ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 6 , June 2021

Classification Model for Predicting the Risk of Parkinson Disease: A Comparative Analysis of Radial Basis Function and Logistic Regression Classifiers

Oyegoke T.O,Amoo A. O, Aramide O. T, E. R. Adagunodo

Lecturer, Department of Computer Science & Engineering, Obafemi Awolowo University, Ile – Ife. Nigeria
Lecturer, Department of Computer Science & Engineering, Obafemi Awolowo University, Ile – Ife. Nigeria
Student, Department of Computer Science & Engineering, Obafemi Awolowo University, Ile – Ife. Nigeria
Professor, Department of Computer Science & Engineering, Obafemi Awolowo University, Ile – Ife. Nigeria

ABSTRACT: This study presents a predictive model for classifying the risk of Parkinson Disease (PD) using machine learning based approaches. A Particle Swarm Optimization (PSO) Algorithm was employed for feature selection of the most relevant variables that can influence PD risk. This was done in a view to identify the key acoustic parameters for PD risk. Two supervised Machine Learning algorithms (Radial Basis Function and Logistic Regression) were used for the classification of the PD risk (Risk or No-Risk). The model was simulated and validated using Python programming language; and Spyder was used as the Integrated Development Environment (IDE) for the simulation. This model for Parkinson disease risk will thus help medical personnel and patients knows early enough the possibility of having Parkinson disease so as to commence preventive care.

KEYWORDS: Parkinson disease, classification, radial basis function, linear regression, particle swarm optimization, machine learning, prediction.

1. INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease, affecting more than one million people in North America and about 2% of the population over the age of 65 years [1]. It is a disease which affects the part of the brain that controls the movement of the whole body. It can be very unnoticeable at first but continually affects the brain and in turn the body. The key abnormality in PD is the deficiency of a hormone called dopamine in the sustantia nigra which is a region in the midbrain [2]. With PD, the cells of the sustantia nigra begins to die gradually and there is no replacement for them as the dopamine level drops. The progression of the disease is formally described as the Hoehn and Yahr staging of PD from stage I to stage V [3]. Its symptoms typically include tremor (shaking) at rest, aches and pains, sleep difficulties, rigid muscles, changes in writing, changes in speech, slow movement (bradykinesia), and gait-posture-related dysfunction [4, 5].

PD is one of the major public health problems in the world. It is a well-known fact that around one million people suffer from Parkinson's disease in the United States whereas the number of people suffering from Parkinson's disease worldwide is around 5 million [6]. Depression is akey factor of the disease, and after getting PD, depression plays a major and dangerous role in PD patients. Depression is so common with this disease but is often overlooked and undertreated. It is estimated that at least 50 percent of those diagnosed with PD will experience some form of depression during their illness, and up to 40 percent will experience an anxiety disorder.[7].A very common symptom of PD is tremor. A tremor is usually a classic, slow, rhythmic and involuntary movement or shakes in one hand, leg and can eventually affect the entire body. A tremor of certain part of the body like the finger, hand, chin, etc., while at rest is a major sign of this disease.

PD cannot be cured, and the cause of Parkinson's disease is unknown, but several factors plays a role in knowing the risk of having the disease and the factors are often referred to as risk factors [8]. Some risk factors include voice, age, family history, gender, ethnicity, environmental pesticides, Head trauma, voice (speech) and lots more. The problem with Parkinson disease is that once it un-earth the body, complete cure is not possible, but its prevalence can be

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 6 , June 2021

controlled to some extent. Predicting Parkinson's disease in early stages is important so that early plan for the necessary treatment can be made. People are mostly familiar with the motor symptoms of Parkinson's disease. However, an increasing amount of research is being done to predict the Parkinson's disease from non-motor symptoms that precede the motor ones. If early and reliable prediction is possible then a patient can get a proper treatment at the right time. Developing machine learning models that can help in predicting the risk of having the disease can play a vital role in early prediction.

Machine learning (ML) is a branch of artificial intelligence that allows computers to learn from past examples of data records [9,10]. Unlike traditional explanatory statistical modeling techniques, machine learning does not rely on prior hypothesis [9]. Many researchers have proposed and developed a couple of machine learning models in predicting the risk of Parkinson and they have use a number of approaches. Some of these include; Naïve Bayes (NB), Random Forest (RF), Linear Discriminant Analysis (KNN), Logistics Regression (LR), Support Vector Machine (SVM), fuzzy logic (FL), Kernel Fisher Discriminant (KFD) and other machine learning algorithms. Machine learning has found great importance in the area of predictive modeling in medical research especially in the area of risk assessment, risk survival and risk recurrence. Its techniques can be broadly classified into: supervised and unsupervised techniques; supervised involves matching a set of input records to one out of two or more target classes while unsupervised is used to create clusters or attribute relationships from raw, non-labeled or non-classified datasets.

Different branches and techniques of machine intelligence have been and are being used in prediction of risk in health sector. Machine learning techniques applied to predicting the risk of Parkinson disease includes but not limited tree classifier, statistical classifier, and support vector machine classifier in identifying the people affected by PD [11]; [12] used standard dimensionless equations and multiple regression normalization to normalize data, Kernel Fisher Discriminant, Support vector machine and Random Forest were used in classifying the accuracy of the models. Non-linear iterative partial least squares were used for data dimensionality reduction and self-organizing map for clustering task and Incremental support vector machine was used to predict Total-UPDRS (Unified Parkinson's Disease Rating Scale) and Motor-UPDRS [13]. [14] used a hybrid of Subtractive Clustering Features Weighting (SCFW) and a Kernel-based Extreme Learning Machine (KELM) –KELM in diagnosing PD. The result showed that the proposed SCFW-KELM significantly outperforms SVM-based, KNN-based and ELM-based approaches.

Feature selection methods are unsupervised machine learning techniques used to identify relevant attributes in a dataset. While the dataset used in this research consists of different attributes, the relevant ones had to be selected using a nature inspired optimization algorithm called the Particle Swarm Algorithm (PSO). It is important in identifying irrelevant and redundant attributes that exist within a dataset which may increase computational complexity and time [15, 16]. Feature selection methods are broadly classified as filter-based, wrapper-based and embedded methods while wrapper based methods are chosen for this study.

Supervised machine learning algorithms can be used in the development of classification or regression models. Classification model is a supervised approach aimed at allocating a set of input records to a discrete target class unlike regression which allocates a set of records to a real value. This research is focused at using both the Radial Basis Function (RBF) and Logistic Regression (LR) to classify PD risk as either risk or no-risk. RBF alone was used without the option of selecting relevant features, features were selected using PSO and RBF was used in classifying, and LR was used to classify using the attributes selected by the PSO. The three models were then simulated and the results were compared.

II. SIGNIFICANCE OF THE SYSTEM

This paper addresses the problem with Parkinson disease risk by developing a model for predicting the risk level of Parkinson disease. The rest of this paper is arranged as follows: Section III discusses the literature Survey, section IV presents the Methodology, experimental result is covered in section V while the future study is discussed in section VI.

III. LITERATURE SURVEY

Parkinson's disease was first medically described as a neurological syndrome by James Parkinson in 1817 [17]. Parkinson is a progressive neurological condition, which is characterized by both motor (movement) and non-motor symptoms. Researchers estimate that one million people in the United States, and four to six million people worldwide, are living with Parkinson's. Non-motor symptoms of Parkinson disease include soft speech, especially resulting from a lack of coordination, disturbances during sleep, depression, and dementia [18].

The exact causes of Parkinson disease are not known but, researchers have identified characteristics that increase a person's risk of developing Parkinson disease. Some including gender, age, race, occupation, diet, ethnicity, head

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 6 , June 2021

trauma, environmental pesticides, Oxidative stress, family history and genetics [19, 20, 21]. However, it is worth noting that the vast majority of cases of PD are considered idiopathic Parkinson's disease. Idiopathic means a condition that arises spontaneously or for which the cause is 10 currently unknown. Major advances in research and science are continuing to reveal more underlying causes for PD.

[22] proposed classification of pathological voice from normal voice for PD risk. It was implemented using Support Vector Machine (SVM) and Radial Basis Functional Neural Network (RBFNN). The normal and pathological voices of children were used to train and test the classifiers. The speech signal was then analysed in order to extract the acoustic parameters such as the Signal Energy, pitch, formant frequencies, Mean Square Residual signal, Reflection coefficients, Jitter and Shimmer. It shows the classification accuracy of RBFNN 91% and SVM 83%. For the model seven (7) acoustic parameters were used for the predictive model.

[23]proposed speech recognized by Mel-frequency cepstral coefficients (MFCC) and Vector Quantization (VQ) for PD risk prediction. The author used MFCC for speech analysis frames in signal to frequent domain and Vector Quantization was the codebook used for calculating the lowest distortion in voices. The model has an accuracy of 90% and twenty (20) phonation's attributes were used.

[24]used an evolutionary algorithm, named estimation of distribution algorithm for the classification of the severity of staging of Parkinson's disease. The study's result show how five different classification paradigms using a wrapper feature selection scheme are capable of predicting each of the class variables with estimated accuracy in the range of 72–92%. In addition, classification into the main three severity categories (mild, moderate and severe) was split into dichotomic problems where binary classifiers perform better and select different subsets of non-motor symptoms. The number of jointly selected symptoms throughout the whole process was low, suggesting a link between the selected non-motor symptoms and the general severity of the disease.

[25] applied feature selection to choose features that have a positive effect so that the performance of the model does not decrease. The results of the study indicate that a model that integrates Decision Tree and Forward Selection provides better performance values. The experiment results show that the application of feature selection can lead to better model performance.

[26] developed a method in which an enhanced fuzzy k-nearest neighbor (FKNN) method for the early detection of PD based upon vocal measurements was used. In this study, CBFO-FKNN, was developed by coupling the chaotic bacterial foraging optimization with Gauss mutation (CBFO) approach with FKNN. The result. indicated the proposed approach outperformed the other five FKNN models based on BFO, particle swarm optimization, Genetic algorithms, fruit fly optimization, and firefly algorithm, as well as three advanced machine learning methods including Support Vector Machine (SVM), SVM with local learning-based feature selection, and kernel extreme learning machine in a 10-foldcross-validation scheme.

[27] proposed a diagnosis system using fuzzy k-nearest neighbor (FKNN) for Parkinson's disease (PD) diagnosis. The proposed FKNN-based system is compared with the support vector machines (SVM) based approaches. The effectiveness of the proposed system was estimated on a PD data set in terms of classification accuracy, sensitivity, specificity and the area under the receiver operating characteristic (ROC) curve (AUC). The experimental results showed that the FKNN-based system greatly outperforms SVM-based approaches.

[28] again proposed an hybrid intelligent system which includes feature pre-processing using Model-based clustering (Gaussian mixture model), feature reduction/selection using principal component analysis (PCA), linear discriminant analysis (LDA), sequential for-ward selection (SFS) and sequential backward selection (SBS), and classification using three supervised classifiers such as least-square support vector machine (LS-SVM), probabilistic neural network (PNN) and general regression neural network (GRNN).The experimental results showed that the combination of feature pre-processing, feature reduction/selection methods and classification gives a maximum classification accuracy of 100%for the Parkinson's dataset.

IV. METHODOLOGY

This study involves the use of supervised machine learning algorithms in the development of predictive model for PD risk using data collected from UCI Machine learning repository from Centre for Machine Learning and Intelligent Systems. Figure 1 shows the methodology framework which was applied in the development of the predictive model for PD risk.

The study began with the identification of the variables monitored during the biomedical voice recording of PD patients by Centre for Machine Learning and Intelligent Systems and the collection of the dataset containing the identified variables for patients in the study location. The dataset collected from UCI Machine learning repository formed the basis of the historical dataset which contains various records of predictive parameters.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 6 , June 2021

Wrapper-based Feature selection methods was used to identify the most relevant and important features among the features collected based on the distribution of the dataset collected from the study location. The reduced feature set was identified to be predictive for PD risk. Following this, the historical dataset containing the reduced feature set was divided into training and testing dataset and fed to each supervised machine learning algorithms proposed for this study using the percentage split evaluation method.

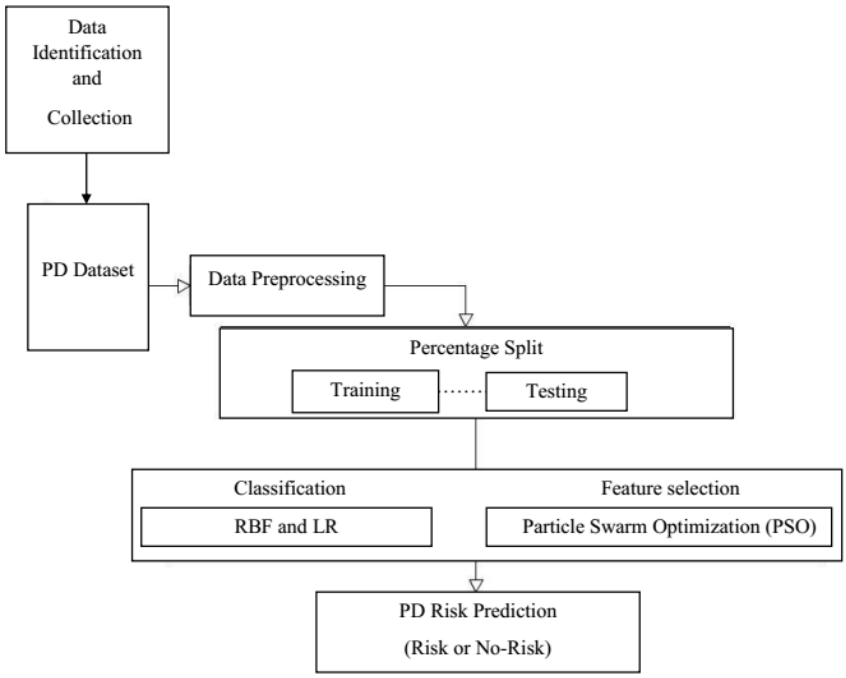

Figure 1: Proposed Predictive Model for PD Risk

The result of the performance of the combination of each wrapper-based feature selection method and the supervised machine learning algorithm was used to identify the most effective and efficient predictive model for PD risk.

Data Description

This part highlights the process involved in identifying the data containing the variables monitored. Each variable name was identified and properly defined with its respective units defined. The method of data collection was also clearly stated showing from whom the data was collected and the instruments of data collection from the data source alongside the identification of the different risk classes in the dataset. A number of features were identified to be monitored during the biomedical voice recording of PD patients. The voice dataset for Parkinson disease has been retrieved from UCI Machine learning repository from Centre for Machine Learning and Intelligent Systems. Table 1 gives a description of the variables that were identified to be monitored in biomedical voice recording.

The description of all the variables identified is discussed as follow:

i. **MDVP:Fo(Hz):** Average vocal fundamental frequency.
ii. **MDVP:Fhi(Hz):**Maximum vocal fundamental frequency.
iii. **MDVP:Flo(Hz):** Minimum vocal fundamental frequency.
iv. **MDVP:Jitter(%),MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ,Jitter:DDP**: Several measures of variation in fundamental frequency.
v. **MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA**: Several measures of variation in amplitude.
vi. **NHR, HNR:** Two measures of ratio of noise to tonal components in the voice.
vii. **RPDE, D2**: Two nonlinear dynamical complexity measures.
viii. **DFA:** Signal fractal scaling exponent.
ix. **spread1, spread2, PPE:** Three nonlinear measures of fundamental frequency variation.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 6 , June 2021

The structured data was collected from https://archive.ics.uci.edu/ml/datasets/parkinsons and it is in ASCII CSV format stored in spreadsheet format following the identification of the variables monitored during the biomedical voice recording. For the purpose of handling the problem as a classification problem, the target class (output variable) was determined using two labels, namely: risk and no-risk.

i. **Risk**: refers to the PD patients that are with the tendency of having PD.
ii. **No**-Risk: refers to those with no risk of having PD.

Table 1: Variables monitored during the biomedical voice recording

Name	Unit of Measurement	Label
MDVP:Fo	Hz	Numeric
MDVP:Fhi	Hz	Numeric
MDVP:Flo	Hz	Numeric
MDVP:Jitter	%	Numeric
MDVP:Jitter	Abs	Numeric
MDVP:RAP	Nil	Numeric
MDVP:PPQ	Nil	Numeric
Jitter:DDP	Nil	Numeric
MDVP:Shimmer	Nil	Numeric
MDVP:Shimmer	dB	Numeric
Shimmer:APQ3	Nil	Numeric
Shimmer:APQ5	Nil	Numeric
MDVP:APQ	Nil	Numeric
Shimmer:DDA	Nil	Numeric
NHR	Nil	Numeric
HNR	Nil	Numeric
RPDE	Nil	Numeric
DFA	Nil	Numeric
Spread1	Nil	Numeric
Spread2	Nil	Numeric
D2	Nil	Numeric
PPE	Nil	Numeric

Formulation of Predictive Model for PD Risk

Following the identification of the most relevant and predictive variables for PD risk prediction, the next phase is the formulation of the predictive model for PD risk using the identified variables. Mathematical expressions called mapping functions were used to express the process of model development (and loss function) following the description of the selected supervised machine learning (SML) algorithms adopted for the purpose of this study. The training dataset S consists of the original features identified at the point of data identification and collection is represented by X_i , where i is the number of features existing in the original dataset of patients whose records were collected (number of PD risk cases), and X'_j consists of the features relevant for predicting the risk of PD where $j \leq i$. The process of feature selection is represented by the mapping:

$$F: X_i \rightarrow X'_j \qquad (1)$$

where X_i are the original set of attributes collected and X'_j are the relevant features selected by the feature selection (FS) method.

Following the process of feature selection, the new PD dataset records, $X_{jk}{}'$ where k is the number of PD patients' record and j is the number of relevant features selected from the original I features. If k datasets were selected for training the supervised machine learning (SML) algorithms adopted for the model using the relevant variables, then the model can be represented by the mapping:

$$\varphi: X_{jk} \rightarrow Y_k$$

Defined as $\varphi(X_{jk}) = Y_k$ for all patients k (2)

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 6 , June 2021

where $X_{jk}$is the set of relevant attributes, j for patient, k and Y_k is the risk class of the PD patient, k given the values of X_{jk}.

The mapping function which describes the predictive model formulated for PD risk using the identified risk factors/variables (relevant features) is:

$$\varphi: X_{jk} = \{^{supervised}_{unsupervised} (3)$$

where $X_{jk}$is as described in equation (2)

Supervised machine learning (SML) algorithms are generally black boxed models; which implies that there is no general equation that can be used to describe the predictive model using a mathematical representation. Although, all SML algorithms have a metric that is used to evaluate how well the model is doing during the training and testing process of model development.

V. EXPERIMENTAL RESULTS

For the purpose of developing the predictive model for the classification of PD risk, the collected data containing information about the values of the identified indicators for PD risk were used to formulate the model. The dataset collected was divided into two parts: training and testing data; the training data was used to formulate the model while the test data was used to validate the model.

The division of the dataset into training and testing was done using percentage splitting and it was done in the folds 60%, 70% and 80% for training and 40%, 30% and 20% for testing respectively.

The results are shown in Table 2.

Table 2: Summary of Result of the Predictive Models

Radial Basis Function classification model

Percentage Split	**Feature Selection**	**Accuracy (%)**	**TP rate**		**FP rate**		**F1-Score**		**Precision**	
			No-Risk	Risk	No-Risk	Risk	No-Risk	Risk	No-Risk	Risk
60/40 Dataset	RBF Only	87.18	0.529	0.967	0.033	0.471	0.64	0.96	0.818	0.881
	PSO + LR	88.46	0.588	0.967	0.033	0.412	0.69	0.93	0.833	0.894
	PSO + RBF	93.59	0.706	1	0	0.294	0.83	0.96	1	0.924
70/30 Dataset	RBF Only	89.83	0.692	0.957	0.043	0.308	0.75	0.94	0.818	0.917
	PSO + LR	89.83	0.692	0.957	0.043	0.308	0.75	0.94	0.818	0.917
	PSO + RBF	93.22	0.769	0.978	0.022	0.231	0.75	0.94	0.909	0.938
80/20 Dataset	RBF Only	89.74	0.7	0.966	0.034	0.3	0.78	0.93	0.875	0.903
	PSO + LR	89.74	0.7	0.966	0.034	0.3	0.78	0.93	0.875	0.903
	PSO + RBF	74.36	0	1	0	1	0.78	0.93	0	0.744

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 6 , June 2021

For the dataset containing all 22 variables, when 60% of the dataset were used for training and 40% for testing, the True Positive (TP) rate values were 0.529 and 0.967, the False Positive (FP) rate values were 0.033 and 0.471, precision rate were 0.818 and 0.881 and F1-score values were 0.64 and 0.96 for no-risk and risk respectively and the model had an accuracy of 87.18%.
When 70% of the dataset was used for training and 30% for testing, the True Positive (TP) rate values were 0.692 and 0.957, False Positive (FP) rate values were 0.043 and 0.308, precision rate were 0.818 and 0.917 and F1-score values were 0.75 and 0.94 for no-risk and risk respectively and the model had an accuracy of 89.83%.
When 80% of the dataset was used for training and 20% for testing, the True Positive (TP) rate values were 0.7 and 0.967, False Positive (FP) rate values were 0.034 and 0.3, precision rate were 0.875 and 0.903 and F1-score values were 0.78 and 0.93 for no-risk and risk respectively and the model had an accuracy of 89.74%.
When all the variables in the dataset was used, the model with the best performance was achieved when 70% of the dataset were used for training and 30% for testing.

Feature selection model using PSO with LR

After feature selection was done, using the 15 variable selected, when 60% of the dataset was used for training and 40% for testing, the True Positive (TP) rate values were 0.588 and 0.967, False Positive (FP) rate values were 0.033 and 0.412, precision rate were 0.833 and 0.894 and F1-score values were 0.69 and 0.93 for no-risk and risk respectively and the model had an accuracy of 88.46%.
When 70% of the dataset was used for training and 30% for testing, the True Positive (TP) rate values were 0.692 and 0.957, False Positive (FP) rate values were 0.043 and 0.308, precision rate were 0.818 and 0.917 and F1-score values were 0.75 and 0.94 for no-risk and risk respectively and the model had an accuracy of 89.83%.
When 80% of the dataset was used for training and 20% for testing, the True Positive (TP) rate values were 0.7 and 0.967, False Positive (FP) rate values were 0.034 and 0.3, precision rate were 0.875 and 0.903 and F1-score values were 0.78 and 0.93 for no-risk and risk respectively and the model had an accuracy of 89.74%.
When feature selection was done using Particle Swarm Optimization (PSO) and Logistic Regression (LR), the model with best performance was achieved when 70% of the dataset were used for training and 30% for testing.

Feature selection model using PSO with RBF

After feature selection was done, using the 15 variable selected, when 60% of the dataset was used for training and 40% for testing, the True Positive (TP) rate values were 0.706 and 1.00, False Positive (FP) rate values were 0.00 and 0.294, precision rate were 1.00 and 0.924 and F1-score values were 0.83 and 0.96 for no-risk and risk respectively the model had an accuracy of 93.59%.
When 70% of the dataset was used for training and 30% for testing, the True Positive (TP) rate values were 0.769 and 0.978, False Positive (FP) rate values were 0.022 and 0.231, precision rate were 0.75 and 0.94 and F1-score values were 0.909 and 0.938 for no-risk and risk respectively and the model had an accuracy of 93.22%.
When 80% of the dataset was used for training and 20% for testing, the True Positive (TP) rate values were 0 and 1, False Positive (FP) rate values are 0 and 1, precision rate were 0 and 0.744 and F1-score values were 0.78 and 0.93 93 for no-risk and risk respectively and the model had an accuracy of 74.36%.
When feature selection was done using PSO and RBF, the model with best performance was achieved when 60% of the dataset were used for training and 40% for testing.
In concluding the discussion on the results, out of the predictive models developed for Parkinson disease risk, it was discovered that the classification done by the PSO with RBF feature selection methods was the most effective predictive model.

VI.CONCLUSION AND FUTURE WORK

Following the use of feature selection methods in identifying the variables relevant to designing the predictive model, the accuracy of the predictive model increases compared to when normal classification was used. The model had the highest level of accuracy (93.59%) after feature selection was done using RBF for classification and 60/40 (training/testing) percentage splitting, followed by an accuracy of 93.22% also after feature selection was done using RBF for classification and 70/30 (training/testing) percentage splitting.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 6 , June 2021

It can be concluded that using RBF for classification after feature selection with PSO provides the predictive model with the highest level of accuracy which makes it the model with the best performance.
This model will increase the awareness of Nigerians on the possibility of having Parkinson disease and makes them take necessary action in treating it and knowing what next to do.
It is recommended that Parkinson disease risk prediction system can be implemented on a web based system or mobile platform in order to make the system available to a large number of people.

REFERENCES

1. Mehri S, Omid M, Masoud E, Vahid S, Fariborz K, Mohammad R. N., Fereshteh A, Ahmad C. (2018). Demographic features and clinical characteristics of patients with Parkinson's disease in Isfahan, Iran. *Iranian Journal of Neurology* Vol. 17(1), pp. 6-10
2. Andrew, J., Ben, S., Yoav, B.A, Daniel D., Hughes H., Susan E., and Lee P.C. (1992). What Features Improve the Accuracy of Clinical Diagnosis in Parkinson's disease: A Clinicopathologic Study. *Neurology* 42 (6), 1142
3. Lewis SJ, Foltynie T, Blackwell AD, Robbins TW, Owen AM, Barker RA. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *J Neurol Neurosurg Psychiatry.* 2005;**76**(3):343–348. doi: 10.1136
4. Hoff, J. I., Wagemans, E. A., and Hilten. J. J. (2001). Accelerometric Assessment of Levodopa Induced Dyskinesias in Parkinson's Disease. *Movement Disorders* 16 (1), 58 - 61.
5. Fahn, S. (2003). "Description of Parkinson's disease as a clinical syndrome." *Annuals of the New York Academy of Sciences* 991 (1), 1 14.
6. K. N. R. Challa, V. S. Pagolu, G. Panda and B. Majhi, "An improved approach for prediction of Parkinson's disease using machine learning techniques," *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 2016, pp. 1446-1451, doi: 10.1109/SCOPES.2016.7955679.
7. Chauncey Spears, Understanding Parkinson. Parkinson's Foundation Movement Disorders Fellow at the University of Florida, a Parkinson's Foundation Center of Excellence. https://www.parkinson.org/Understanding-Parkinsons/Symptoms/Non-Movement-Symptoms/Depression accessed 13th June, 2021.
8. Jorge Garza Ulloa (2019). Update on Parkinson's Disease. *American Journal of Biomedical Science and Research.* DOI: 10.34297/AJBSR.2019.02.000614
9. Quinlan, J.R. (1986) Induction of Decision Trees. Machine Learning, 1, 81-106. http://dx.doi.org/10.1007/BF00116251
10. Cruz, J.A. and Wishart, D.S. (2006). Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics* 2: 59 – 75.
11. Geeta Yadav, Yugal Kumari, Gadadhar Sahoo (2011). Predication of Parkinson's Disease Using Data Mining Methods: A Comparative Analysis of Tree, Statistical and Support Vector Machine Classifiers. *Indian Journal of Medical Sciences*, Vol. 65 (6).
12. Ferdous Wahid, Rezaul Begg, Chris J. Hass, Saman Halgamuge, David C. Ackland (2015). Classification of Parkinson's Disease Gait using Spatial-Temporal Gait Features. *IEEE Journal of Biomedical and Health Informatics*. DOI:10.1109/JBHI.2015.2450232
13. Mehrbakhsh Nilashi, Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi, Mohammadreza Farahmand (2017). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. *Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Science. Published by Elsevier* doi.org/10.1016/j.bbe.2017.09.002
14. Chao Ma, Jihong Ouyang, Hui-Ling Chen, Xue-Hua Zhao (2014) 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* doi.org/10.1155/2014/985789
15. Yildirim, P. (2015). Filter-Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease. *International Journal of Machine Learning and Computing* 5(4): 258 – 263.
16. Hall, M.A. (1999) Correlation-Based Feature Selection for Machine Learning. PhD Thesis, University of Waikato, Hamilton.
17. Christopher G. Goetz (2011). The History of Parkinson's Disease: Early Clinical Descriptions and Neurological Therapies. *Cold Spring Harbour Perspectivs in Medicine.*Vol 1(1). doi: 10.1101/cshperspect.a008862
18. Olanow, C.W. (2014). Manganese-Induced Parkinsonism and Parkinson's Disease. *Annals of the New York Academy of Sciences.* DOI: 10.1196/annals.1306.018
19. Jaya Sanyal, D.P. Chakraborty, Biswanath Sarkar, Tapas Kumar Banerjee, Subhash Chandra Mukherjee, Bidhan Chandra Ray, V.R. Rao (2010). Environmental and Familial Risk Factors of Parkinsons Disease: Case-Control Study. *The Canadian Journal of Neurological Sciences*. Vol. 37: 637-642
20. Rybkina, V.L., Azizova, Tamara (2018). The risk factors of Parkinson's disease. *Nevrologicheskii Zhurnal*. DOI: 10.18821/1560-9545-2017-22-6-287-294
21. Tufail M. (2019). Clinical Features and Risk Factors of Parkinson's Disease in a Population of Khyber Pakhtunkhwa, Pakistan: A Case-Control Study. *Neurological disease Journal.* Doi.org/10.1159/000506742
22. Sellam, V., & Jagadeesan, J. (2014). Classification of Normal and Pathological Voice Using SVM and RBFNN. Journal of Signal and Information Processing, 2014.
23. Kapoor, T., and Sharma, R. K. (2011). Parkinson's disease Diagnosis using Mel frequency Cepstral Coefficients and Vector Quantization. International Journal of Computer Applications, 14(3), 43-46.
24. Rubén Armananzas, Concha Bielza, Kallol Ray Chaudhuri, Pablo Martinez-Martin, Pedro Larranaga. Unveiling relevant non motor Parkinson's disease severity symptoms using a machine learning approach. *Artificial Intelligence in Medicine* 58(3), 195-202.
25. Yulianti, A. N. Syapariyah, A. Saifudin, T. Desyani., (2020). Feature Selection Techniques to Choose the Best Features for Parkinsons Disease Predictions Based on Decision Tree. *Journal of Physics: Conference Series.* doi:10.1088/1742-6596/1477/3/032008

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 6 , June 2021

26. Zhennao Cai , Jianhua Gu, Caiyun Wen, Dong Zhao, Chunyu Huang, Hui Huang , Changfei Tong, Jun Li, and Huiling Chen (2018). An Intelligent Parkinson's Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach. Doi.org/10.1155/2018/2396952
27. Hui-Ling Chen, Chang-Cheng Huang, Xin-Gang Yu, Xin Xu, Xin Sun, Gang Wang, Su-Jing Wang (2013). An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Systems with Applications, Published by Elsevier*. Vol 40 pp. 263–271.
28. M. Hariharana, Kemal Polat, R. Sindhu (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*. Vol. 113(3)