[1]Rahul Kumar

[2]Rajeev Paulus

[3]Bireshwar Dass Mazumdar

Metaheuristic Methods for Predicting and Classifying Crop Diseases using Machine Learning

Journal of Electrical Systems

Abstract: - Lack of knowledge frequently causes delays in treating agricultural diseases, which can deteriorate the situation. The results of several researchers' use of machine learning techniques to crop disease diagnosis are promising. Certain crop illnesses are predicted by some approaches, whereas a broad range of crop diseases are predicted by others. To indicate the class label, all the target attributes are used. Predictive qualities that are not un-labelled are only present in labelled features. The primary objective of the classification process is to use categorization models developed from labelled examples (historical data) to accurately classify un labelled data. Making a training model with the right class (or goal values) to develop it is the first stage. Several Machine Learning methods are used to construct an autonomous system for crop disease identification. utilizing the PSO-SVM classifier, classification is carried out utilizing the features that were taken out of the crop dataset. The suggested systems diagnose the illness 100% of the time. Several Machine Learning methods are used to construct an autonomous system for crop disease identification. utilizing the PSO-SVM classifier, classification is carried out utilizing the features that were taken out of the crop dataset. The suggested systems diagnose the illness 100% of the time.

Keywords: Deep Learning, Logistic regression, PSO-SVM, Machine Learning, Supervised Learning.

I. INTRODUCTION

The lack of knowledge on agricultural disease causes worsening circumstances. The results of several researchers' applications of machine learning integrated AI approaches to crop disease diagnosis are promising. Certain crop diseases are predicted by some approaches of machine learning, whereas a broad range of crop diseases are predicted by others conventional techniques. A common job in agricultural science is data collecting and classification, which is also a computationally challenging task in computer science [1]. Creating an efficient strategy to combat crop disease requires analyzing a sizable data set.

Metaheuristic approaches are commonly used to solve optimization difficulties. Computational models can be made more accurate by using metaheuristic techniques. Prediction, severity assessment, and early disease diagnosis are common applications of Metaheuristic methods. Divyanshu Varshney et al. introduced machine learning solutions for plant disease identification in 2021 [2]. They said that the economies of agricultural continents depend on the early identification of plant diseases in agriculture, which is essential to protecting the crops. Farmers should treat illnesses appropriately based on AI based prediction results [3]. While human efforts, such as necked eye examinations, require expertise, they are not a viable strategy or remedy. As a result, we need an automated, speedy, precise, and affordable solution. Numerous applications have been created for things like leaf identification, fruit disease diagnosis, and lead ailment detection. Using image analysis and artificial intelligence, they provided a study of different approaches for crop disease prevention and treatment programs to simply, randomly, automatically, and accurately diagnose and identify plant leaf disease. Reports state that the KNN machine learning approach can identify and select diseases with an accuracy rate of 96.76 percent. This simple method queries instances using K-nearest neighbor to classify examples.

Simple machine Learning methods can be used to classify diseases such as canker, anthracnose, leaf label, bacterial blight, and alternate. Naive Bayes is among the most straightforward and successful classification techniques. Naive Bayes method provides the probabilistic model. It is therefore capable of forecasting depending on the probability

[1] *Corresponding author: VIAET, Sam Higginbottom University of Agriculture Technology and Sciences, Naini, Prayagraj, rahkam@gmail.com

[2] Professor, VIAET, Sam Higginbottom University of Agriculture Technology and Sciences, Naini, Prayagraj,rajeev.paulus@shiats.edu.in

of specific objects. Gathering information, color transformation, segmentation, and feature extraction are the important processes in these methods.

Support vector machines, or SVMs, are a technique for two-group classification problems. SVM uses an activation function, a mathematical equation, to compare new data with the best image from the learning pattern to estimate the mark of an intangible.

When it comes to disease identification, SVM outperforms in comparison with other two classifiers Random Forest and Logistic Regression in terms of outcomes. Artificial neural networks, or ANNs for short, are said to be used as computer components to drive analysis that is hard or impossible to answer by human or mathematical standards. Symptoms of prevalent diseases are the biggest obstacles to the limitations of identification processes, and the suggested solutions depend on optimal capture settings in order to work.

In 2020, Lovemore Chipindu et al. presented a maize kernel abortion recognition and classification experiment using deep convolutional neural networks and a binary classification machine learning algorithm [6]. Metaheuristics The field of engineering, operation research, bioinformatics etc. are generating a huge number of information day by day. These vast number of data leads to the optimization problem. But unfortunately, most of them are unsolvable in a polynomial time, and for that, these are determined as the NP-hard problems. Thus, very few of these types of problem can be answered by usual exact mathematical or numerical techniques [11]. But without giving up, the researchers think of a possible approximation technique which can find a quality solution in an equitable time. These approximation techniques are categorized as heuristic and metaheuristic. The main difference between these two methods is that metaheuristics are more problem independent than heuristics techniques. It can also be said that the metaheuristics can be applied to almost all types of the optimization problem and not like the heuristics, which can be applied to a specific problem only.

In simple word, metaheuristics can be defined as an efficient guide to the searching process to find (near)optimal solution. The term “Metaheuristics” was first introduced by Glover F. in 1986 [8] to explain the heuristic technique but without problem-specific nature. Precisely it can be said that metaheuristics are the high-level, heuristic-based soft-computing algorithms that can be applied to different types of optimization problem by detecting the general structure of a specific problem with few desired modifications [12]. In metaheuristics is not guaranteed that optimal solutions can be found by the algorithm, but in a reasonable amount of time, a quality solution of a difficult optimization problem can be found. In metaheuristics, it is assumed that the algorithms will work as desired most of the time but not every time.

Basically, metaheuristics are algorithmic architecture which can be applied to heterogeneous optimization problems with few modifications to solve the given problem. Some fundamental characteristics of metaheuristics are summarized below [9]

Metaheuristic are not for only one or a particular problem.

- This is not precise but more approximate.
- Metaheuristics are to found ‘good enough’ or quality solution.
- Metaheuristics can be extended to the advance learning technique from basic local search.

Getting the bi-cluster from a gene data matrix is “NP-hard”. For that reason, normal or conventional mathematical ways are ineffective on the bi-clustering problem. But now many metaheuristics are available to address this problem. However, metaheuristics will not guaranty to provide accurate or perfectly desired results but within a sensible amount of time, an efficient result can be discovered.

II. LITERATURE SURVEY

The authors in [8] constructed a classifier model that back-eliminates disease datasets. Experience has shown that the feature selections have enhanced categorization methods and decreased the number of inputs. Only a 25% reduction in the size of the dataset used in this study was necessary to get the desired results. This new dataset contains only four different features and is 83% better than the previous data. Redundant characteristics have been demonstrated to enhance classifier performance in earlier studies.

A fuzzy master system based on the surge development approach is described [5]. This approach was specifically created to handle UCI machinery datasets. To determine the most crucial elements for the best possible diagnosis and course of treatment, decision tree algorithms are employed. The output data is generated by fuzzy rules. The

result is obtained by the use of fuzzy approximation. The accuracy rate of an expert system that utilizes the particle swarm optimization method is 93.27%. When opposed to other classification techniques, which make it challenging to comprehend the output model provided by fuzzy expert systems, this system has a significant benefit.

Authors in [3] suggested a firefly-based approach based on rough sets as the basis for a precise prediction system. Including both imprecise and approximate theoretical concepts can help to mitigate the high level of complexity and uncertainty that come with crop disease datasets. The roughest-based fuzzy learning approach makes it possible to identify the best solutions with the least amount of computer power. These outcomes are unmatched by support vector machines and artificial neural networks for the prognosis of crop-related diseases and the prescription of medications.

Crop pathologists and crop breeders have found managing disease in crops to be a difficult area of work. A thorough investigation of the epiphytotic aspects of illness is necessary. It has been noted that the sickness is uncontrollable once it has spread throughout the field. As a result, the majority of advised management techniques focus on preventative efforts to lessen pathogen accumulation in the field [4]. Researchers emphasized a Naïve Bayes classification method for illness detection. To accurately predict and diagnose disease, the Naive Bayes classifier is utilized in conjunction with statistical approaches. It manages the intricate and large-scale data collection process by using data preparation techniques. A discretization technique is used to classify crop diseases. In this instance, the discretization method employed was directed variation with equal frequencies. The results demonstrate that compared to other methods, our strategy offers more precise measurements [5].

The linear SVM classifier model operates as described in [8]. This difference classifier outputs an ideal hyperplane after isolating a hyperplane using categorized instructional tests from a provided dataset. This allows for a further classification of the newly produced instances of the input data model [15]. A line that divides a hyperplane into two sections in two dimensions is called a hyperplane. Every class is on both sides of the divides. SVM, to put it briefly, divides classes.

In an innovative study, the authors of [6] used deep learning and a linear SVM classifier. In deep convolutional networks, a linear combinational machine takes the place of the soft-max layer. Because margin-based loss is more effective than cross-entropy loss, it may be employed in its place. Studies that have been published claim that the SVM can be used to yield different layers of a deep complex network. SVM is swapped out for a deep convolution network on the second layer.

The statistical method is examined in detail by the authors of [7] using logistic regression data analysis. A statistical method for assessing two dependent variables is called logistic degeneracy. Logistic regression is an effective regression analysis technique. Logistic regression techniques are employed to estimate the parameters of the logistic model. Independent components are used by logistic models to determine the likelihood that an event will occur [14].

Logistic regression models yield the most accurate categorization outcomes and are widely applied across numerous domains. Using this method, predictive models for crop disease are regularly assessed. By avoiding overfitting, this strategy produces more accurate results. Conversely, nonlinear connections increase complexity and time requirements. Furthermore, this method works well as an assessment tool for healthcare organizations as opposed to a classification model.

The best results were found when K-means and a priori were combined [10]. Using k-means collecting, the dataset is initially obtained. The most regularly recurring item sets are then found using the a priori approach. The Boolean association rule employs a "bottom-up" strategy to produce better outcomes. The heart disease prediction system's real-world scenarios pose a series of difficult queries to patterns. For the purpose of effectively identifying and mapping the input data, predictive analysis depends on classification. The data can be divided into two categories: tagged data and untagged data. The labelled data include many predictive factors in addition to the single target quality. The class label is indicated using all the target characteristics. Predictive qualities that are not unlabeled are only present in labelled features. The primary objective of the classification process is to use categorization models developed from labelled instances (historical data) to accurately classify not-labelled data. Making a training model with the right class (or goal values) to develop it is the first stage.

III. METHODOLOGY

An example of how crop disease might be anticipated using genetics is shown in Figure 1. A database of crop disease cases forms the basis of this technique. The process known as PCA is used to extract features from data. Several machine learning techniques are applied to classify the data. Disease prediction can be enhanced with the help of the combination of machine learning and metaheuristics methods. Healthy and unhealthy crop leaf datasets are categorized based on the derived textural properties. Only significant characteristics with sufficient information for classification will be chosen; in certain cases, all the features may not be significant for classification. In these cases, the features with little to no predictive information are removed. In this case, the illnesses are categorized according to the predetermined guidelines provided for doing so. This work applies a linearly separable condition to fourteen texture-based characteristics that are utilized as input for a Support Vector Machine (SVM) classifier. In essence, SVM hyperplane reduces mistakes by classifying the data iteratively. As a result, the training vectors with labels will maximize the figure's margin of separation between the two classes.

The ID-3 method is the first ever developed conclusion tree-based strategy. Measures of information gain and entropy are employed in this approach. Starting with a nodule, the entropy of the functional attributes is computed iteratively. The subsets of a dataset that have been split according to the feature that has the highest information gain and the lowest error rate (entropy) are referred to as split attributes. The algorithm iterates over each subset of data if its target classes are not accurately classified. The last subset of a branch's nonterminal nodes is known as the terminal nodes in a decision tree. Class labels are indicated by terminal nodes, while nonterminal nodes are specified by the split attribute. Early detection of cardiac issues is possible with an ID-3-based conclusion tree approach described in [8].

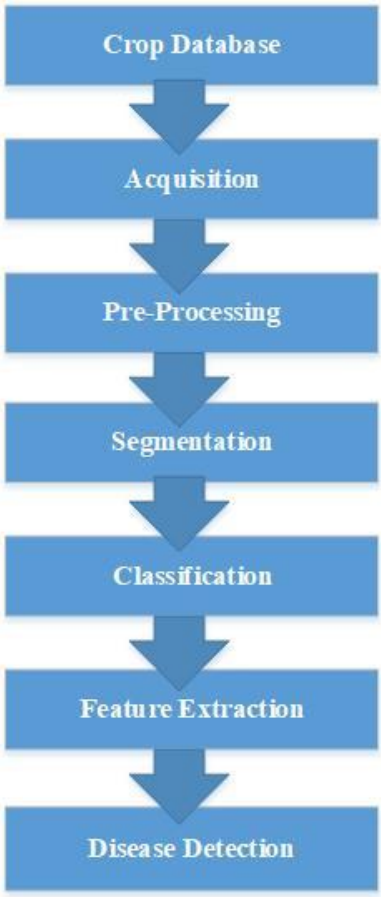

Fig. 1. Machine Learning and metaheuristic

The most reliable technique in the field of supervised machine learning techniques, according to Peterson et al. [13], is to use a K-nearest neighbor strategy for pattern discovery and data categorization. K-nearest neighbor algorithms use similarity metrics or distance functions. The K-Nearest Neighbors (KNN) algorithm is a supervised machine

learning method employed to tackle classification and regression problems. A similarity metric is used to categorize newly defined instances, and each instance is retained. It utilizes the instance-based learning approach to efficiently classify. Every new instance of the dataset is given a category based on the votes of the classes that are right next to it. The distance measure is calculated using the training and testing datasets. Once k has been selected, the algorithm calculates the distance between the two instances. The pseudocode of the KNN classification algorithm is provided in algorithm 1

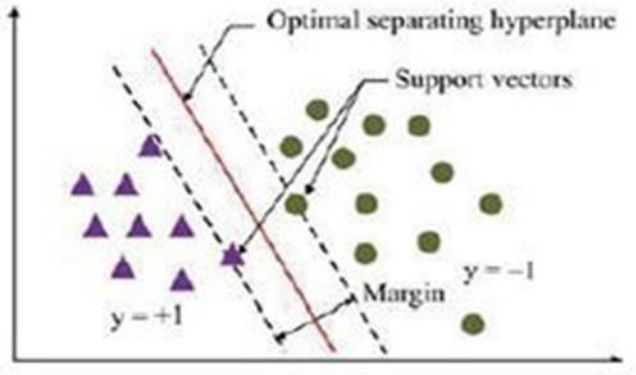

Fig. 2. SVM in linearly separable condition methods for classification and detection of crop diseases data

Algorithm 1 KNN classification Algorithm

- **Given:**

  -training examples {x_i, y_i}

  x_i……attribute-value representation of examples

  y_i…….class label: {ham,spam}, digit {0,1,…..9} etc.

  -testing point x that we want to classify

- **Algorithm:**
  - Compute distance D (x_i, x_i) to every trainig examples xi
  - Select k closest instances x_{i1}……x_{ik} and their labels y_{i1}…..y_{ik}
  - Output the class y* which is most frequent in y_{i1}……y_{ik}

The ideal option is to use a PSO-SVM (particle swarm optimization) support vector machine as a non-probabilistic binary linear classification method [6]. Samples can be sorted into one or more target classes using this method. Every item of data is represented by a single point. It grows wider with every new group due to the obvious differences. The new occasions' target classes are remapped based on which side of the space they occupy. The absence of tags in the input datasets can lead to nonlinear categorization. Since the instances can't be assigned to target classes, the support vector machine uses an apart learning method to classify the data. Once the function-based clusters are built, further instances are added. The flow diagram of PSO-SVM model construction is shown in Figure 3.

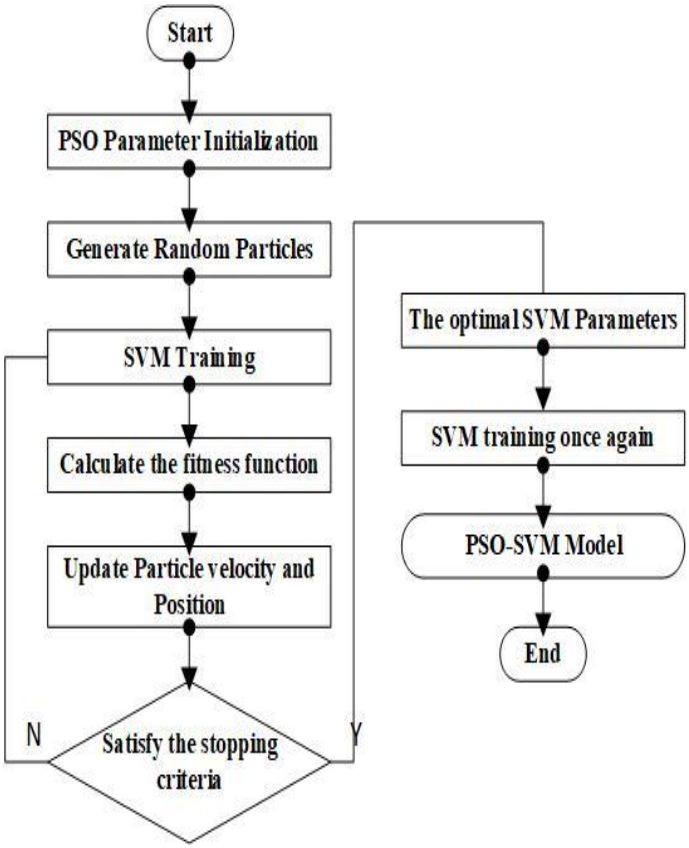

Fig 3: The flow diagram of PSO-SVM

The process involves initializing particles and setting PSO parameters, evaluating their fitness, updating their velocity and position, checking stopping criteria, and determining optimal parameters. The initial particles are composed of SVM parameters, while the PSO parameters include population size, maximum iteration, inertia weight, and acceleration coefficients. The fitness value is then calculated, and the optimal parameters are obtained. The training and verification process is completed, and the PSO-SVM model is constructed.

IV. RESULT AND ANALYSIS

We carried out analytical research using data from the crop dataset [5]. The crop database contains 175 records that are used as input by the ID-3, KNN, and SVM algorithms. The incoming data collection is pre-processed using Weka. Because of this preprocessing, the data is now more accurately represented. There were 175 cases total; 150 of those were used to train machine learning predictors, while the remaining 25 were utilized to test the predictions. The performance of machine learning predictors based on various performance comparison factors is displayed in Figures 4–5.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

Where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

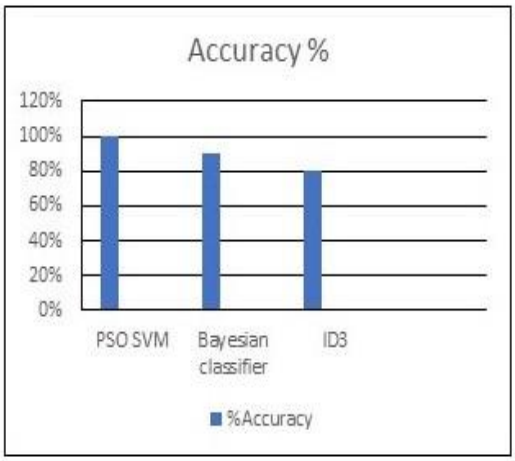

Fig. 4. Accuracy of different algorithms for crop disease detection

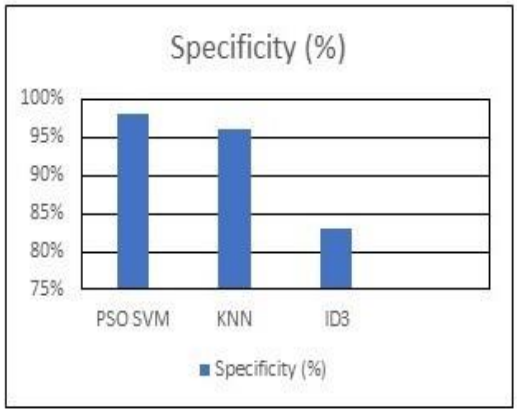

Fig. 5. Specificity of classifiers for crop disease dataset

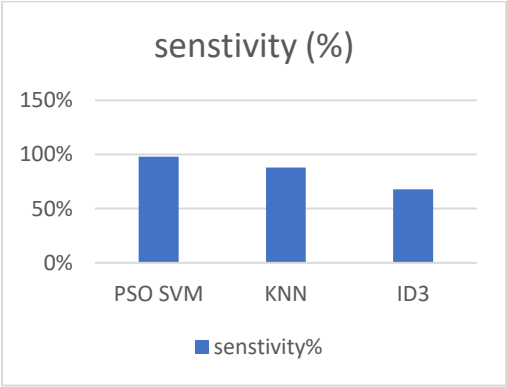

Fig. 6. Sensitivity of classifiers for crop disease dataset

V. CONCLUSION

Optimization problems often solved with the use of metaheuristics methods. Several machine learning methods are used to construct an autonomous system for crop disease identification. Utilizing the PSO-SVM classifier

classification is carried out utilizing the features that were taken out of the crop data set. The suggested systems diagnose the illness 98% accurate.

The application of metaheuristic methods can increase the precision of computing models. Artificial Intelligence is widely used in prediction, severity rating and early disease diagnosis. To detect crop disease early and accurately, this study offered machine Learning and Metaheuristic techniques. Compared to other machine Learning predictors the PSO-SVM performs better.

DATA AVAILABILITY

The data used to support the findings of the study can be obtained from the corresponding author upon request.

Table 1. Experimental Descriptions

Definition	**Input Data**	**Output**
True Negative	Healthy	Healthy
True Positive	Unhealthy	Unhealthy
False Positive	Healthy	Unhealthy
False Negative	Unhealthy	Healthy

Table 2. Experimental parameters and Values

Parameter	**Value**
Total Image Used	175
Trained Image	150
Tested Image	25
Failed Case	03
Successful Case	22
True Negative	10
True Positive	12
False Positive	02
False Negative	01
Sensitivity	98%
Specificity	98%
Accuracy	99%

REFERENCES

[1] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and electronics in agriculture, 147, 70-90.

[2] Divyanshu Varshney Burhanuddin Babukhanwala Javed Khan, “Machine Learning Techniques for Plant Disease Detection”, IEEE Xplore Part Number: CFP21J32-ART; ISBN:978-1-6654-1571-2.

[3] Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Computers and Electronics in Agriculture, 198, 107119.

[4] H. Z. Almarzouki, H. Alsulami, A. Rizwan, M. S. Basingab, H. Bukhari, and M. Shabaz, “An internet of medical things-based model for real-time monitoring and averting stroke sensors,” Journal of Healthcare Engineering, vol. 2021, Article ID 1233166, 9 pages, 2021.

[5] K. G. Liakos; P. Busato; D. Moshou; S. Pearson; D. Bochtis, “Machine learning in agriculture: A review”, Sensors 2018, 18, 2674.

[6] K.P. Ferentinos, "Deep learning models for plant disease detection and diagnosis" Computers and Electronics in Agriculture, vol. 145, pp. 311-318,2018.

[7] M. A. Hussein and A. H. Abbas, "Plant Leaf Disease Detection Using Support Vector Machine, "Al- Mustansiriyah Journal of Science, vol. 30, no. 1, pp. 105- 110,2019.

[8] Sörensen, K., & Glover, F. (2013). Metaheuristics. Encyclopedia of operations research and management science, 62960-970.

[9] Karimi-Mamaghan, M., Mohammadi, M., Meyer, P., Karimi-Mamaghan, A. M., & Talbi, E. G. (2022). Machine learning

at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art. European Journal of Operational Research, 296(2), 393-422.

[10] Lovemore Chipindu, Walter Mupangwa, Jihad Mtsilizah, Isaiah Nyagumbo and Mainassara Zaman-Allah "Maize Kernel Abortion Recognition and Classification Using Binary, Classification Machine Learning Algorithms and Deep convolutional Neural Networks",

[11] S. Sharma. Kaur and N. Dhillon, "Plant Disease Classification with KNN-SVM classification," International Journal of Advance Engineering and Research Development, vol. 5, no. 05, pp. 752-758, 2018.

[12] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning-based recommender system," ACM Computing Surveys, vol. 52, no. 1, pp. 1–38, 2020.

[13] Nematzadeh, S., Kiani, F., Torkamanian-Afshar, M., & Aydin, N. "Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases". Computational biology and chemistry, 97, 107619 ,2022.

[14] Shinde, P.P., Oza, K.S. & Kamat, R.K. RETRACTED ARTICLE: Systematic acuity of medicinal big data: need of health industry. Pers Ubiquit Comput 27, 941–954 (2023).

[15] Elia Ramirez-Asis, Magna Guzman-Avalos, Bireshwar Dass Mazumdar, D Lakshmi Padmaja, Manmohan Mishra, Deepali S Hirolikar, Karthikeyan Kaliyaperumal, "Metaheuristic Methods for Efficiently Predicting and Classifying Real Life Heart Disease Data Using Machine Learning", Mathematical Problems in Engineering, vol. 2022, Article ID 4824323, 5 pages, 2022.