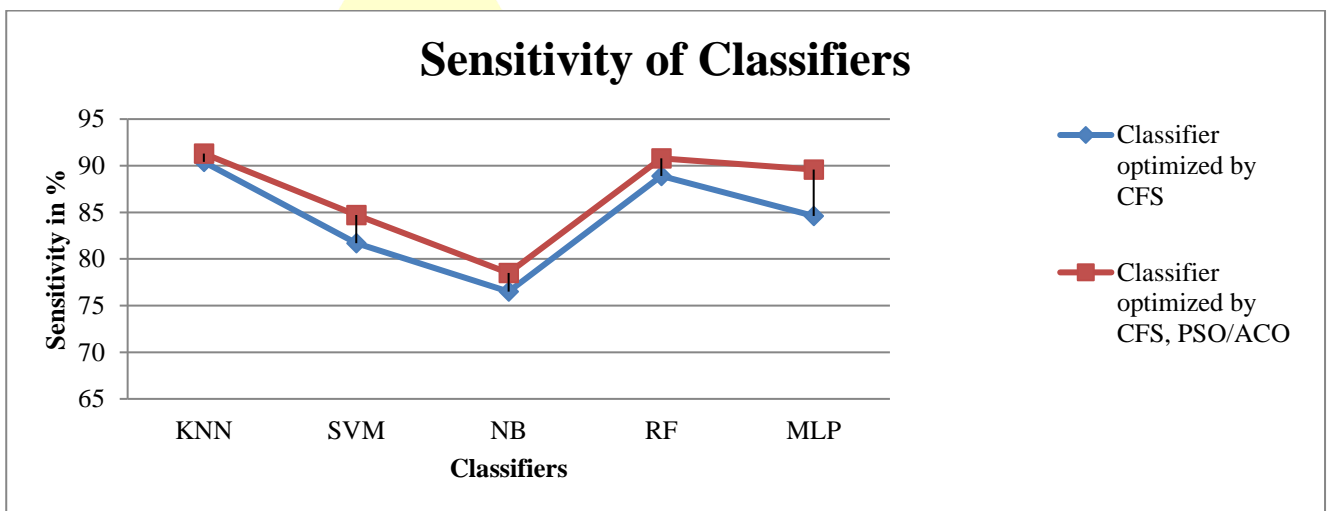*Figure 2: Accuracy of classifiers*

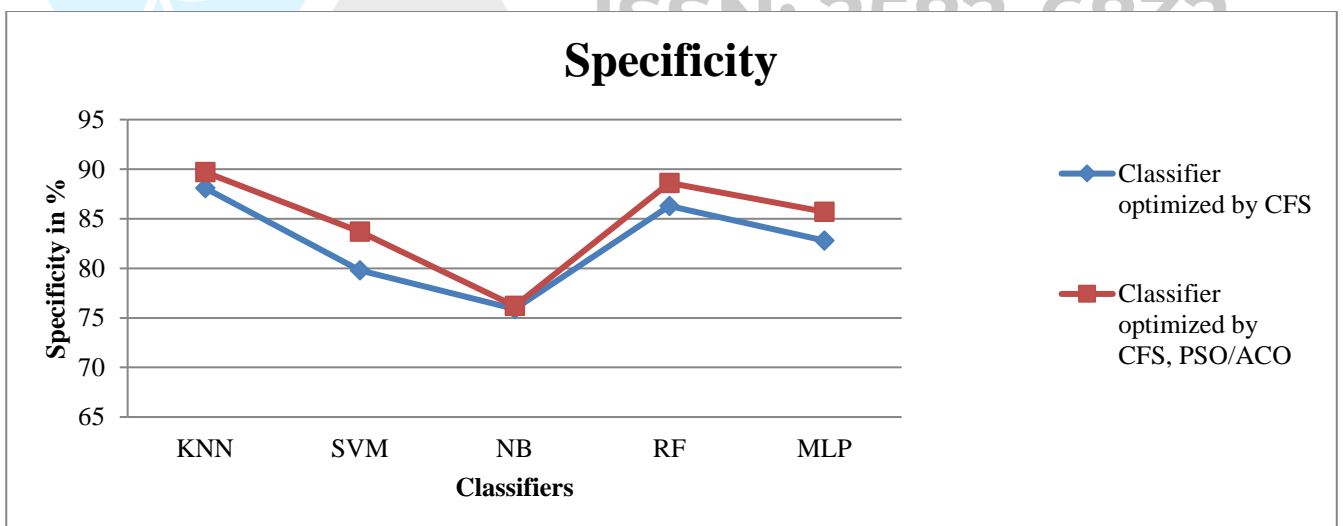*Figure 3: Sensitivity of clssifiers*

Figure 4: Specificity of classifiers

V. CONCLUSION AND SUMMARY

In this research work, analyze the automatic prediction of parkinson's disease prediction from UCI data set to improve the overall disease prediction accuracy. During the parkinson's data recognition process, data is collected from patient using by optimized machine learning approaches. According to the discussions, parkinson's disease is predicted with the help of correlation feature selection based various classifier algorithms (Random Forest, NB, SVM, KNN and MLP

classifier) and CFS wit PSO/ACO feature selection based various classifier algorithms (Random Forest, NB, SVM, KNN and MLP). The result shows that feature selection using CFS along with PSO and ACO gives better results when compared to feature selection with CFS alone. Also the KNN classifier outperforms the other classifiers such as RF, NB, SVM and MLP in predicting the Parkinson's disease.

REFERENCES

[1] Rustempasic, Indira, and Mehmet Can. "Diagnosis of parkinson's disease using fuzzy c-means clustering and pattern recognition."Southeast Europe Journal of Soft Computing 2.1 (2013).

[2] Berus, Lucijano, et al. "Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks."Sensors 19.1 (2019): 16.

[3] Ramani, R. Geetha, G. Sivagami, and ShomonaGracia Jacob. "Feature relevance analysis and classification of parkinson disease tele-monitoring data through data mining techniques."International Journal of Advanced Research in Computer Science and Software Engineering 2.3 (2012)

[4] Armañanzas, Rubén, et al. "Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach."Artificial intelligence in medicine 58.3 (2013): 195-202.

[5] Tsanas, Athanasios, et al. "Novel speech signal processing algorithms for high-accuracy classification of Parkinson" s." (2011).

[6] Tsanas, Athanasios, et al. "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity."Journal of the royal society interface 8.59 (2011): 842-855.

[7] Agarwal, Aarushi, SprihaChandrayan, and Sitanshu S. Sahu. "Prediction of Parkinson's disease using speech signal with Extreme Learning Machine."2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE, 2016.

[8] Sellam, V., and J. Jagadeesan. "Classification of normal and pathological voice using SVM and RBFNN."Journal of signal and Information Processing 2014 (2014).

[9] Shukla, Amit, et al. "Classification of Postural Response in Parkinson's Patients Using Support Vector Machines."ASME 2013 Dynamic Systems and Control Conference. American Society of Mechanical Engineers Digital Collection, 2013.

[10] Salvatore, Christian, et al. "Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy."Journal of Neuroscience Methods 222 (2014): 230-237.

[11] Xiao, Hanguang. "Diagnosis of Parkinson's disease using genetic algorithm and support vector machine with acoustic characteristics."2012 5th International Conference on BioMedical Engineering and Informatics. IEEE, 2012.

[12] Gupta, Deepak, et al. "Optimized cuttlefish algorithm for diagnosis of Parkinson's disease."Cognitive systems research 52 (2018): 36-48.

[13] Bind, Shubham, et al. "A survey of machine learning based approaches for Parkinson disease prediction."International Journal of Computer Science and Information Technologies 6.2 (2015): 1648-1655.

[14] Rustempasic, Indira, and Mehmet Can. "Diagnosis of Parkinson's disease using principal component analysis and boosting committee machines."Southeast Europe journal of soft computing 2.1 (2013).

[15] Behroozi, Mahnaz, and Ashkan Sami. "A multiple-classifier framework for Parkinson's disease detection based on various vocal tests."International journal of telemedicine and applications 2016 (2016).

[16] Sharma, Aprajita, and Ram NivasGiri. "Automatic recognition of Parkinson's Disease via artificial neural network and support vector machine."International Journal of Innovative Technology and Exploring Engineering (IJITEE) 4.3 (2014): 2278-3075.

[17] Sonu, S. R., et al. "Prediction of Parkinson's disease using data mining."2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). IEEE, 2017.

[18] Ma, Chao, et al. "An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach."Computational and mathematical methods in medicine 2014 (2014).

[19] Prashanth, R., et al. "Parkinson's disease detection using olfactory loss and REM sleep disorder features."2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014.

[20] Przybyszewski, Andrzej W. "Applying Data Mining and Machine Learning Algorithms to predict symptom development in Parkinson's disease."AnnalesAcademiaeMedicaeSilesiensis. Vol. 68. No. 5. 2014.