Prediction of Plant diseases using feature optimization and Machine Learning Techniques

Dr. S. Sivakumar[1]

Assistant Professor, Department of Computer Science, Thanthai Hans Roever College (Autonomous), Perambalur, India

Dr. Sharmila Banu Sheik Imam[2]

Lecturer in CCSIT, King faisal university, Al Ahsa, Saudi Arabia, simam@kfu.edu.sa

Abstract

Plant diseases are adverse elements that significantly reduce crop quality and yield. Observing plants for disease with the naked eye is a common practice among seasoned biologists and farmers, although it can be inaccurate and take a lot of time. The purpose of this project is to create and build an intelligent classification system for plant diseases using feature optimization and artificial intelligence approaches. In this study, two approaches are used, and the results of their simulations are contrasted to assess performance. The Plant disease dataset is subjected to feature optimization utilizing optimization techniques in the first section. These features are classified using a Bayesian-optimized support vector machine classifier in the second half of this work, and the results obtained in terms of precision, sensitivity, f-score, and accuracy. The above-said methodologies will enable farmers all over the world to take early action to prevent their crops from becoming irreversibly damaged, thereby saving the world and themselves from a potential economic crisis.

Keywords: Plant diseases, optimization, Machine Learning, GWO

DOI Number: 10.14704/nq.2022.20.9.NQ440035 **Neuro Quantology 2022; 20(9):344-355**

1. Introduction

Agricultural productivity is something on which economy highly depends. This is the one of the reasons that disease detection in plants plays an important role in agriculture field, as having disease in plants are quite natural [1]. If proper care is not taken in this area then it causes serious effects on plants and due to which respective product quality, quantity or productivity is affected. For instance a disease named little leaf disease is a hazardous disease found in pine trees in United States. Detection of plant disease through some automatic technique is beneficial as it reduces a large work of monitoring in big farms of crops, and at very early stage itself it detects the symptoms of diseases i.e. when they appear on plant leaves.

Plant disease identification by visual way is more laborious task and at the same time, less accurate and can be done only in limited areas. Whereas if automatic detection technique is used it will take less efforts, less time and become more accurate. In plants, some general diseases seen are brown and yellow spots, early and late scorch, and others are fungal, viral and bacterial diseases. Therefore, the complete process sometimes takes so much time which become time taking for the disease to be eradicated and also very tedious in case of large areas [2]. Data analysis and optimization techniques are used for assessing pant disease and to classify the different types of diseases in the various crops. Advanced Techniques like similarity identification and machine learning based classification techniques are available for cotton crop disease predication in respect of time saving and exactness than the traditional methods, which will help the farmers to improve the quality of crops, also bringing a lessening of disease occurrence by early detection and timely curing them.

This work aims to focus on feature optimization and classification algorithms that are useful for effective prediction of plant diseases. This area is

of utmost importance because a good prediction method will significantly improve the cotton production. The feature optimization algorithms such as Particle Swarm Optimization (PSO), WHALE Optimization and Grey Wolf Optimization (GWO) Algorithms are applied for optimize the features then the supervised machine learning algorithms are applied in optimized feature set for classification process. In addition, this work tested several measures with given data in order to a priori determine the expected performance of the compared classifiers for plant disease datasets.

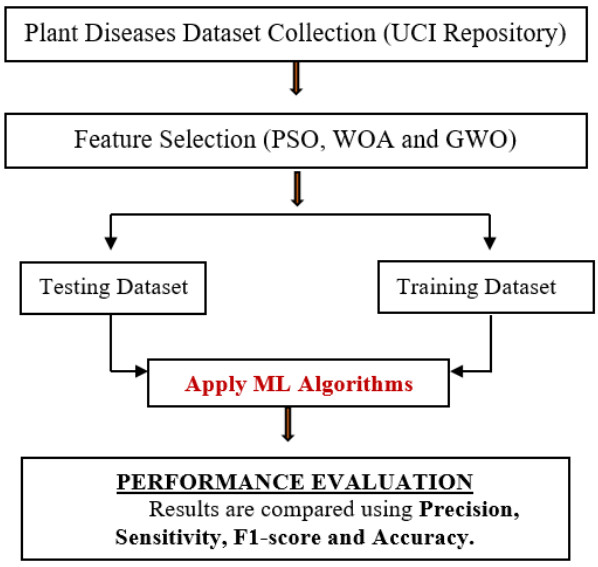


Fig.1. System Architecture

The above figure shows the block diagram for the proposed experimentation. The proposed architecture consists of Dataset collection, Feature Selection and classification model. The main aim of this research is to explore feature optimization based ML model and indicate their importance in the agricultural field. ML algorithms have large number of characteristics and these properties effects the performance of the model. The selection of significant features is the most challenging part. Indeed, it influences the performance of the algorithm by reducing the time and complexity of calculations. The work presented in this paper is based on the description of the pattern of the plant diseases, and the identification of the type and characteristic of the major disease problems during the recent years in India.

1.1 Contribution

The production and the quality of various products in agriculture are affected significantly by the various disease, which leads great financial loss for farmers. As the lack of basic knowledge of plant diseases prediction and protection, the farmers are yearning for the guidance from experts since the occurrence of new and large variety of diseases. An advanced computer based intelligent prediction system for plant Diseases could realize the diagnosing and early predication during the whole growing process.

2. Literature Survey

Programming digital computers to understand dataset is very difficult. To do this, an advanced computer aided system solves the problem with various techniques such as data mining and machine learning (Agrawal, A., et al., 2019) [3]. Due to the complex multi factorial nature of the cotton crop disease, it is very difficult to detect at early stage. Computer Aided prediction system that provides the better understanding of disease progression as well as can help in its early detection. To achieve a more reliable and accurate diagnosis, data mining and classification approaches have been developed in this research.

Pantazi et al. (2015) demonstrated that the combination of data fusion modelling and clustering methods was able to improve the quality of management zone delineation. Specifically, they compared k-means clustering with the Self-Organizing Map (SOM) for delineating management zones maps for variable-rate N application. Furthermore, a hybrid SOM algorithm in combination with k-means was compared with k-means in terms of cluster separation and management zone formation based on data fusion of NDVI and soil parameters [4].

The FAHP method is used to handle the various criteria included for the selection of the agricultural approach for turkey. Two fuzzy methodologies, AHP and ideal point has been compared for land suitability evaluations and fuzzy AHP was considered as better than ideal point. As the application areas are growing enormous, the FAHP become an impressive research area for many researchers. This approach has been applied successfully in different applications like job selection, producer

selection, personnel selection, measuring instrument selection, project selection, supplier selection and other applications [5]. The Fuzzy AHP is a best decision making and dominant tool in providing both quantitative and qualitative decisions (Yucenur G. N., 2017).

Godliver Owomugisha et al (2018) developed automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease "Color histograms are extracted and transformed from RGB to HSV, RGB to L*a*b. Peak components are used to create max tree, five shape attributes are used and area under the curve analysis is used for classification. They used nearest neighbors, Decision tree, random forest, extremely randomized tree, Naïve bayes and SVM classifier. In seven classifiers extremely, randomized trees yield a very high score, provide real time information provide flexibility to the application [6].

Benos et al., (2021) proposed a model to detect whether the crops are infected or not, and if they are further particular disease is identified according to the crop. The highest Accuracy achieved is for VGG16 i.e., 97.53 %. For the crops recommendation system, we have trained a Content-based filtering technique on the dataset. In which we are suggesting crops considering various parameters. By taking the user's location and the season for cultivation as an input they were recommending crops by training model on their collected dataset [7].

Bhagat et al. (2021) proposed a system for detecting plant leaf disease. SVM is a machine learning algorithm that is being utilized to generate a system for classifying and identifying plant diseases. The method would assist farmers in obtaining a valuable method for precise disease identification with less computational effort. The precision of the SVM classifier was 80%, and when it was combined with Grid Search hyperparameter tuning, it rise to 84% [8].

Mukhopadhyay et al. (2021) for identifying the disease area in tea leaves, image clustering based on the Non-dominated Sorting Genetic Algorithm (NSGA-II) is proposed. Then, for feature reduction and determining the disease in the tea leaves, PCA and multi-class SVM are used, respectively. The results show that the suggested algorithm has an average accuracy of 83% in detecting the type of disease present in tea leaves [9].

3. Disease Prediction Factors

Disease prediction in cotton plants is mainly executed to attain a good yield in the agriculture productivity. Many techniques have been implemented with simple predictions based on the past information and then integrating the computer technologies to afford inclusive information about the multi-dimensional analysis of the disease causing factors affecting the growth of the cotton plant [10]. Diagnosis of the disease in the plant at earlier stage can help the farmers in controlling the disease by providing pesticides over the cropping area. Machine learning algorithms are used to provide exact prediction for augmenting the cost-effective production systems. There are no present systems which approve the occurrence in cotton diseases based on multiple factors like climatic, physiological factor, management factors and so on. The various diseases of cotton crop in the dataset is listed below.

- Angular Leaf Spot or Black Arm Disease
- Vascular Wilt Disease
- Grey Mildew or Dahiya Disease
- Anthracnose Disease
- Root Rot Disease
- Boll Rot Disease
- Leaf Spot or Blight Disease
- Reddening or Lalya Disease
- Injury Disease

In this proposed system, various factors such as temperature, amount of rainfall, speed of the wind, moisture, Dew, leaf wetness, healthiness of the plant, crop geometry are considered by the monthly report that gives a throughput value for the specific field. In this study, three years data are collected and used for the training model. The data are pre-processed by removing NULL/NAN values and cleaning. The missing values in the monthly report are filled by the disjoint process to generate the series of data and also replaced by the average value of the subsequent data. The factors like rainfall, temperature for the unknown periods are counterfeited. The monthly data attained are evaluated with the cotton crop cultivation and provided a revealed efficacy in the prediction of diseases [11].

Attributes plays very important role in the classification process. Based on the value of the attribute class labels are identified. In our dataset

attributes are divided into two different categories. They are

i. ***Weather parameters:*** Temperature, Moisture, Wind, Rainfall, Dew, Leaf Wetness , Climatic Factors.

ii. ***Physic crop parameters:*** Plant Healthiness, Crop Geometry, Seed Treatment, Irrigation System, Weed management, Nutrient Management, Right Variety, Right Season, Tolerance Nature, Physiological Factors, Management Factors and Nature of Variety [12].

4. Feature Selection

The number of features in the retrieved features was reduced by deleting unnecessary and superfluous features to obtain acceptable classification accuracy using the feature selection approach. In order to attain better accuracy and faster convergence, feature selection is a common and important method in classification problems. This section covers optimizers for selecting the finest features from an obtained dataset [13]. PSO and whale feature selection methods are exposed here to compare the suggested optimizer's performance. The optimizers for successful feature selection are discussed in this section. The preliminary backdrop of the PSO, whale, and the working mechanism of the suggested optimizer is discussed using GWO approaches.

Feature selection will be transformed into a more reliable and appropriate form of input for the classifier to classify the various feature categories. The chaotic based GWO optimised characteristics for machine algorithms to categorise human activities were introduced in this research. Meanwhile, the suggested method has been compared to the machine learning algorithms PSO, GWO, and Whale in terms of accuracy.

i. PSO

It is an optimization method inspired by groups of birds. Respectively, every bird in the swarm regulates the search model depending on their newly acquired information [10]. The PSO algorithm is a hybrid of evolutionary algorithms and swarm system.The particles (Birds) of the swarm are free to fly across the multidimensional search space. Each particle acquires its own velocity and location throughout the voyage. The entire population is updated by updating each particle [13]. The particles gradually concentrate around the upper goal function value as the swarm arrangement forces itself toward it. The steps in particle swarm optimization are as follows:

Step 1: Initialization – The swarm particles are within the velocity and location ranges that have been pre-defined.

Step 2: Velocity Updating – The velocity of the swarm particles are determined at the end of each cycle by:

$$\vec{V}_i = W\vec{V}_i + c_1R_1(\vec{P}_{i,best} - \vec{P}_i) + c_2R_2(\vec{g}_{i,best} - \vec{P}_i)$$

Where $\vec{P}_i$ = position of particle 'i'
$\vec{V}_i$ = velocity of particle 'i'
$\vec{P}_{i,best}$ = finest position reached by the particle
$\vec{g}_{i,best}$ =best location remembered by the particle individual
'W' = parameter controlling the flying elements

R_1, R_2 =random numbers among 0 and 1
c_1, c_2 =cognitive learning factor and social learning factor

The PSO, or correctness in searching facility, is created by including variables from each particle. The weighting aspects c1, c2 prevent particle collision (individuals). To avoid collision, the velocity v and random number r are proven to be protected within the range supplied after updating particle I.

Step 3: Position updating - Because there is a time interval between repetitions, the positions of the particles change.

$$\vec{P}_i = \vec{P}_i + \vec{V}_i$$

$\vec{P}_i$ Must be checked and within the permitted range after refreshing.

Step 4: Memory updating – Using the formula, update $\vec{P}_{i,best}$ and $\vec{g}_{i,best}$

$$\vec{P}_{i,best} = \vec{P}_i \; if \; f(\vec{P}_i) > f(\vec{P}_{i,best})$$

$$\vec{g}_{i,best} = \vec{g}_i \; if \; f(\vec{g}_i) > f(\vec{g}_{i,best})$$

Where ($\vec{x}$) is the point function that has been extended.

Step 5: Checking of Destination **–** Once completed, the technique repeats steps 2 through 4 until definitive end states are attained, for a predetermined number of iterations. Estimating $\vec{g}_{i,best}$ and and $\vec{P}_{i,best}$yields the result.

PSO algorithms do not take fitness values into account. When the population is large, this is a significant computational benefit over GA. The real-number arithmetic operation is used to calculate velocity and position. The PSO is based on the ideas of altering velocity (accelerating) toward pbest (Personal Best) and lbest (Global Best) locations (local form of PSO) and accelerating toward pbest and lbest locations [15].

Personal Best: The best solution (fitness value) is effectively achieved when individual particle maintains the path of its coordinates in feature space. This value is called pbest.

Global Best: Particle, while maintaining all the features as its topological neighbor, this value is the gbest (global best). When a particle locates with reference to its neighbor it is called lbest.

ii. WHALE Optimization Algorithm

WOA has recently sparked a spike of attention. This hunt and progress estimate is based on a scientific model of humpback whale behaviour and development as they sustained their search for foodstuff and accommodations. The Bubble-net-Assaulting technique, in which whales attack fish by twisting air pockets around them down to 12 metres below the surface, then swimming back up to grasp and catch their victim, was the inspiration for this method.

In this computation, the inquiry technique is employed to address the unpredictable pursuit of food, which can be scientifically evaluated by recreating historical arrangements rather than randomly choosing unique configurations. Despite this peculiar behaviour, WOA differs from other improvement estimates in that it just requires the modification of two parameters. These traits make switching between the abuse and inquiry forms a breeze [15].

Fig.2. Humpback Whale Encircling Attack Prey Searching Methodology

In the next section, we'll go over the mathematical models of surrounding prey, pursuing prey, and spiral bubble-net foraging. Prey found nearby: Humpback whales update their position as they circle their prey by increasing the number of repeats from zero to maximum. This behaviour can be mathematically stated as:

$$\text{If } (p<0.5 \text{ and } \text{mod } (U) <1)$$

The following equations then update the location of the candidate position X (t+1).

$$D= \text{mod } \{(C.X)-X (t)\}$$

$$X (t+1) = [X (t) - \{U.D\}]$$

Where X (t+1) is the optimal location in the current scenario and p =0.1 (constant). The following equations are used to compute U and D.

$$U= \text{mod } \{2.a.r-a\}$$

$$C=2.r$$

Where r is a randomly selected vector and a decreases linearly from 2 to 0.

Searching of Prey: The random variables replace X in the prey search method. The following are the Xrandom and mathematical equations:

$$D= \text{mod } \{(C.X_{random})-X(t)\}$$

$$X(t+1) = [X_{random}(t) - \{U.D\}]$$

In the whale optimization algorithm's exploring phase [15], the prey was encircled and spiral updates were performed. The following equation gives the mathematical phrase for updating a new position throughout the spiral process.

$$X(t+1) = D^{l}.e^{bl}.\cos(2\pi l) + X^{*}(t)$$

The distance between the new and updated positions in the new generation is D, and b is a constant ranging from 0 to 1.

iii. Grey Wolf Optimization (GWO)

It's a new meta-heuristic optimization algorithm based on the social structure and cunning hunting strategy of grey wolves. In their natural habitat, grey wolves are frequently at the top of the food chain. Grey wolves live in packs of 5–12 individuals. Grey wolves, in particular, have a well-organized social system. A man and female wolf lead a grey wolf group (alpha). They make decisions for their pack, including as where to sleep, hunt, and wake up. The remainder of the pack must follow Alpha's lead.

The social order of grey wolves, on the other hand, is democratic (alpha may follow other individuals of the pack). People that keep their tails down in groups support the alpha's decision. It's also worth mentioning that the alpha doesn't necessarily have to be the pack's strongest member.

The major responsibility of the alpha is to lead the pack. A pack of grey wolves need discipline and organization. In the grey wolf social system, beta is the level below alpha, and its function is to aid alpha in making decisions. When one of the alpha wolves matures or dies, beta is the most likely candidate to succeed him. The beta has power over the alpha, which must be acknowledged. Beta serves as alpha's advisor and oversees the pack's discipline. Beta gives alpha feedback and reinforces alpha's directives.

Omega, the weakest rank in a grey wolf group, is made a scapegoat. Omega wolves are the final wolves allowed to eat and must obey the directives of other wolves [15]. Omega may appear to be the least important member of the pack, yet his absence could cause internal turmoil and other problems. This is due to the omega's role in channeling other wolves' rage and hatred, allowing other wolves to be satisfied and grey wolf supremacy to be maintained. Omega periodically looks after the pack's children. The remaining wolves are known as subordinates, in addition to alpha, beta, and omega (delta).

i. Tracking, chasing, and approaching the prey.
ii. Pursuing, encircling, and harassing the prey till it stops moving.
iii. Attacking the prey.

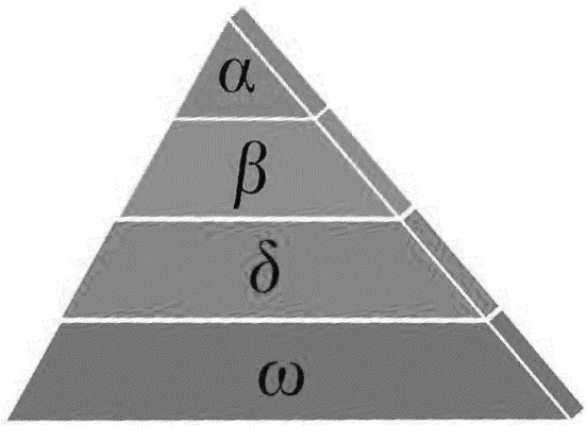


Fig.3. Social hierarchy of grey wolves

The graphic above depicts the common grading of grey wolves. In the pyramid, there are four hierarchical tiers. The pyramid's leader, alpha (), is at the top, followed by beta () and delta ().

Omega is at the bottom of the above pyramid (). At the most fundamental level, each of these types understands the subgroups. They must, for example, accept while yet having the ability to demand. As a result, in various ways, and are the pack's dominant wolves.

$$D^{t} = \left| C^{t} \cdot X_{p}^{t} - X^{t} \right|$$
$$X^{t+1} = X_{p}^{t} - A^{t} \cdot D^{t}$$

In the above equation, current iteration is represented as t, X is the grey wolf's position vector, and X_p is the prey's position vector. The estimated distance vectors between the prey and the grey wolf are represented by D. The coefficient vectors are A and C. A = 2ϕ • r1 − ϕ, C = 2 • r2. is gradually reduced from 2 to 0 throughout the course of the rounds. Members of r1 and r2 are picked at random from the range [0, 1].

The first three best solutions produced thus far are taken into account when designing GWO, and they are implemented accordingly [15]. The remaining search agents are thought to be. The optimization procedure is directed by, while following individual's dominant wolves to update their location. The following is the formula for hunting behavior:

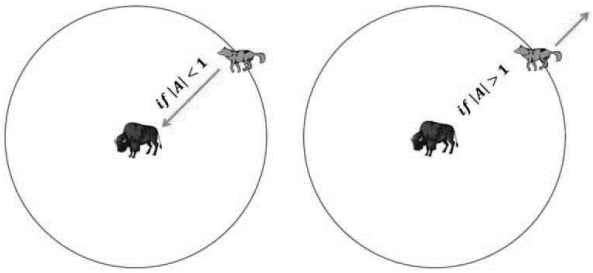
$$D_\alpha^t = \left|C_1^t \cdot X_\alpha^t - X^t\right|$$
$$D_\beta^t = \left|C_2^t \cdot X_\beta^t - X^t\right|$$
$$D_\delta^t = \left|C_3^t \cdot X_\delta^t - X^t\right|$$

$$X_1^t = X_\alpha^t - A_1^t \cdot D_\alpha^t$$
$$X_2^t = X_\beta^t - A_2^t \cdot D_\beta^t$$
$$X_3^t = X_\delta^t - A_3^t \cdot D_\delta^t$$
$$X^{t+1} = \left(X_1^t + X_2^t + X_3^t\right)/3$$

The position vectors of α, β, and δ are represented as X_α, X_β, and X_δ. D_α, D_β, and D_δ signify the approximate distances between the present solution α, β, and δ respectively. The estimated positions of the prey are based on the positions of and X1, X2, and X3. Finally, the algorithm takes into consideration all of the estimated positions by α, β, and δ, and to calculate the position of the prey.

The GWO optimization method begins with the creation of a random grey wolf population (candidate solutions). The wolves a, b, and d estimate the likely position of the prey across the iterations (optimum solution). Grey wolves constantly adjust their postures in relation to their prey. Parameter should be reduced from 2 to 0 near emphasis exploration and exploitation during the search process. Candidate solutions diverge from the prey if |A| > 1; if |A| < 1, candidate solutions converge to the prey. |A| < 1 forces the wolves to assault the prey, as shown in Fig. 3(a).

- If |A| < 1 then attacking prey
- If |A| > 1 then searching for prey



(a) **(b)**

Fig.4. Attack towards the prey

Grey wolves seek for prey based on the positions of the alpha, beta, and delta wolves. They separate up to hunt their prey before coming back together to attack. This system induces the search agent to diverge from the prey by using A with random values higher than 1 or less than -1 to mathematically represent divergence. This encourages discovery and allows the GWO algorithm to search the entire world. Grey wolves must diverge from their prey when |A| > 1 in order to find better prey, as shown in Figure 4. (b). The GWO algorithm's pseudo code is defined in the steps below.

Input: Features, GWO Parameters
Output: Best Solution

Begin
Initialize the grey wolf population X_i (i = 1, 2… n)
Initialize a, A and C
Calculate the fitness of each search agent
X_α = the best search agent
X_β = the second best search agent
X_δ = the third best search agent
While (t < Max number of iterations)
For each search agent
Update the position of the current search agent
End for
Update a, A and C
Calculate the fitness of all search agents
Update X_α, X_β, and X_δ
t = t+ 1
End while

Return X_α

End

This process continues until the halting requirements are met, at which point the GWO algorithm is terminated.

5. Machine Learning (ML) Techniques

Machine learning investigates how computers can learn based on data. ML is an interdisciplinary study, which has focused on algorithm design and computers utilize these algorithm for learning purpose. Learning is nothing but learning from the feature dataset. Commonly, ML system are designed and implemented in that manner, they have permitted the expert system to produce the solution for the diagnostic problem by using previous data. The various learning techniques existing for the classification task such as supervised classification, unsupervised classification and reinforcement learning. Classification is a one of the supervised learning processes and the classification tool is used to predict the target class.

In this research, supervised ML algorithms are used to solve the tumor classification as a regular classification problem that makes use of scientific features. We have a set of training records D={X_1, X_2... X_n} where each record labeled to a class 'A' and class 'B'. The classification model is related to the features in the testing record to one of the class labels. Then for a given instance/features of unknown class, the model is used to predict a class label for it. The supervised learning methods based on the existence of labeled training features of the input image.

Supervised Classification

Supervised learning involves learning a mapping between a set of input variables X and an output variable Y and applying this mapping to predict the outputs for unseen data. The majority of practical ML techniques uses supervised learning. All data is labeled and algorithms learn to predict the output from the input data. This research applies the some ML classification algorithms such as NB, KNN and LR and demonstrates all classification algorithm's performance on selected features.

- Naïve Bayes (NB) Algorithm
- K-Nearest Neighbor (KNN)
- Logistic Regression (LR)

6. Result and Discussion

Proposed classification accuracy of the system is calculated with the various features. We perform the disease identification task for various plants with various properties (features). The performance of the proposed algorithms has been evaluated by different cases which are shown in the below section. The features used in this experiments can be seen in below,

1. *Weed Management*
2. *Nutrient Management*
3. *Right Variety*
4. *Right Season*
5. *Tolerance Nature*
6. *Climatic Factors*
7. *Physiological Factors*
8. *Management Factors*
9. *Nature of Variety*
10. *Temperature*
11. *Moisture*
12. *Wind*
13. *Rainfall*
14. *Dew*
15. *Leaf Wetness*
16. *Plant Healthiness*
17. *Crop Geometry*
18. *Seed Treatment*
19. *Irrigation System*

This dataset consists in a collection of various features extracted from leaf specimens originating from a total of 40 different plant species, which is downloaded from UCI repository.

6.1 Experimentation details:

On a 64-bit macOS system with an Intel 2.6GHz 8-core i7 processor, 16GB DDR4 RAM, and a Radeon Pro 560X 4GB GPU, the test is run. All of the programmes use Python 3.8 and run on the MATLAB 2019a platform.

6.2 Performance Metrics and Evaluation:

The evaluation is carried out for the different algorithms with the following parameters such as TP, FP, TN, FN, and Precision, Recall, F1_Score and classification accuracy [15]. TP is True Positive Values, TN is True Negative Values, FP is False Positive and FN is False Negative values.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Fig.5. Confusion Matrix Format

Metrics are calculated to evaluate the performance of the proposed architecture, as given in the table below.

Table.2. Performance Metrics

SL.NO	Performance Metrics	Mathematical Expression
01	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
02	Sensitivity or recall	$\frac{TP}{TP+FN}$ x100
03	Specificity	$\frac{TN}{TN+FP}$
04	Precision	$\frac{TN}{TP+FP}$
05	F1-Score	$2.\frac{Precison * Recall}{Precision + Recall}$

5.3 Results and Findings:

In the suggested architecture, various feature selection and ML techniques are applied. Metrics are calculated to evaluate the performance of the proposed architecture, as given in the table below.

Table.3. Performance Analysis with Feature selection Algorithms

Algorithm Details	Performance Metrics				
	Accuracy	Sensitivity	Specificity	Precision	F1-Score
NB	81.5	80.4	81.0	79.6	80.4
KNN	84.5	82.5	80.5	82.3	82.7
LR	85.56	83.67	81.89	83.37	83
NB-PSO	82.6	82.45	82.74	80.46	81.78
KNN-PSO	85.14	83.5	81.78	83.67	83.23
LR-PSO	86.13	84.1	82.87	84.51	84.15
NB-WOA	83.76	83.9	83.61	81.78	82.67

KNN-WOA	86.22	84.55	82.0	84.6	84.78
LR-WOA	87.8	85.85	83.86	85.89	85.88
NB-GWO	88.89	86.85	86.86	85.89	85.88
KNN-GWO	89.5	88.56	87.5	86.56	86.45
LR-GWO	92.9	91.8	92.23	91. 89	91.88

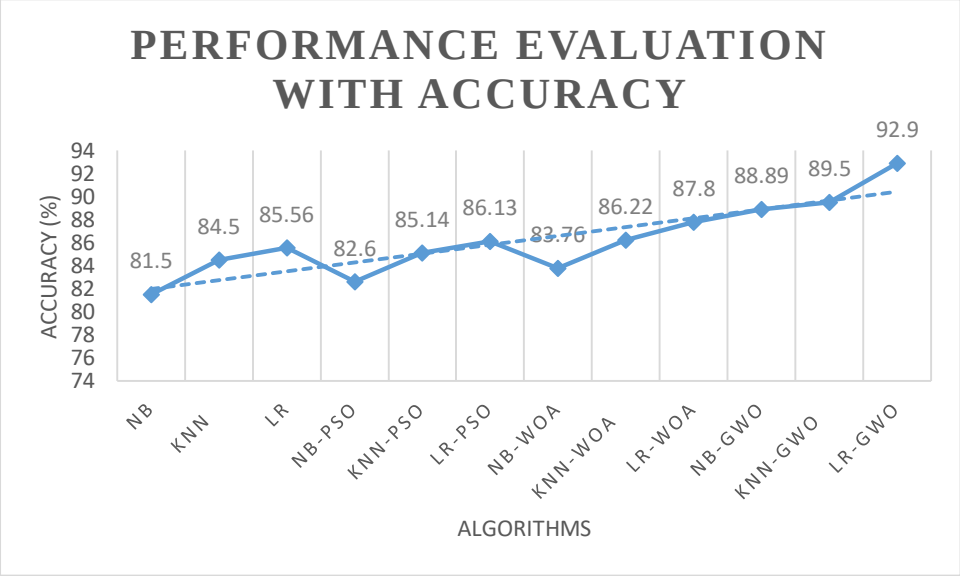
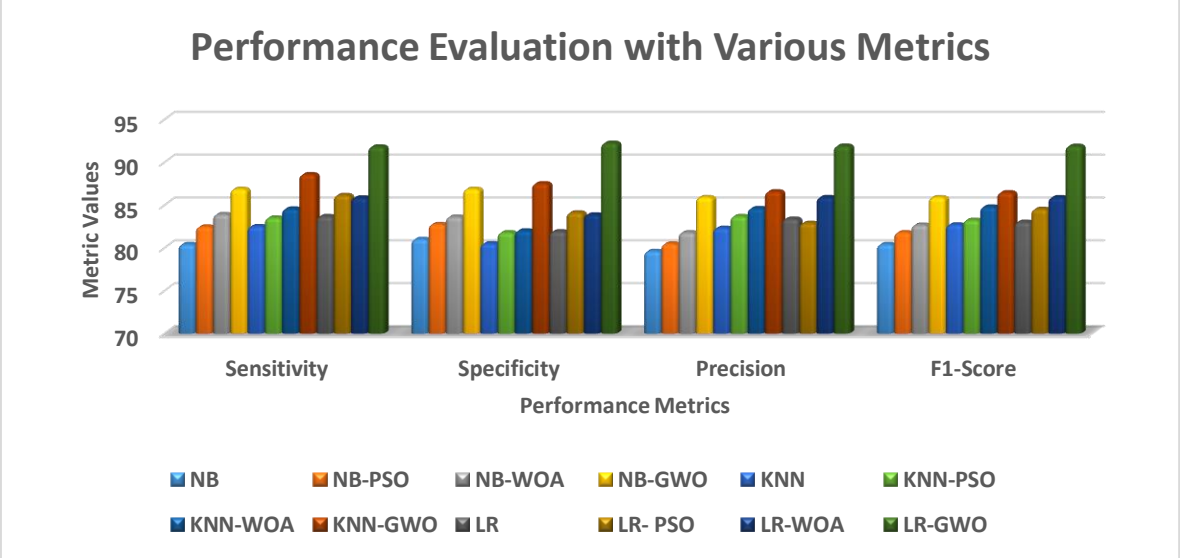


Fig.6. Performance Analysis with FS-ML Algorithms



From the statistics, it is obvious that the LR-GWO method has a higher detection ratio than the other available algorithms. The suggested algorithm clearly beat the other existing algorithms, as seen in the graph.

7. Conclusion

The experimental findings show that the LR-GWO feature optimization approach outperforms the other algorithms. To evaluate the performance of feature selection methods, a few machine learning techniques are applied to selected features. According to the results, numerous ML methods perform well in the majority of classification problems for this dataset. As a result, the next phase in this research will be to investigate the performance of novel machine learning methods. This research shows that the LR-GWO algorithm can perform well with adequate feature tuning and statistical testing to determine the relevance of the results. As a result, machine learning algorithms are crucial in the development of intelligent expert systems that improve accuracy, detection, and reduce false positives. The various machine learning techniques utilized for plant disease detection system such as SVM, KNN, NB and so on.,. In this research, the dataset size is larger. So, work need to be improved by build an accurate models with novel machine learning models.

Reference:

1. Bhagat, M., Kumar, D. A comprehensive survey on leaf disease identification & classification. Springer, Multimed Tools Appl (2022).
2. Kaur, S., Pandey, S. & Goel, S. Plants Disease Identification and Classification Through Leaf Images: A Survey. Arch Computat Methods Eng 26, Springer, 507–530 (2019).
3. Agrawal, A., Gans, J., & Goldfarb, A. (2019). The impact of artificial intelligence on innovation. In: The economics of artificial intelligence. University of Chicago Press. pp 115–148.
4. Pantazi, X.; Moshou, D.; Mouazen, A.; Alexandridis, T.; Kuang, B. Data fusion of proximal soil sensing and remote crop sensing for the delineation of management zones in arable crop precision farming. In Proceedings of the CEUR Workshop, Kavala, Greece, 17–20 September 2015; pp. 765–776.
5. Yucenur G. N. (2017). A producer selection problem: A case study from Turkish food industry. Beykent university journal of science and engineering, 10(2), 185–212, doi: 10.20854/bujse.316308.
6. Godliver Owomugisha,Ephraim Nuwamanya, John A. Quinn, Michael Biehl, Ernest Mwebaze "Early detection of plant diseases using spectral data" APPIS 2020: Proceedings of the 3rd International Conference on Applications of Intelligent Systems, 2019,Pages 1– 6.
7. Benos, Lefteris, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, and Dionysis Bochtis. 2021. "Machine Learning in Agriculture: A Comprehensive Updated Review" Sensors 21, no. 11: 3758.
8. Bhagat, S.K., Paramasivan, M., Al-Mukhtar, M. et al. Prediction of lead (Pb) adsorption on attapulgite clay using the feasibility of data intelligence models. Springer, Environ Sci Pollut Res 28, 31670–31688 (2021).
9. Mukhopadhyay, S., Paul, M., Pal, R. et al. Tea leaf disease detection using multi-objective image segmentation. Springer, Multimed Tools Appl 80, 753–771 (2021).
10. Singh, A., Kaur, H. (2022). Comparative Study on Identification and Classification of Plant Diseases with the Support of Transfer Learning. In: Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, vol 1387. Springer, Singapore.
11. Fenu, G.; Malloci, F.M. Forecasting Plant and Crop Disease: An Explorative Study on Current Algorithms. Big Data Cogn. Comput. 2021, 5, 2.
12. Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant Disease Detection and Classification by Deep Learning. Plants 2019, 8, 468.
13. El-Shafiey, M.G., Hagag, A., El-Dahshan, ES.A. et al. A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. Multimed Tools Appl 81, 18155–18179 (2022).
14. A. Pahurkar and R. Deshmukh, "PDMBM: Design of a High-Efficiency Plant Disease Classification Method using Multiparametric Bio Inspired Modelling," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 1607-1615.
15. A. Pahurkar and R. Deshmukh, "PDMBM: Design of a High-Efficiency Plant Disease Classification Method using Multiparametric Bio Inspired Modelling," IEEE, International

Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 1607-1615.
16. Kraiem, H.; Aymen, F.; Yahya, L.; Triviño, A.; Alharthi, M.; Ghoneim, S.S.M. A Comparison between Particle Swarm and Grey Wolf Optimization Algorithms for Improving the Battery Autonomy in a Photovoltaic System. Appl. Sci. 2021, 11, 7732.