International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Comparative Analysis Of The Early Detection Of Parkinson's Disease

J. Olivia Capitola

Assistant professor, Computer Application Department & Affiliated to Bharathidasan University, Bishop Heber College, Tiruchirappalli

Email: oliviacapitola.ca@bhc.edu.in

Abstract: In medical science, the data mining approach is now used to evaluate vast amounts of medical data. This study aims at examining Parkinson's diseases using the feature selection process. Parkinson's disease is a central nervous system degenerative condition that primarily affects the motor system because of dopamine loss, a chemical which transmits a message to the brain part for motion control. Parkinson's early identification is really challenging, so we Modified Whale Optimization (WOA) is used to select the significant feature from the dataset. These process can select the important and relevant data from large dataset, it help to classify the PD easily. In this study we mainly analysis the detection of Parkinson diseases by using different classifiers such as K-NN, DT, NB, GMM and K-means. We take this classifier to classify the diseases in idea of comparative study among this classifier model. In this different classifier performance parameters are measured by using different parametric calculations. By finally, conclude that the K-means classifier provide the better performance result than other classifier as accuracy of 95.64% respectively.

Keyword: Modified Whale Optimization (WOA), classifier, Parkinson diseases, MATLAB and K-NN classifier.

1. INTRODUCTION

Speaking is an important symptom of Parkinson's disease (PD), therefore, recording and automated analysis of speech signals is the calmest and maximum reasonable way to detect the disease early. [1] Investigators are focused on using this technique to learn more about the nature of the disease and how to examine its signs using data mining procedures. In 2011, he created the Ford Oxford PD Discovery dataset from the Irwin Data Mining (UCI) collection at the University of California and compared some secret collections. [2]. they found that the RF algorithm categorizes the dataset with accuracy and is 100% accurate. R. Arfi Shirvan and E. Hami showed the FS using the genetic algorithm and used the K-NN algorithm for the classification [3] the top accuracy obtained with the top 9 features was 98.2%. In 201 In FS using genetic procedures was applied to the Parkinson's dataset and [4] was used for SVM classification. The uppermost accuracy attained with the top 4 functions was 94.5%. In 2014, Harinish, Gracie Annamari [5] associated the presentation of random plants, M5 rules and ANN algorithms. Demonstrates the highest accuracy of the ANN algorithm.
In, all irrelevant parameters and outlets in the OPD dataset were removed during the study. Comparing the performance of K-NN, Random Forest K-NN found the highest accuracy to be 90.26%. According to the study, the features were selected in 2015 and the datasets were

classified using SVM [6]. The maximum accuracy obtained was 100%. In a second study in 2012, two additional target classes were added to Parkinson's dataset [7]. The main features were then screened using main component analysis (PCA) and comparatively different classification rates. Automatic minimum SVM optimization provides maximum ura %%% accuracy. In a second study in 2012, Parkinson's datasets were first generalized, followed by analysis of various methods for selecting and classifying features [8]. It was found that 98.97% accuracy was obtained using the wrapper selection method and K-NN classification. P. Kriparapun and s. United additional NN and more effectual classification of OPD datasets for generalization, amorphous [9]. P. Secundo Durga, V. Soot Jebkumari and D. Mir, who classified the OPD dataset, showed that MLP has the highest 97.78% purity compared to 48 and NB algorithms [10].

2. LITERATURE SURVEY

Zahari Abu Bakar et.al [11] The analysis is based on two algorithms. These are neural networks from Lewenbergmarkart (LM) and SCG Multi Layer Perceptron (MLP) in PDI to detect partial discharges. The data in this project is generated from the PD data record. The LM algorithm was found to have achieved better training accuracy and test accuracy with 25 hidden modules compared to other hidden modules, but 97.86 percent during the learning phase and 92.96 percent during the experimental phase. LM. Did a good job with a classification accuracy of 92.95% and SKG achieved an accuracy of 78.21%.

Hausdorff et al. [12] focussed on the dynamic of the gait, which assesses the impact of RAS, the application of musical stimulation in order to increase the gait output of neurological persons. It has been shown that RAS facilitates more automated motion and decreases step-by-step variability in PD topics.

Joshi et al.[13] proposed a technique that merged wavelet analysis with SVM to separate Parkinson from stable subjects using the variability in the gait period. The findings revealed that the approach to wavelet transformation contributed to a grade rate of 90%.

Dr. R. Geetha Ramani et.al [14] A proposal was made to classify patients with PD and non-PD using the following approaches: linear differential analysis (LDA), RND tree, and SVM. The dataset contains PD data from the UCI collection. The training data set contained 197 samples from 22 characteristics of patients. The Fisher filter function selection algorithm has proven to be an efficient system for classifying features. The edge tree algorithm achieved a classification accuracy of 100%, while LDA, C4.5, CS-Mac4 and K-NN achieved results with an accuracy of over 90%. The C-plus algorithm achieved the lowest accuracy of 69.74%.

Maryi et al. [15] introduced a system focused on wearable on-shoe sensors and a processing algorithm for assessing PD symptoms of timed up and moving (TUG) agility and gait checking. The research used the following spatio-temporal parameters: swing distance, turning, route longitude and uncertainty between periods.

3. PROPOSED SYSTEM

In this paper, for early diagnosis of PD a novel approach is introduced which is defined in the fig.1. The proposed method consist of three major blocks such as pre-processing, feature selection and classification. In the first stage the data set data’s are readied and the null values are removed in the preprocessing stage. To improve the classification accuracy in this paper a Modified Whale Optimization (WOA) FS technique is used and finally, based given classifier early diagnosis has been predicted.

3.1. Data Source

The dataset used in this research experiment includes features obtained from speech signals from 31 people at the Voice Speech Center in Max Little of Oxford University created the dataset and presented it at the UCI ML Repository. Of the 31, 23 are PDs and 8 are control groups. The dataset contains 195 biomedical voice measurements. Table 1 displays the voice solutions used in the experimentations. The status in the file describes the class of the column and gets 0 for health, 1 for PD. The distribution of the classes of the dataset is exposed in Figure 2. There are 48 healthy phonetic and 147 phonological PDs for 48 people.

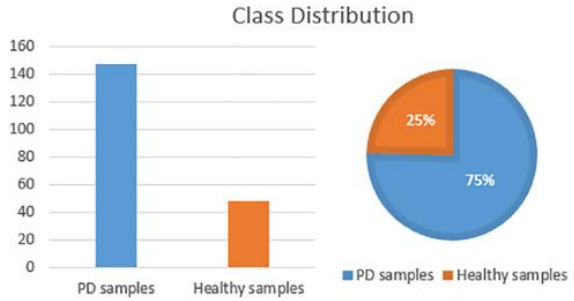

Fig.2. Recorded phonetics Class distribution in the dataset.

Feature no	Voice measure	MEANING
1	MDVP:Fo (Hz)	Average frequency
2	MDVP:Fhi (Hz)	Maximum frequency
3	MDVP:Flo (Hz)	Minimum frequency
4	MDVP:Jitter (%)	Numerous measures of difference in
5	MDVP:Jitter (Abs)	fundamental frequency
6	MDVP:Shimer	Numerous measures of variation in amplitude
7	MDVP:Shimer (dB)	
8	Shimer:APQ3	
9	Shimer:APQ5	
10	MDVP:APQ	
11	Shimmer:DDA	
12	NHR	Two ratio measures of noise to tonal
13	HNR	components in the voice
14	RPDE	Two nonlinear dynamical difficulty
15	D2	measures

3.2. Feature selection

The definition of brightness is also indicated by the selection of an element, the selection of an attribute, or the selection of a subset of variables for the development of the model, which makes it difficult to select a subset of relevant key points. In this projected structure, the Modified Whale Optimization (WOA) algorithm is used to determine the inclusion.

3.2.1. MWO Algorithm feature selection

In this segment, the WOA changes the benchmark to accommodate other kinds of methods. Three changes proposed and detailed in the MWOA.

A key problem for large scale worldwide enhancement (LSGO) cover by metaheuristic computing (MAS) is that most of them are rapidly converging in the direction of the optimal

neighborhood due to the rapid decrease of differential diversity, and the first WOA is not a superior case. In previous studies, the Levy flight course has been widely used in MA to prevent the close agreement of Optima and accelerate integration in light of worldwide hunting productivity. Therefore, levy flight is used to escape near-optimal at MWOA, which differentiates population diversity..

The Lévy flight is a sort of non-Gaussian randompractice with step length subsequent a Lévyassumption. Anupfront power-law vision of the Lévy conveyance is:

$$L(s) \sim |s|^{-1-\beta}, 0 < \beta \leq 2 \quad (1)$$

Where β an index, s is is the step length of the Lévy flight. Mantegna's procedure is applied to calculating

$$s = \mu/|\vartheta|^{1/\beta} \quad (2)$$

Where, μ And ϑ obey normal distribution, i.e.

$$\mu \sim N\left(0, \sigma_\mu^2\right), \vartheta \sim N\left(0, \sigma_\mu^2\right) \quad (3)$$

$$\sigma_\mu = \left[\frac{\tau(1+\beta).\sin(\pi.\beta/2)}{\tau\left(\frac{1+\beta}{2}\right).\beta.2^{\frac{(\beta-1)}{2}}}\right]^{1/\beta} \quad (4)$$

$$\sigma_\vartheta = 1 \quad (5)$$

A step size avoiding the Lévy flight leaping out of the design field is adopted. It is defined by:

$$Levy = random\left(size(D)\right) \oplus L(\beta) \sim \frac{0.01\mu}{|v|^{\frac{1}{\beta}}(X_i - X^*)} \quad (6)$$

If dimension (D) is the scale of the problem,$\oplus$ it indicates the initial multiplication, Xi is the ^{i}th vector of the solution. Due to the unlimited fluctuations in the circulation of the levy, the levy flight sometimes does the development of a long separation to increase research capacity, while the development of a short separation is done to increase performance. Obviously, this legality can guarantee that MA will recover to nearby Optima. At MWOA, the procurement tool is replaced by Levy's trip to discover the research space more and more skillfully. The novel location is updated in the same way.

$$X(t+1) = X(t) + \frac{1}{sqrt(t)}.sign(rand - 0.5) \oplus Levy \quad (7)$$

Where 1 / sqrt (t) is the factor associated with the in progress iteration number$t$, and $sqrt$ () is the sqrtprocess. In this regard, an earlier search may be performed at an earlier stage, while a slighter one is used in a later passé. $Sign\ (rand - 0.5)$signifies a sign function with only three values -1, 0, 1, which makes the search additional random. The MWOA exploration phase is summarized as follows:

$$X(t+1) = \begin{cases} X(t) + \dfrac{1}{sqrt(t)}.sign(rand - 0.5) \oplus Levy & if\ p < 0.5 \\ D'.e^{bl}\cos(2\pi l) + X^*(t) & if\ p \geq 0.5 \end{cases} \quad (8)$$

Classification Techniques

The classification strategies utilised in this analysis are briefly defined in this portion.

- K-NN is one of the simplest controlled approaches to classification. It is a method of classification which is not supervised by parameters. In K-NN, prior to the classification stage no overt or modelling process exists. K-NN grouping includes two key stages: (1) a distance measurement is rendered between the current sample and all

training samples (usually Euclidean distance), and (2) a new sample is allocated with the most class of the next samples using a range from K next to the next neighbour.

- · The vector support method is a popular supervised learning machine model used mainly for binary prediction issues. The theory behind this model is based on the theories of a hyperplane and its perimeter. The learning process is to find a linear separator (or hyperplane) that separates the training data while maximising the distance between the hyperplane and those training data. The training process consists of In some situations, in their original representation, SVM cannot automatically find a linear division between the data. Thus a training data transition suggested by Vapnik from its original domain into another higher dimensional space is carried out to find a linear seperator among the groups. The kernel function including the gaussian, quadratic or polynomial kernel functions may be added to render this transition.
- · The decision tree is a simple, accurate and easy-to-interpret supervised classification process. A DT considers non-linear connections among the device inputs and outputs. A DT is an iterative grader which divides variables into branches and nodes. The nodes consist of a root node and numerous inertial nodes and leaves.
- · The random forest is another controlled Breiman machine learning. A random forest is built from a set of DTs, as its name implies. Each tree is built by a randomly generated training subset using the Bootstrap technique from the original dataset. The randomised collection from the partitioning of the data nodes during dt creation is therefore merged in the RF model.
- The NB is another simple monitoring machine model based on the Theorem of Bayes with assumptions of independence between observation data. The key benefit of NB is that its models are simple, without the need for a complicated calculation of the iterative parameter. The NB model will execute more complex machine learning models, considering its simplicity.
- THE GAUSEAN MIXTER MODEL is a controlled and unattended model of probabilistic learning. This model presents the data of the training as finite Gaussian densities for weighted sums. The data are seen by an or multi-Gussian distribution and defined by the covariance and mean vector matrix. The parameter evaluation for this model is focused on log probabilities maximised using the expectation-maximization (EM) algorithm (the covariance matrices of the Gauzan variable and the mean vectors).
- Towards K imply another basic machine learning model is still unregulated. The training data is grouped into k-means clusters. The goal is to minimise the overall gap in the intracluster and the cost estimation. The K-means model defines and assigns the data iteratively to the different cluster centres depending on their distance (e.g. Euclidean), before a convergence takes place.

4. RESULTS AND DISCUSSIONS

To evaluate the various performance metrics such as accuracy, class error, specificity, and sensitivity were used to evaluate the presentation of the classifier. The time required to implement the model was also measured. All calculations were done in Python using an Intel (R) Core ™ i5-2400CPU at 3.10 GHz. The main results of the projected research are the following:

4.1. Performance Evaluation Metrics

In this study, various presentation metrics were used to test the effectiveness of classification. We used a random matrix, evaluating each observation in the test sample, even in one field. As a repo class of 2, it is a 2 × 2 matrix. In addition, it offers two types of accurate predictions and two types of false predictions. Table 3 presents the confusion matrix.

Table 3: Confusion matrix.

	Projected HD patient (1)	Predicted healthy person (0)
Real HD patient (1)	TP	FN
Actual healthy person (0)	FP	TN

Classification accuracy: accuracy indications that the complete performance of the classification system as follows

$$classification\ Accracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (9)$$

Classification error: it is the complete improper classification of the classification ideal which is considered as follows:

$$\text{Classification error} = \frac{FP + FN}{TP + TN + FP + FN} \times 100 \quad (10)$$

Sensitivity: This is the proportion of newly confidential patients with heart disease out of the total number of patients with heart disease.

$$\text{recall or true positive rate} = \frac{TP}{TP + FN} \times 100 \quad (11)$$

Specificity: an analytical test is negative and the being is healthy and is accurately written as follows:

$$\text{specificity(Sp)} = \frac{TN}{TN+FP} \times 100 \quad (12)$$

Precision: the equation is given as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 1 \quad (13)$$

Classification Performance

In this unit we discussed the performance evaluation of the some previous existing scheme with our proposed method. In below table 2 shows the presentation of the different models outcomes. K-NN, DT, RF, SVM, NB, GMM and K-means

Table.2. Classification Performance.

Method	**SE (%)**	**SP (%)**	**AC (%)**
K-NN	72.25	83.24	66.52
DT	100	94.65	64.35
RF	98.47	96.58	65.35
SVM	97.25	96.67	66.47
NB	96.47	97.23	67.80
GMM	94.65	97.56	70.46
K-means	99.31	98.87	95.64

In this figure 2 and table 2 signifies that the performance measures of different classifier. Initially, K-NN classifier achieved low parametric value of 66.52% accuracy and DT classifier achieved the accuracy of 64.35%, it is better than K-NN classifier. RF classifier achieved the accuracy of 65.35% and SVM classifier achieved the accuracy of 66.47%. And also K-mean classifier attained the better classification value of 95.64%. By this comparisons, K-means attained the better accuracy results than other classifier model.

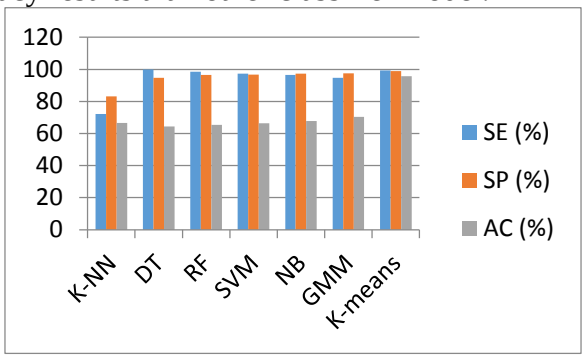

Fig.2.Classification Performance.

5. CONCLUSION

Parkinson's disease is a central nervous system degenerative condition that primarily disturbs the motor system because of dopamine loss, a chemical which transmits a message to the brain part for motion control. Parkinson's early identification is really challenging, so in this study aims at examining Parkinson's diseases using the feature selection process by using the Modified Whale Optimization (WOA) to select the important feature from the large data's, it help to classify the PD easily. In this study we mainly analysis the detection of Parkinson diseases by using different classifiers such as K-NN, DT, NB, GMM and K-means. We take this classifier to classify the diseases in idea of comparative study among this classifier model. In this different classifier performance parameters are measured by using different parametric calculations. By finally, conclude that the K-means classifier provide the better performance result than other classifier as accuracy of 95.64% respectively than the other classifier models.

6. REFERENCES

[1] "Parkinson's Foundation: Better Lives. Together.".
[2] R. G. Ramani, "Parkinson Disease Classification using Data Mining Algorithms," vol. 32, no. 9, pp. 17–22, 2011.
[3] R. A. Shirvan and E. Tahami, "Voice Analysis for Detecting Parkinson's Disease Using Genetic Algorithm and KNN Classification Method," no. December, pp. 14–16, 2011.
[4] M. Shahbakhi, D. T. Far, E. Tahami, and M. Shahbakhi, "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine," J. Biomed. Sci. Eng., vol. 7, no. 7, pp. 147–156, 2014.
[5] S. Hariganesh and G. Annamary, "A Survey of Parkinson's Disease Using Data Mining Alogorithms," Int. J. Comput. Sci. Inf. Technol., vol. 5, 2014.
[6] Saloni, R. K. Sharma, and A. K. Gupta, "Detection of Parkinson Disease Using Clinical Voice Data Mining," vol. 9, pp. 320–326, 2015.

[7] Caesarendra W, Putri FT, Ariyanto M, Setiawan JD. Pattern recognition methods for multi stage classification of parkinson's disease utilizing voice features. In2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM) 2015 Jul 7 (pp. 802-807). IEEE.
[8] M. S. Wibawa, H. A. Nugroho, N. A. Setiawan, J. Grafika, and Y. Indonesia, "Performance Evaluation of Combined Feature Selection and Classification Methods in Diagnosing Parkinson Disease Based on Voice Feature," Int. Conf. Sci. Inf. Technol., pp. 126– 131, 2015.
[9] Kraipeerapun P, Amornsamankul S. Using stacked generalization and complementary neural networks to predict Parkinson's disease. In2015 11th International Conference on Natural Computation (ICNC) 2015 Aug 15 (pp. 1290-1294). IEEE.
[10] P. Durga, V. S. Jebakumari, and D. Shanthi, "Diagnosis and Classification of Parkinsons Disease Using Data Mining Techniques," ISSNOnline) Int. J. Adv. Res. Trends Eng. Technol., vol. 3, no. 14, pp. 2394–3777, 2016.
[11] Zahari Abu Bakar, Nooritawati Md Tahir, Ihsan M Yassin, "Classification Of Parkinson's Disease Based On Multilayer Perceptrons Neural Network', Ieee Colloquium In Signal Processing And Its Applications (Cspa), 2010.
[12] Hausdorff, J.M.; Lowenthal, J.; Herman, T.; Gruendlinger, L.; Peretz, C.; Giladi, N. Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. Eur. J. Neurosci. 2007, 26, 2369–2375.
[13] Joshi, D.; Khajuria, A.; Joshi, P. An automatic non-invasive method for Parkinson's disease classification. Comput. Methods Programs Biomed. 2017, 145, 135–145.
[14] Dr. Geetha R Ramani, G Sivagami,"Parkinson Disease Classification Using Data Mining Algorithms", International Journal Of Computer Applications 32(9):17-22, October 2011.
[15] Mariani, B.; Jiménez, M.C.; Vingerhoets, F.J.; Aminian, K. On-shoe wearable sensors for gait and turning assessment of patients with Parkinson's disease. IEEE Trans. Biomed. Eng. 2013, 6, 155–158.