Nature Inspired Binary Grey Wolf Optimizer for Feature Selection in the DETECTION of NEURODEGENERATIVE (PARKINSON) DISEASE

Vamsidhar Enireddy [1], Kalyani Gunda[2], N Lakshmi Kalyani[3], Kolla Bhanu Prakash[4]

[1,2,4]Department of CSE, Koneru Lakshmaiah Education Foundation
[3]Department of CSE, VNR Vignana Jyothi Institute of Engineering & Technology
[1]enireddy.vamsidhar@gmail.com, [2]kalyani47124@gmail.com,[3]neerukondait@gmail.com and
[4]drkbp@kluniversity.in

ABSTRACT

As the Parkinson's disease detection is the topmost chronic disease which was on the 3rd position. By using the parameters of both Parkinson and Non-Parkinson's person, we can build a model using algorithms which can easily detect the disease. For dealing with the dataset which is of high dimensions, that can lead to noise or redundancy which may affect the model performance. For choosing the Optimum Features that are required for the model building is to be selected using the Nature Inspired Optimization Techniques like Particle Swarm Optimization (PSO), Gravitational Search Algorithm (GSA), Differential Evolution (DE), Evolutionary Programming (EP), and Evolution Strategy (ES).Among all Binary Grey Wolf Optimizer is the best method for optimum features for building a model and similarly the machine learning techniques like K-nearest neighbors, Decision Tree Classifier, Naive Bayes, Support Vector Machine and implemented with kernelized, Ensemble Learning techniques which having more than one classifier like Bagging, Boosting, Adaboosting are used for best performance, Among all Extreme Gradient Boosting algorithm achieves the best performance characteristics and obtained the satisfactory results

Key words: Adaboosting, Binary Grey Wolf algorithm, Ensemble learning techniques, Extreme Gradient Boosting algorithm, Feature selection, Parkinson's disease, Support vector machine.

1.INTRODUCTION

Parkinson is a progressive, irreversible, chronic, devastating neurodegenerative disease, a central nervous system disorder which it was affected about 10 million people worldwide with more than 70,000 Australians and almost one million are affected with this disease in the United States itself [1]. In brain a chemical called as Dopamine helps to build coordination in the movement of the person, this chemical shortage in the person's brain are affected by this situation and it also progress with time. There are certain symptoms for Parkinson's disease such as slowness in movement, rigidity, tremor, and dysphonia. Other issues related to this disease include depression, pain, constipation, and fatigue that have a huge impact on the daily routines of a person [2].

It can be known that 60,000 new cases recorded in this year. Unfortunately, there is no proper cure for this and recent study from Mayo Clinic Showed the Parkinson's disease incidence which significantly incremented during the period 1976-2005.Especially, the men with age 70 and above takes a major role in the list. PD is unique [4] to each person and the treatment should. Therefore, be as patient-centered as possible for all the patients. Detection in early stages can help to reduce the symptoms using the drugs. Furthermore, for People with Parkinson's (PWP), the condition is different every day and having personalized treatment plans can greatly improve their lives [3]. The methods such as walking and vocal analysis are mostly used in the diagnosis of the disease. The improvement in the machine learning algorithms helps to diagnose the disease [4].

As PD is subsequent most regular neurological condition after Alzheimer's, a need of Computer-based evaluation of Parkinson's disease is important that it focus is on advanced research, to help the people suffering from the disease. This application is crucial, as currently the methods using the gold standard test — known as the Unified Parkinson's Disease Rating Scale (UPDRS) [5]. The research is an important step for looking to a cure. The peculiar property in this disease is that it initiates on one side of body (hemi parkinsonism) and progress to the other side. The distinguishing proof and measurement of the manifestations of Parkinson's illness is a non-paltry issue, i.e., mostly reliant on numerous elements identifying with the patient's physical and with medical condition [6].

Neurological clinical assessment is the most common method to diagnosis this disease. In order to affirm and recognize Parkinson's disease with the other comparative diseases, there is no lab parameter, but nuclear imaging study can help to distinguish them. The patients suffering with the Parkinson's disease, brain patterns are studied using atomic imaging methods which employ the gamma radiations of various substances at different places in the cerebrum [12]. The imaging contemplates have accomplished significant levels of acknowledgment rates, however they are convoluted and open patients to a small portion of radiation [13,14,15]. In most of cases, treatment with a chemical called levodopa can deliver great medical reaction and affirms the analysis of Parkinson's disease [16].

2.MATERIALS AND METHODS

2.1 Data and Material

The most accepted and accessible datasets of this kind are "Parkinson's Data Set" [7] and "Parkinson's Telemonitoring Data Set" [8], both accessible from UCI Machine Learning Repository. These signs and symptoms are grouped into Motor Symptoms and Non-Motor symptoms. Motor symptoms are those which affect movement and muscles. Non-motor are neurobehavioral cognitive problems, sleep problems, sensory problems, and autonomic neuropathy (dysautonomia). Many multiple speech recordings per subject is put together in the speech datasets that are utilized in field of Parkinson's disease [9]. There are two classifications in the datasets

(i). It consists of one term that is repeated

(ii). These consists of various vocal terms.

In these many datasets belongs to the first category. Consequently, most of the examinations on PD analysis are led on these datasets [8]; be that as it may, none of them could get 100% grouping exactness. For accomplishing ideal classification accuracy forecast, the patients with Parkinson are watched and assembled the highlights.

PD is hard to recognize ahead of schedule because of the unobtrusive starting manifestations. There is a critical weight to patients and the human services framework because of deferrals in analysis [10]. The trouble in early PD conclusion has propelled researchers to create screening devices depending on automated algorithms to separate sound controls from individuals with PD.

2.2 Data Description

Parkinson's Disease Classification Data Set used in the study is obtained from UCI Repository[11]. The dataset consists of 5876 subjects of PD Patients with 20 features and 188 patients are PD patients and rests of them are healthy. The data set is a binary classification problem since it is composed of two

classes as healthy and PD patient. The dataset created with the twenty different sound recordings for each subject and it contains features both with the time and frequency in speech are gathered along with the different types of vocals recordings [8].

2.3. Data Pre-Processing

The data is to be pre-processed to clean and remove the redundant features or noisy data to improve performance+. It is an important step to for the data transformation into a required format. Feature selection is the pre-processing that chooses a subset of original features according to a set of evaluation criterion. The features in Parkinson's disease dataset are taken for the analysis and can be estimated in histogram representation for the features A histogram in Figure 1 represents the shape and spread of continuous data fall in specific range.
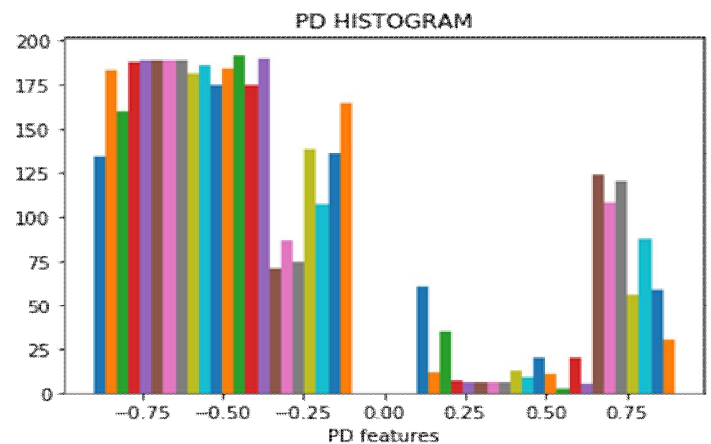

Figure 1: Histogram represents the shape and stretch of uninterrupted data fall

Total mean of the features are -0.45009599942297557. Each color represents the count or density of a features.

2.4. Feature Selection

A large number of features are extracted and a high dimensional data is formed, but these large number of features also increases the complexity of the classifier when given as input, hence the feature selection had become a good solution for these type of high dimensional data where the features are selected by removing the unwanted features and repetitive features [17].Feature choice endeavors to locate the most discriminative subset of highlights for some classifier. Customarily, feature selection is a compelling method to assess the most useful highlights, which adds to execution improvement and feature reduction [18]. Given an issue with d highlights, we have 2d potential solutions, which make a comprehensive quest impracticable for high dimensional component spaces. the errand of selecting the features as a optimization issue, in which the fitness function can be a

proportion of feature space distinct or even some classifier's. Feature choice techniques can be utilized to distinguish and evacuate unneeded, unimportant, and repetitive properties from information that don't add to the exactness of a prescient model. Accuracy and number of features chosen are utilized to assess and contrasted and dataset from the UCI archive.

Feature selection approaches can be grouped as follows

- filter approach,
- wrapper approach
- Embedded approach.

These optimization models have some limitations to solve the problems, so the nature inspired algorithms are the substitution to solve these problems and for searching the optimum result of the problem [19]. Taking the inspiration from the nature many nature-inspired algorithms are designed from it. These include biological, and social behavior of animals, fish, birds, bat, bees, and wolves and are apply the nature processes to solution [19]. A large number of nature-inspired algorithms are available in literature namely genetic algorithm, bat algorithm, cuckoo search algorithm, firefly algorithm [20], bee swarm [21], fish swarm[22], binary gravitational algorithm[23], and etc. Many models have been developed for feature selection [29, 30]. In this paper binary grey wolf algorithm is used to find a best feature subset. This algorithm is a nature-inspired algorithm, which replicate hunting procedure of wolves in wildlife [24].

2.5. Feature Transform

Before building any machine learning algorithms, the features are observed; the variables are different types like ordinal, continuous, nominal, and dichotomous. For better handling features in the feature, there is a need to transform the features to a specific format for optimum model performance. Feature transformation can be done by scaling the given feature to a given range. In order to translate and scale each feature in the given range mostly between zero and one the estimator calculates the values for each feature individually. Some of the transformation techniques are

- Scaling or normalizing features within a range, either 0 to 1.
- Principle Component Analysis and its variants.
- Random Projection.
- Neural Networks.
- SVM also transforms features internally.
- Transforming categorical features to numerical.

For transforming, the machine learning techniques like Standard Scaler, Min-Max Scaler, Robust Scaler and Normalizer are used. In the detection of demented person, Min-Max Scaler is used to transform the features which are used to build the machine learning algorithm for optimum results.

2.6. MinMaxScaler

An elective way to deal with Z-score standardization (or normalization) is the alleged Min-Max scaling. In this methodology, the information is scaled to a fixed range [0 to 1]. The expense of having this limited range - as opposed to normalization - is that we will wind up with littler standard deviations, which can stifle the impact of exceptions.

A Min-Max scaling equation is represented as

$$X_{SC} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where, X_{min} is the minimum value in the data, X_{max} is the maximum and X is the current value which it was dealing to transform

2.7. Grey wolf optimization (GWO)

It is one of the recent bio-inspired algorithm developed for optimization, based on the grey wolves and their hunting skills developed by themselves in nature[24]. This Metaheurstic algorithm is inspired by dark wolves also called as Canis Lupus scientifically. This Grey wolf algorithm emulates the command and hunting components of the grey wolves [25].The group of grey wolves is called as a pack. The grey wolf pack comprise of four types Alpha, Beta, Delta and Omega wolves, in these alpha wolves are dominant and they take the decisions regarding hunting, selecting rest place, time of waking up. The beta wolves help the alpha wolves in the activities. The delta wolves comprise of the care takers, hunters, scouts, sentinels and they submit to the above wolves. The least considered wolves are the omega wolves where they surrender to the all other categories[25].

2.7.1 Social Hierarchy

In the scientific model for the GWO the fittest arrangement is known as the alpha (α). Beta (β) and delta (δ) are second and third best, leaving the rest as omega (ω).
The fundamental phases of Gray Wolf Hunting's are

- Track, chase and move towards the prey
- Surround, pursue and bothering until a movement starts in it.
- Attack the prey in the stationary circumstances [24].

In the mathematical form the best solution in the algorithm is alpha(α). Beta(β), and delta (δ) gives the next best solutions after the alpha(α). The omega (ω) represents the remaining solutions. The hunting process in GWO is initiated and taken forward by α, β, δ and ω follows them.

Generally, Grey Wolf Optimization algorithm performs well for the continuous data

Here a binary model is used in two different approaches

- In the principal approach to find the updated positions for the three best solutions binary and stochastic crossover is applied.

- In the subsequent methodology, to crush the continuous refreshed position, at that point stochastically limit these qualities to discover the updated positions of the binary grey wolf. These two methodologies is used for selecting the best feature subset.

2.7.2. Binary grey wolf optimization (BGWO)

Feature selection is an example of a binary optimization problem, hence binary version of GWO is used. In a research article Emary et al. [25] demonstrated the method of solving the feature selection problem using two binary grey wolf optimizations BGWO1 and BGWO2.

The 0's and 1's represent the features in the BGWO, where 0 denotes the lack of a feature and 1 denotes the existence of a feature. Since it is the binary value, continuous values are converted to binary values, the binary version of grey wolf optimization for feature selection [25] is performed using by using the sigmoid function

$$\text{Sigmoid function} = \frac{1}{1+e^{-10(x-0.5)}} \quad (2)$$

2.7.2.1 Model 1 BGWO1

In the principle approach crossover operators are utilized for updating the wolves position and it is defined as :

$$C(t+1) = \text{Crossover}(P1, P2, P3) \quad (3)$$

Where, Crossover (P1, P2, and P3) is crossover process among the solutions, and P1, P2, and P3 are the binary vectors exaggerated by the progress of alpha, beta, and delta wolves.

InBGWO1, P1, P2, and P3 are defined using

$$P_1^d = \begin{cases} 1, & if(C_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0, & otherwise \end{cases} \quad (4)$$

Where, Cd represents position of alpha, dimension of search space given by d, and $bstep_\alpha^d$ represents binary step represented as

$$bstep_\alpha^d = \begin{cases} 1, & if\ cstep_\alpha^d \geq r_3 \\ 0, & otherwise \end{cases} \quad (5)$$

Where, r3 represents random vector in [0, 1], $cstep_\alpha^d$ is continuous valued step size given as

$$cstep_\alpha^d = \frac{1}{1+exp\left(-10\left(A_1^d.D_\alpha^d-0.5\right)\right)} \quad (6)$$

Where A_1^d and D_α^d are calculated as .

$$P_2^d = \begin{cases} 1, & if(C_\beta^d + bstep_\beta^d) \geq 1 \\ 0, & otherwise \end{cases} \quad (7)$$

Where, Pd gives position of beta, dimension of search space given by d, and $bstep_\beta^d$ denotes the binary step stated as

$$bstep_\beta^d = \begin{cases} 1, & if\ cstep_\beta^d \geq r_4 \\ 0, & otherwise \end{cases} \quad (8)$$

Where, r4 represents random vector in [0, 1], and $cstep_\beta^d$ is continuous valued step size given by

$$cstep_\beta^d = \frac{1}{1+exp\left(-10\left(A_2^d.D_\beta^d-0.5\right)\right)}, \quad (9)$$

Where A_2^d and D_β^d are given as

$$P_3^d = \begin{cases} 1, & if(C_\delta^d + bstep_\delta^d) \geq 1 \\ 0, & otherwise \end{cases} \quad (10)$$

Where Pd is the location of delta, d is dimension of search space, and $bstep_\delta^d$ represents the binary stride represented as

$$bstep_\delta^d = \begin{cases} 1, & if\ cstep_\delta^d \geq r_5 \\ 0, & otherwise \end{cases} \quad (11)$$

Where, r5 is a random vector in [0, 1], and $cstep_\delta^d$ denote a uninterrupted valued step size that is represented as

$$cstep_\delta^d = \frac{1}{1+exp\left(-10\left(A_3^d.D_\delta^d-0.5\right)\right)}, \quad (12)$$

After obtaining P1, P2, and P3, the positions are updated as follows:

$$Cd(t+1) = \begin{cases} P_1^d & if\ r_6 < \frac{1}{3} \\ P_2^d & if\ \frac{1}{3} \leq r_6 < \frac{2}{3} \\ P_3^d & otherwise \end{cases} \quad (13)$$

Dimension of search space is represented by d, and r6 is uniformly distributed between [0, 1].

At first, the number of inhabitants assigned either zero or one. Thereafter, for each wolf fitness is assessed and the

arrangements are characterized as alpha, beta and delta for the first, second and third positions. P1, P2, and P3 are figured by conditions (4), (6) and (9). This process repeats and the values are refreshed until the algorithm meets the stopping criteria. The final result is chosen based on the alpha group which gives the optimal feature subset.

2.7.2.2. Pseudo code of BGWO1

1. Begin

2. Randomly initialize the population of grey wolves, X

3. Initialize the parameters A, b and c.

4. Calculate fitness function of wolves f(c)

5. Set $X\alpha$ = position of the best wolf

6. $X\beta$ = position of the second-best wolf

7. $X\delta$ = position of the third best wolf

8. for t=1 to max no.of iteration, T

9. for i=1 to number of wolves, N

10. Compare Y1Y2 &Y3 using equations (4), (7) and (10).

11. Generate X_i^{new} by applying crossover between Y1Y2 &Y3 in equation (13)

12. next i

13. Calculate the fitness of all grey wolves, $f(C_i^{new})$

14. Update the position of α, β and δ.

15. Update the parameter A, b and c

16. t=t+1

2.7.3. Model 2 BGWO2

In the next move, BGWO2 algorithm refreshes location of wolf by changing the location into a binary vector,

$$C^d(t+1) = \begin{cases} 1 & \text{if } S\left(\frac{X1+X2+X3}{3}\right) \geq r_7 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Here r7 represents a random vector in range [0, 1], d represents search space, and S represents sigmoid function, and it is given by

Sigmoid function $S(x) = \dfrac{1}{1+e^{-10(X-0.5)}}$ (15)

2.7.3.1. Pseudo code of BGWO2

In the first state, the underlying populace of wolves is haphazardly instated (either bit 1 or 0) then the fitness is assessed and based on fitness alpha, beta, and delta, are

chosen. P1, P2, and P3 are figured individually, then the new location of the wolf is updated. Thereafter, the wellness of wolves is assessed, and the situation of alpha, beta, and delta are refreshed. The algorithm runs till the stopping criteria is met and at last optimal feature subset is given by alpha solution.

2.7.3.2. Pseudo code of BGWO2

1. Begin

2. Randomly initialize the population of grey wolves, X

3. Initialize the parameters A, b and c.

4. Calculate the fitness function of wolves f(c)

5. Set $P\alpha$ = position of the best wolf

6. P_β = position of the second-best wolf

7. P_δ = position of the third best wolf

8. for t=1 to max. no. of iteration, T

9. for i=1 to number of wolf, N

10. Compare P1 P2 &P3 using equations (4), (7) and (10).

11. Generate C_i^{new} by applying crossover between P1 P2 &P3 in equation (13)

12. next i

13. Evaluate the fitness of all grey wolves, $F(X_i^{new})$

14. Update the position of α, β and δ.

15. Update the parameter A, b and c

16. next $C^d(t+1)$ using (14)

17. calculate sigmoid function using equation (15)

2.8. Fitness Function(F(X))

The fitness function is used to evaluate an individual grey wolf position is

$Fitness = \dfrac{\#CCS}{\#N}$ where

#CCS Number of samples classified correctly,

#N are the total number of Samples

Since depending on the accuracy, the fitness is evaluated, the threshold value of the fitness function is fixed within 100%. In order to evaluate the algorithm the results are compared with other existing popular algorithms.
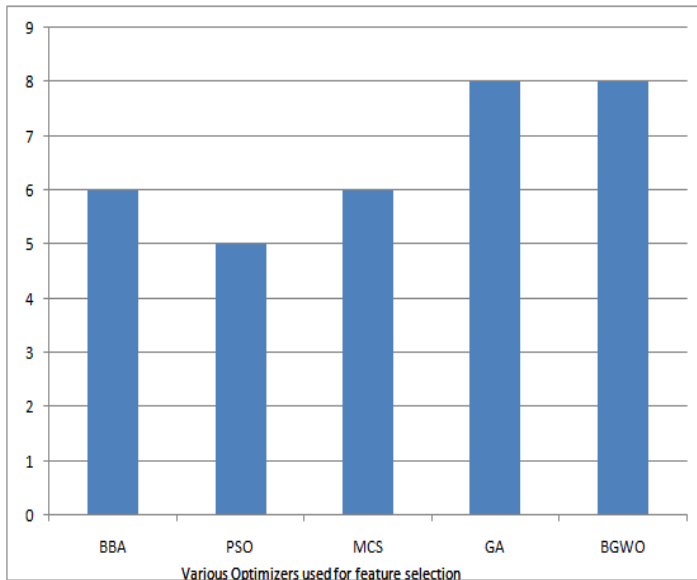
Figure 2: Comparison of features with various optimizers

The outcome as shown in Figure 2, displays that GWO algorithm is able to provide very competitive results compared to these well-known meta-heuristics [23]. Four common optimizers namely Binary Bat Algorithm(BBA), Particle Swarm Optimization (PSO), Modified Cuckoo Search Algorithm (MCS) and Genetic Algorithm(GA), are compared with Binary Grey Wolf Optimization performed resulted the with a satisfactory performance.

2.9. Extreme Gradient Boosting (XGB) Classifier Model

Extreme Gradient Boosting is one of the Ensemble machine learning technique of highly versatile and flexible that can work through classification, regression, and ranking problems. XGBoost is a decision-tree-based ensemble machine learning technique which used a gradient boosting technique in prediction, for data other than the unstructured data such as images, text artificial neural network (ANN) is outperformed where for structured and tabular data; decision tree based model is implemented for better results [26,27].

There is a wide choice of applying the model as these are used to perform classification, ranking, regression, and required estimate analysis, for scalability (Portability), language scalability and cloud integration(AWS, Azure, Yarn clusters).it performs well because the principle of gradient boosting algorithm improves based on GBM framework through system optimization and algorithmic enhancement.

2.9.1. Objective Function of XGBOOST

The objective function also called as loss function and regularization is minimized at iteration t

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(X_i)\right) + \Omega(f_t) \tag{16}$$

Where y_i,real value label known fron training data set
and $\hat{y}_i^{(t-1)}$ is seen as f(x+$\triangle$x) where x= $\hat{y}_i^{(t-1)}$

"Cannot be optimized using traditional optimization methods in Euclidean space".

2.10 Taylor's Theorem and Gradient Boosted Trees

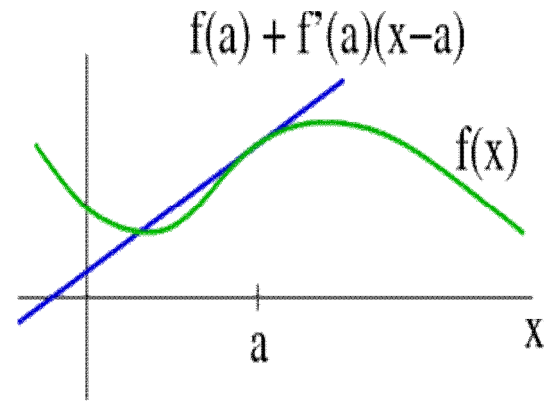The finest linear approximation(figure 3) for a function f(x) at point a as shown in Figure 3.

Figure 3: Displays Taylor linear estimation of a function in the region of a point 'a'

2.10.1. Requirement of Taylor approximation

To use in the conventional optimization technique the objective function has to be transformed into a function in Euclidean domain and this can be achieved by applying Taylor's theorem around the function f(x)[25,27].

$$f(x) \approx f(a) + f'(a)(x - a) \tag{17}$$

The loss function in the f(x) is represented by L, a denotes preceding stage (t-1) expected value and to add in every cycle t a new learner Δx is included. The objective function is calculated in every iteration. The term 'a' at the step (t-1) gives the prediction, new learner is (x-a). To limit the goal a step(t) is included. Hence the second-order Taylor approximation, we have:

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 \tag{17}$$

$$L^{(t)} \approx \sum_{i=1}^{n}\left[l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_t(X_i)\right) + \frac{1}{2}h_i f_t^2(x^i)\right] + \Omega(f_t)$$

$$\tag{18}$$

Using XGBoost objective

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \text{ and } h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \quad (19)$$

Here, g_i and h_i are the Ist and IInd order gradient statistics of the loss function simplified objective to minimize at step t:

$$L^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(x^i) \right] + \Omega(f_t) \quad (20)$$

The above equation shows a XGBoost simplified objective

Learners that reduce the loss function at iteration t.

$$\text{argmin}_x G_x + \frac{1}{2} Hx^2 = -\frac{G}{H}, H > 0 \ \min_x G_x + \frac{1}{2} Hx^2 = \frac{1}{2}\frac{G^2}{H} \quad (21)$$

2.11 Constructing next learner

A learner is to be build that can achieve a maximum reduction in loss at an iteration t

"Quality scoring function" above returns the base misfortune esteem used for a provided tree formation, implying that first misfortune work is assessed through utilizing the ideal weight esteems. Along these lines, for some random tree structure the ideal weights in leaves are calculated. For building next learner, the steps are considered

Begin with only one root
Iteration over all features and their values, and assess every conceivable split loss minimization

Gain = Loss (father instances) - (Sum of the Losses of both left and right branch)

The calculated gain value of best split would be a non-negative value and it should be greater than the gain parameter, else end the growth of the branch

2.12 Binary classification with log loss optimization

Binary classification and log loss objective function:

$$y \ln(p) + (1 - y) \ln(1 - p) \text{ where } p = \frac{1}{(1 + e^{-x})} \quad (22)$$

Here term y in {0,1} is a real label and probability score is given by p. The value of p is computed by applying sigmoid function to the output of GBT model x. By summing up across the CART learners the output x is calculated. The first and second derivate is calculated with respect to x to limit the log loss function [27].

2.12.1. Key parameters in XGBoost

N-estimators = number of runs XGBoost,
learning rate = learning speed,
early_stopping_rounds = over fitting prevention,
 stop early when improvement [40].
System optimization techniques may be a Parallelization, tree pruning and hardware optimization and for Algorithmic Enhancement techniques, Regularization, Sparsity Awareness weighted Quantile sketch and cross-validation. When algorithms like logistic regression, random forest, standard Gradient Boosting and Extreme Gradient Boosting (XGBoost) performed on dataset, The XGBoost algorithm model performs well in accuracy metrics and processing time [27].

3.EVALUATION METRICS

Evaluating a model gives a satisfying result in the performance. In many cases, accuracy metric is used to measure the model performance. Performance in a Classification model depends on the metrics. Some of the popular metrics are Confusion matrix, accuracy, F-score.
Confusion Matrix:The arrangement of the labels is represented in a matrix form as shown in Figure 4.

Figure 4: Confusion Matrix

Parkinson's disease detection is based on Binary classification with a prediction of resultant as "YES" or "NO", can be labeled as '1' or '0'. For this, the confusion matrix well performed in prediction.
True Positive (TP): value, PD Patients correctly classified as Parkinson's patient.
True Negative (TN): value of patients is labeled as healthy.
False Negative (FN): value of patients, model falsely labeled as healthy.
False Positive (FP): value of healthy patients labeled as having PD.

Based on the Positives and Negatives.

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

4. RESULTS AND DISCUSSIONS

For selecting the best classifier for the detection of Parkinson's disease, the machine learning algorithms are best useful. In machine Learning, the supervised learning Techniques are the Classification algorithms that are applied on a data having Labeled data that are classified based on the target.[30].

Some of the classifier Algorithms in the Supervised Learning are K-nearest neighbors, Decision Tree Classifier which selects the best optimum node known as Greedy Algorithm as shown in Table 1. The probability based Naive Bayes, The optimum Decision margin of Support Vector Machine and implemented with kernelized, In the supervised learning Ensemble Learning is the best technique which having more than one classifier like Bagging, Boosting, Adaboosting are used for best performance [31].

Table 1: Classifiers

S.No	Classifier	Accuracy
1	K- Neighbors Classifier	0.9387755102040816
2	Decision TREE CLASIIFIER	0.9591836734693877
3	Naïve Bayes Classifier	0.6938775510204082
4	Support vector classifier (SVM –linear)	0.8979591836734694
5	KERNEL SVM (polynomial)	0.7959183673469388
6	RBF kernel SVM	0.8775510204081632
7	Sigmoid Kernel SVM	0.8775510204081632
8	Artificial Neural Networks	0.8615
9	Bagging Classifier	0.7776315789473685
10	AdaBoost Classifier	0.7678947368421053
11	Gradient Boosting Classifier	0.9591836734693877
12	XGB Classifier	0.9795918367346939

In Gradient Boosting Classifier, the accuracy of the classification depends much on so the different learning rates are considered for the accuracy classification and validation. In this model the learning rates are Learning rate [0.05,0.075,0.1,0.25,0.5,0.75,1] as shown in Table 2.

Table 2: Training and Validation accuracy score

S.No	Learning Rate	Accuracy Score(Training)	Accuracy Score(Validation)
1	0.05	0.870	0.878
2	0.075	0.904	0.857
3	0.1	0.904	0.857
4	0.25	0.973	0.918
5	0.5	1.000	0.918
6	0.75	1.000	0.918
7	1	1.000	0.959

In this, accuracy metrics, we mainly focused on the validation. The model having high score represents best performance. Extreme gradient also performed for the accuracy classification [32].The Figure 5 shows the comparison of various classifiers can be plotted.
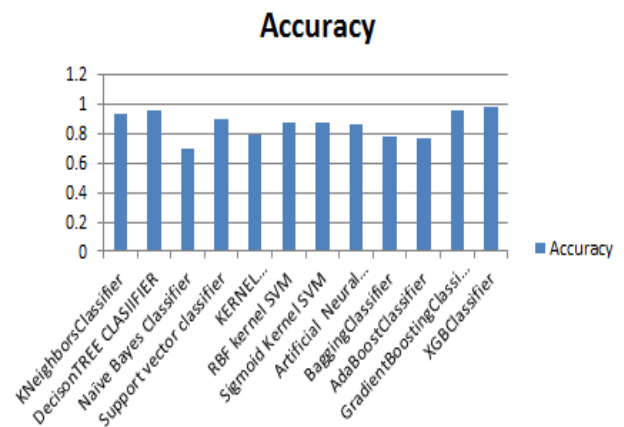

Figure 5: Accuracy of various classifiers

Among all the classifiers Extreme Gradient Boosting algorithm acheives the best performance characteristics [33]. The recall, precision and F-1 measured on training set for model selection is also used for the validation of model performance which are measured using the mathematical equations as

$$\text{Recall} : \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}}$$

$$\text{Precision} : \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{F-1measure}: \frac{2(\text{Recall} + \text{Precision})}{\text{Recall} + \text{Precision}}$$

Classification Accuracy: Accuracy works well when sample of classes are equal. It is defined as a ratio of Number of correct Predictions to the total number of samples expressed as

$$\text{Accuracy} = \frac{\text{number of correct predictions}}{\text{total number of samples}}$$

$$Accuracy : \frac{TP + TN}{TP + TN + FP + FN}$$

of a classifier on the dataset is the percentage of test data that are being classified accurately and mathematically represented as a known metric in machine learning for estimating quality of binary classifier is MCC [28].it is consistent where TP, TN, FP, FN are taken into consideration. It is the correlate as shown in Table 3 on coefficient is the metric, which measured between observed (Actual) labels to predicted values by binary classifier. [34]

$$MCC: \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Gives a value range from -1 to 1. In this, +1 indicate perfect prediction, The value 0 signifies that classifier is no better than randomly predicted and -1 indicates disagreement (not in related) between actual and predicted [35].

Table 3: Confusion matrix representation

	Truth:PD	**Truth: Healthy**
Truth: PD	True positives	True negatives
Truth: Healthy	False positives	False negatives

These metrics of accuracy, precision, recall are measured for different classifiers, among all the classifiers Extreme Gradient Boosting algorithm achieved the best performance characteristics [36].

The dataset shape, information and description about features are referred [37]. The optimum features are obtained using the Binary Grey Wolf Optimization, after the best feature Selection, the data is splitted to training and testing for the analysis [38].Training and testing sets are considered from the features [39].Training dataset being trained for the prediction and testing data is for sample evaluating metric measures. After Training and Testing is performed; the Extreme Gradient Boosting algorithm classifier [40] which performs ensemble technique (Boosting) with Gradient Search which is a heuristic algorithm for optimum (Best) results. This classifier model is used to detect the subject having Parkinson which is set to be a Target [41].

XGB Classifier:0.9795918367346939

The classification accuracy approach is 97.95918367346939approximately equal to 98% of prediction about the Parkinson disease accurately. The metric reports are analysed using python packages from sklearn, the recall, precision, F- measures are obtained as shown in Table 4.

Table 4: Metric report

Precision	Recall	F-1 Measure	Support
1.00	0.90	0.95	10
0.97	1.00	0.99	39

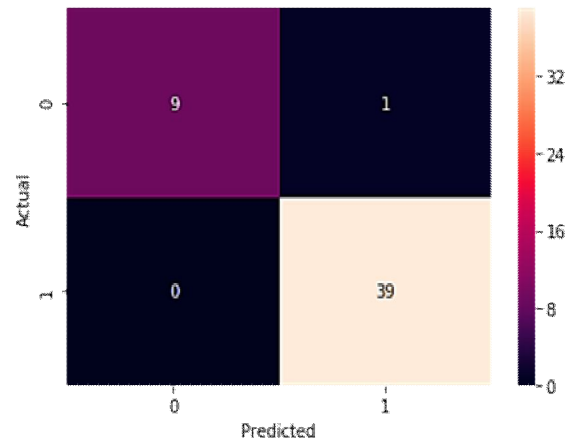|				
Accuracy	0.9849			
Macro avg	0.99	0.95	0.97	49
Weighted avg	0.98	0.98	0.98	49

Figure 6: Confusion Matrix

The model can also be built without feature selection, but the model performance is prone to redundancy. As the features with less performance leads to poor accuracy, we focus mainly on the Classification accuracy of a model which decides the model evaluation. The model can be compared with and without feature selection. The Figure 6 shows the confusion matrix and Figure 7 gives the comparison results for feature detection with and without BGWO accuracy.
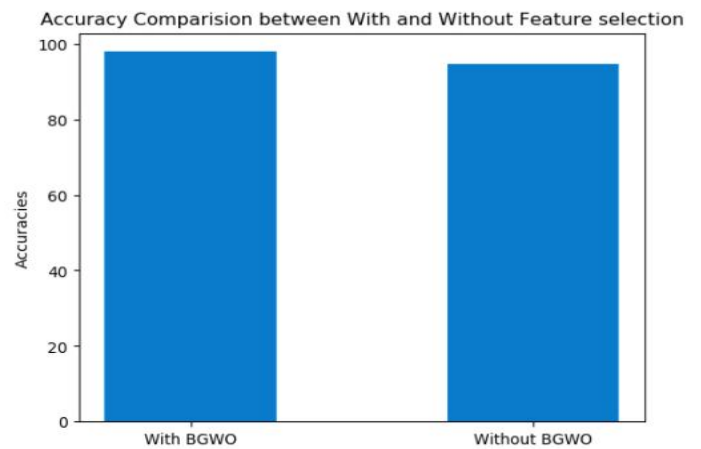

Figure 7: Accuracy comparison Plot for feature selection with and without BGWO

5.CONCLUSION

Parkinson's disease is one of the most advancing disease in the developed countries, identifying and quantifying the symptoms of this disease is difficult as they are dependent on many factors corelating to patient's mental and physical conditions [42].These can be represented in a format will be applied to a machine learning algorithm which performs well for the early detection of a Person with Parkinson's (PWP) can be easily evaluated with best performance results [43].

REFERENCES

1. "Statistics." Available at: http://parkinson.org/Understanding Parkinsons/ Causesand -Statistics/ Statistics, Last accessed May 2018, Jan 2018.
2. Prashant Shrivastava, Anupam Shukla, Praneeth Vepakomma, Neera Bhansali, Kshitij Verma, "A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease", Computer Methods and Programs in Biomedicine (2016), http://dx.doi.org/doi: 10.1016/j.cmpb.2016.07.029.
3. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/ Parkinsons -Disease.
4. X. Liu and H. Fu, "PSO-based support vector machine with cuckoo search technique for clinical disease diagnoses," Scientific World Journal, vol. 2014, Article ID 548483. https://doi.org/10.1155/2014/548483
5. M. A. Little, P.E.McSharry, S. J. Roberts, D. A. E. Costello, and I.M.Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," BioMedical Engineering OnLine, vol. 6, article 23, 2007 https://doi.org/10.1186/1475-925X-6-23
6. G. Pal and C. G. Goetz, "Assessing Bradykinesia in Parkinsonian Disorders," Frontiers in Neurology, vol. 4, p. 1859–1863, 2011.
7. A. Tsanas,M.A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 884–893, 2010.
8. L. O. Ramig, S. Sapir, C. Fox, and S. Countryman, "Changes in vocal loudness following intensive voice treatment (LSVT) in individuals with Parkinson's disease: a comparison with untreated patients and normal age-matched controls," Movement Disorders, vol. 16, no. 1, pp. 79–83, 2001.
9. [9]. B. E. Sakar, M. E. Isenkul, C. O. Sakar et al., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828–834, 2013 https://doi.org/10.1109/JBHI.2013.2245674
10. D. M. Huse, K. Schulman, L. Orsini, J. Castelli-Haley, S. Kennedy, and G. Lenhart, "Burden of illness in parkinson's disease," Movement disorders, vol. 20, no. 11, pp. 1449–1454, 2005.
11. Sakar, C.O., Serbes, G., Gunduz, A., Tunc, H.C., Nizam, H., Sakar, B.E., Tutuncu, M., Aydin, T., Isenkul, M.E. and Apaydin, H., 2018. "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform" Applied Soft Computing, DOI: https://doi.org/10.1016/j.asoc.2018.10.022
12. Arena J., Stoessl A.J. "Optimizing Diagnosis in Parkinson's Disease": Radionuclide Imaging. Parkinsonism Relat. Disord. 2015;22:S47–S51. doi: 10.1016/j.parkreldis.2015.09.029.
13. Weingarten C.P., Sundman M.H., Hickey P., Chen N. Neuroimaging of Parkinson's Disease: Expanding Views. Neurosci. Biobehav. Rev. 2015;59:16–52. doi: 10.1016/j.neubiorev.2015.09.007.
14. Oliveira F.P.M., Faria D.B., Costa D.C., Castelo-Branco M., Tavares J.M.R.S. "Extraction, Selection and Comparison of Features for an Effective Automated Computer-Aided Diagnosis of Parkinson's Disease Based on [123I]FP-CIT SPECT Images". Eur. J. Nucl. Med. Mol. Imaging. 2017;45:1–11. doi: 10.1007/s00259-017-3918-7.
15. Oliveira F.P., Castelo-Branco M. "Computer-Aided Diagnosis of Parkinson's Disease based on [123I]FP-CIT SPECT Binding Potential Images, Using the Voxels-as-Features Approach and Support Vector Machines". J. Neural Eng. 2015;12:26008. doi: 10.1088/1741-2560/12/2/026008.
16. Rizzo G., Copetti M., Arcuti S., Martino D., Fontana A., Logroscino G. "Accuracy of Clinical Diagnosis of Parkinson Disease. Neurology". 2016;86:566–576. doi: 10.1212/WNL.0000000000002350.
17. G.Chandrashekar, F.Sahin, "A survey on feature selection methods", Computers and Electrical Engineering January 2014 https://doi.org/10.1016/j.compeleceng.2013.11.024
18. Shoghian S, Kouzehgar M. "A comparison among wolf pack search and four other optimization algorithms". World Academy of Science, Engineering and Technology. 2012; 6(6):12–[19]. AshrafDarwish "Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications" , Future Computing and Informatics JournalVolume 3, Issue 2, December 2018, Pages 231-246, https://doi.org/10.1016/j.fcij.2018.06.001
19. H.Banati, M.Bajaj, "Firefly based feature selection approach" Int J. Comput. Sci. Issues 8(4)(2011)473-480.
20. Akbari, R., Mohammadi, A., Ziarati, K., 2010. "A novel bee swarm optimization algorithm for numerical function optimization". Commun. Nonl. Sci. Numer. Simul., 15(10):3142–3155. [doi:10.1016/j.cnsns.2009.11.003].

21. Xiao Lei LI,Zhi Jiang SHAO,JiXin QIAN. "An Optimizing Method Based on Autonomous Animats: Fish-swarm Algorithm" [J]. Systems Engineering - Theory & Practice, 2002, 22(11): 32-38.

22. Esmat Rashedi, Hossein Nezamabadi-pour, and Saeid Saryazdi, "BGSA: binary gravitational search algorithm", Natural Computing, September 2010, Volume 9, Issue 3, pp 727-745.

23. SeyedaliMirjalil, Seyed MohammadMirjalili, AndrewLewis, "Grey Wolf Optimizer", Advances in Engineering SoftwareVolume 69, March 2014, Pages 46-61, https://doi.org/10.1016/j.advengsoft.2013.12.007

24. Nykamp DQ, "Introduction to Taylor's theorem for multivariable functions."
From MathInsight. http://mathinsight.org/taylors_theor em_multivariable_introduction.

25. Tianqi Chen, Carlos Guestrin: XGBoost: A Scalable Tree Boosting System, https://arxiv.org/abs/1603.02754

26. Dimitris Leventis, XGBoost Mathematics Explained https://towardsdatascience.com/xgboost-mathematics-e xplained-58262530904a

27. Matthews, B. W. (1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". BiochimicaetBiophysicaActa (BBA) - Protein Structure. 405 (2): 442–451. doi:10.1016/0005-2795(75)90109-9

28. ENIREDDY, V., KUMAR, R.K. Improved cuckoo search with particle swarm optimization for classification of compressed images. Sadhana 40, 2271–2285 (2015). https://doi.org/10.1007/s12046-015-0440-0.

29. Vamsidhar, Enireddy & PhaniKumar, D.V.V.S. & Kishore, Dr. (2016). Application of Fisher Score and mRMR Techniques for Feature Selection in Compressed Medical Images. International Journal of Engineering and Technology. 7. 2109-2121.

30. Prakash K.B., Rangaswamy M.A.D., Raman A.R., Statistical interpretation for mining hybrid regional web documents,2012,Communications in Computer and Information Science,292,CCIS,503-512.

31. Ismail M., Prakash K.B., Rao M.N.,Collaborative filtering-based recommendation of online social voting,2018,International Journal of Engineering and Technology(UAE),7,3,1504-1507.

32. Prakash K.B., Rajaraman A.,Mining of Bilingual Indian Web Documents,2016,Procedia Computer Science,89,514-520.

33. Prakash K.B.,Content extraction studies using total distance algorithm,2017,Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016,7912085,673-679.

34. Prakash, K.B., Mining issues in traditional indian web documents,2015,Indian Journal of Science and Technology,8(32),1-11.

35. Prakash, K.B., Dorai Rangaswamy, M.A., Ananthan, T.V., Rajavarman, V.N.,Information extraction in unstructured multilingual web documents,2015, Indian Journal of Science and Technology,8(16).

36. Prakash, K.B., Rangaswamy, M.A.D., Raja Raman, A.,ANN for multi-lingual regional web communication,2012,Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),7667,LNCS,PART 5,pp-473-478.

37. Babitha, D., Ismail, M., Chowdhury, S., Govindaraj, R., & Prakash, K. B. (2020). Automated road safety surveillance system using hybrid cnn-lstm approach. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1767-1773. doi:10.30534/ijatcse/2020/132922020

38. Babitha, D., Jayasankar, T., Sriram, V. P., Sudhakar, S., & Prakash, K. B. (2020). Speech emotion recognition using state-of-art learning algorithms. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1340-1345. doi:10.30534/ijatcse/2020/67922020

39. Bharadwaj, Y. S. S., Rajaram, P., Sriram, V. P., Sudhakar, S., & Prakash, K. B. (2020). Effective handwritten digit recognition using deep convolution neural network. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1335-1339. doi:10.30534/ijatcse/2020/66922020

40. Prakash, K., Lakshmi Kalyani, N., Vadla, P. K., & Naga Pawan, Y. V. R. (2020). Analysis of mammography for identifying cancer cells using convolution neural networks. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1184-1188. doi:10.30534/ijatcse/2020/44922020

41. Reddy, A. V., Vege, H. K., & Prakash, K. B. (2020). Efficient and accurate hybrid deep learning model for multimodal disease risk prediction. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1262-1267. doi:10.30534/ijatcse/2020/55922020

42. Vadla, P. K., & Prakash, K. B. (2020). Residue based adaptive resource provisioning through multi-criteria decision and horizontal scaling of vm's in agent-based model for federated cloud. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1610-1622. doi:10.30534/ijatcse/2020/108922020

43. Prakash, K. B., Dorai Rangaswamy, M. A., & Ananthan, T. V. (2014). Feature extraction studies in a heterogeneous web world. International Journal of Applied Engineering Research, 9(22), 16571-16579.