Survey of Machine Learning Algorithms for Plant Diseases Detection

[1]Nitish Kumar, [2]Keshav Kishore, [3]Ramesh Kumar

[1]Research Scholar, [2]IT-HEAD, [3]HOD, School of Advanced Computing, *Alakh Prakash Goyal Shimla University*

Shimla, India. [1]nitishchauhan946@gmail.com, [2]mails4keshav@gmail.com, [3]ramesh@agu.edu.in

***Abstract*— Agriculture is the most essential quarter in Indian economy. India occupies the 2nd rank in farm outputs in the world. Early Disease Detection and preventions are significant for better yield and nature of harvests. With Reduction in the quality of the agrarian Product, Disease Plant can prompt the enormous economic losses to the Individual ranchers. In-country like India whose significant Population is engaged with Agriculture It is vital to track down the sickness at beginning phases. Quicker and exact expectation of plant infection could assist with lessening the misfortunes. Machine Learning (ML) could be an urgent point of view for procuring genuine world and operative options for crop yield issues. Identification of these infections expect individuals to specialists notwithstanding a bunch of hardware and it is costly as far as time and money, Therefore, the advancement of a computer-based framework helps the detection of the plants' diseases which is very useful for farmers. The main aim of this study is to focus on, administered Machine Learning calculations like SVM, ANN, RF, LR, CNN, NB, and KNN with image processing strategies and furthermore analysis of different machine learning results with one another and tracks down the best calculation for plant infection grouping.**

Keywords—Machine Learning, Predictive analysis, algorithms analysis, Accuracy analysis, Crop & Plant Diseases prediction.

I. INTRODUCTION

A Machine Learning Algorithm are the programs that can read hidden patterns in data, predict output, and improve the performance of their own experiences.

Computers are designed to process the information over specific work and continuously improves the performance and accuracy by executing the task repeatedly. Thus, the computer based intelligent systems acquires the ability to make decisions, predictions, and predictions based on available data. For example, a machine can detect diseases detection in crops and plants by using the images of the crops & plants the efficiency of the machine can be improved by using the variety of images datasets.

The commonly used methods in Machine Learning are: classification, regression and clustering. Based on the features and availability of the training data set, we apply the supervised, unsupervised and reinforcement algorithms for the same.

Types of Machine Learning Algorithms:

There are some differences in how we can define the types of machine learning algorithms but they can usually be categorized according to their purpose and the main categories are as follows:

- **Supervised Learning**
- **Unsupervised Learning**
- **Semi-Supervised Learning**
- **Reinforcement Learning**

1. Supervised Learning: Supervised learning is when you have input variables (x) and output variables (Y) and you use an algorithm to read map activity from input to output.

$$Y = f(X)$$

The user provides a machine learning algorithm with a known dataset that combines the inputs and outputs you want, and the algorithm should find a way to determine how it gets to that input and output.

For Example: If the algorithm is to distinguish between fruits, the data must be labeled or separated by different fruits in the collection. Data is divided into classes in supervised reading.

 DOI : 10.35291/2454-9150.2022.0088 © 2022, IJREAM All Rights Reserved.

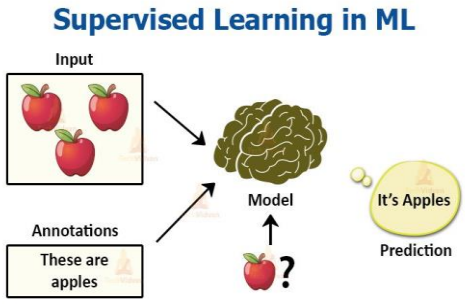

Figure 1: Supervised Learning in ML

List of Common algorithms in supervised Learning are as:

- Support Vector Machine
- Linear Regression
- Naïve Bayes.
- Decision trees.
- Neural Network.
- Nearest Neighbor.

2. Unsupervised Learning: In unsupervised learning, you just have input information (X) and no corresponding output values. The objective for unsupervised learning is to show the fundamental structure or dispersion in the information to look further into the information. For Example: if we have data for dogs and cats, the model will process and train with the data. Since it does not have the pre-sensitivity of the data, it will create clusters based on the similarity of the features.

For example: if we have data for dogs and cats, the model will process and train with the data. Since it does not have the pre-sensitivity of the data, it will create clusters based on the similarity of the features. Features like dogs will end up in the same group, and the same goes for the cat.

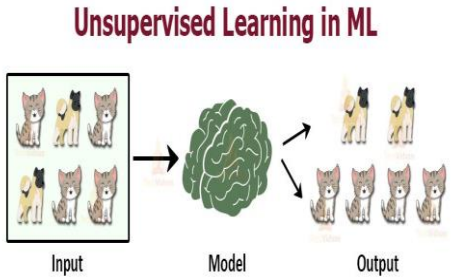

Figure 2: Unsupervised Learning in ML

List of Common algorithms in Unsupervised Learning are as:

- k-means clustering
- Association rules.

3. Semi-Supervised Learning: Problems where you have a lot of information (X) and just a portion of the information is named (Y) are called semi-supervised learning problems.

These problems lies between the both learning types i.e. supervised and unsupervised learning.

In this case, the model begins with training under unsupervised learning. This ensures that most unlabeled data is fragmented. For data left without labels, label production takes place and editing goes smoothly. This process is very useful in areas such as speech recognition and analysis, protein classification, text classification, etc. This is a kind of mixed or hybrid learning problem.

4. Reinforcement Learning: Reinforcement learning focuses on structured learning processes, in which the machine learning algorithm is provided with a set of actions, parameters, and end values. By defining the rules, the machine learning algorithm then attempts to evaluate different and possible options, monitor and evaluate each outcome to determine which one is perfect. Reinforcement learning teaches machine testing and error. He learns from the past and begins to adjust his approach to the situation in order to achieve the best possible result.

For example: Suppose you have a dog and you try to train your dog to sit up. You could give the dog some instructions to try to get him to learn. If the dog does well with this instruction, he will receive a biscuit as a reward. If not, it will get nothing. The dog learns from this after trying to find a biscuit while sitting.

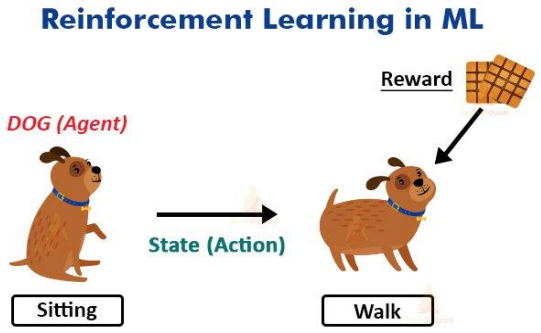

Figure 3: Reinforcement Learning in ML

Suppose you have a dog and you try to train your dog to sit up. You could give the dog some instructions to try to get him to learn. If the dog does well this instruction, he will receive a biscuit as a reward. If not, it will get nothing. The dog learns this after trying to find a biscuit while sitting. This is what we called a Reinforcement learning.

The Most Common Algorithms that are used for the Diseases Prediction:

In this section we covers the some most popular Machine Learning Algorithms that are used widely used for the crop or plant diseases detection are as follows below:

1. Support Vector Machine (SVM).
2. Artificial Neural Network (ANN).
3. Random Forest (RF).
4. Convolution Neural Network (CNN).
5. K-Nearest-Neighbour (KNN).

 DOI : 10.35291/2454-9150.2022.0088 © 2022, IJREAM All Rights Reserved.

A. *Support Vercot Machine (SVM)*

The "Support Vector Machine" (SVM) is a supervised learning algorithm that can be used for both editing and deceleration challenges. However, it is widely used in separation problems. In the SVM algorithm, we set each data object as a point in the n-dimensional space (where n is the number of attributes you have) by the value of each element which is the value of a particular combination. Then, we make a distinction by finding a hyper-plane that best divides the two categories.

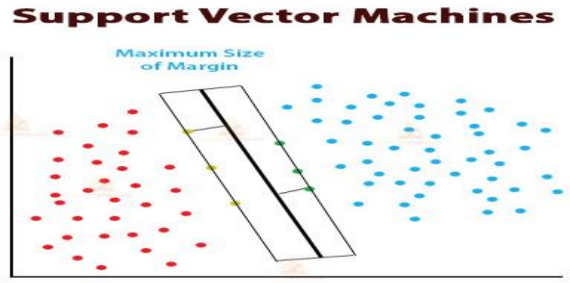

Figure 4: SVM Method in ML

B. *Artificial Neural Network (ANN) .*

Artificial neural networks (ANN) take input as a random image using a functional calculation model and convert it into corresponding output labels. It requires a little pre-training effort that can be trained to learn the features needed for classification purposes.

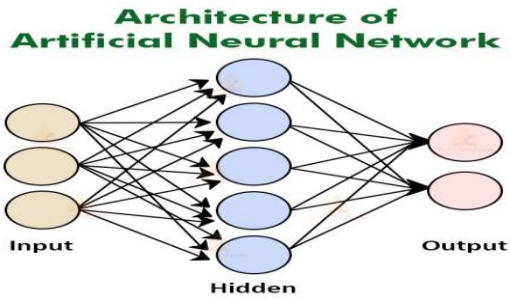

Figure 5: ANN Method in ML

C. *Random Forest (RF) :*

Random Forest is a popular machine learning algorithm that is part of a supervised Machine learning. It can be used for both Regression and Classification problems in ML. It is based on the concept of integration learning, which is the process of combining multiple class dividers to solve a complex problem and improve model performance.

As the name suggests, "The Random Forest is a subdivision that contains a number of decision trees for the various datasets set and takes measures to improve the prediction accuracy of that database." Instead of relying on a single decision tree, the random forest takes a prediction from each tree and is based on multiple predictable votes, and predicts the final outcome.

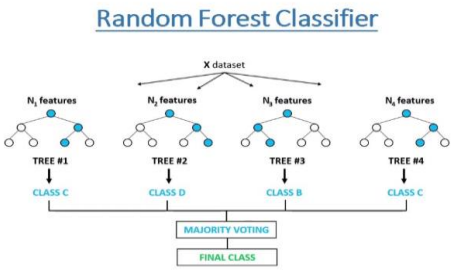

Figure 6: Random Forest in ML

D. *Convolution Neural Network (CNN) :*

The convolutional neural network is a feed-in neural network that is often used to analyze visual images by processing data with a grid-like topology. Also known as ConvNet. The convolutional neural network is used to detect and classify objects in the image. For example:

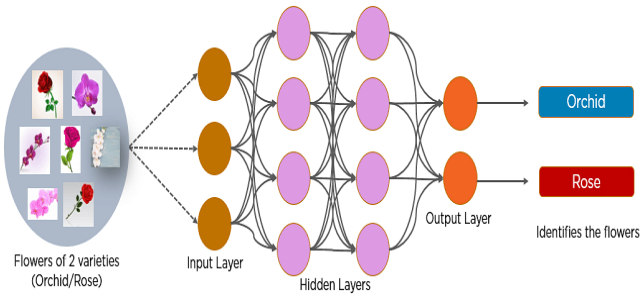

Figure 7: CNN Model in ML

E. *K-Nearest-Neighbour (KNN) :*

K-Nearest Neighbors is one of the simplest machine learning algorithms used for classification. It divides the data point based on the segmentation of its neighbors. It saves all available cases and classifies new cases based on the same features.

Suppose we have a picture of a creature like a cat and a dog, but we want to know if it is a cat or a dog. So in this analysis, we can use the KNN algorithm, as it works at the same rate. Our KNN model will get the same features of the new data set in pictures of cats and dogs and based on the very similar features it will place in the cat or dog category.

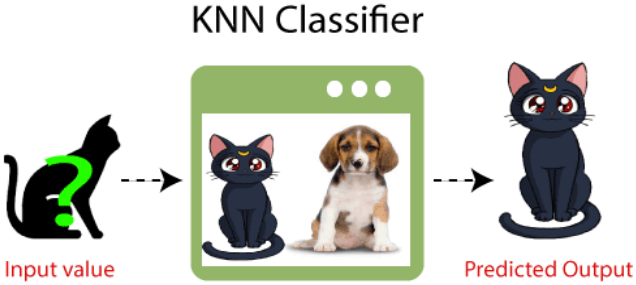

Figure 8: KNN Model in ML

 DOI : 10.35291/2454-9150.2022.0088 © 2022, IJREAM All Rights Reserved.

II. LITERATURE REVIEW

[1] (N. Kanaka Durga, 2019) in this research the researcher have taken the data list contains 200 images of tomato and maize leaves. In those 50pictures there is a healthy tomato leaf and corn photos and 110 photos of corn leaf pictures part training and part testing. Similarly, 40 leaf images are used for testing purposes. In Artificial Neural Networks (ANN), we have hidden layers and output layers. The Keras is a Python API belongs to the neural network. In this the tomatoes and corn both crops are tested using SVM and ANN machine learning algorithms and the result can be a tomato crop through SVM provides 60-70% and through ANN provides 80-85%. With corn, using SVM gives 70-75% again using ANN gives 55-65%.

And finally, we concluded that the one of the ML algorithm works well for which crop based diseases detection and on the accuracy rate.

[2] (Mohammed A. Hussein, 2019) the first part of the proposed plant disease diagnostic program is a training process. Then the picture previous processing techniques used in these images. After pre-processing, the useful image elements are extracted using the element background method which will be used as samples to train the vector machine algorithm (the proposed machine learning algorithm in this system). In this study, images were collected from three plants, wheat, tomatoes, and cucumbers, three different diseases per plant, and the health status of each plant. In this study, we divided the whole sample into 80% training and 20% test. From the total database (799 images) 80% was used to train the SVM separator and 20% was used for testing. After completing the training, he separates will use 20% of the total sample to assess system accuracy depending on the confusion matrix. It was diagnosed with an accurate diagnosis of 90.6.1% and accurate detection of a healthy case 77.4 with a median accuracy of the diagnostic system 88.1%.

[3] (Sk Mahmudul Hassan 1, 2021) In this research the Implementing appropriate techniques to identify healthy and Unhealthy leaves helps to control crop losses and increase productivity. This category covers exceptions to existing machine learning techniques for diagnosing plant diseases.

For training and testing purposes, we have used the standard PlantVillagedataset for open access, which includes 54,305 numbers of healthy and infected plant leaves. Detailed information on the site, the number of classes and pictures in each class, their common and scientific names. The site contains 38 different categories of 14 different plants pictures of healthy- and disease-affected leaves. All photos were taken under laboratory conditions. After dividing the database into 80–20 (80% of complete training data, 20% of complete test images), we have benefited the best accuracy rate of 99.56% in the EfficientNetB0 model.

[4] (Kamalalochana. S, 2019) Database from Plant Village was used to train the Random Forest section. Pre-processed photos are separated into different texts. The effect of separating images by different restraint methods is given below. The following parameters used for tuning Random forest section measurements (number of trees) Leaf size Random setting for default settings, the accuracy is: 73.7% Effects of precision.

[5] (Sumit Kumar, 2021) in this research the purposed system is based on image perception. We explored the three main Architecture of the Neural Network: Based on the Fast Region Convolution Neural Network (Faster R-CNN), Regional-based Fully CNN (R-CNN) and Single-shot Multibook Detector (SSD) .The Plant Village database contains 54303 healthy and unhealthy leaf images. 38 categories by species and diseases. . We analyzed more than 50,000 images of plant-labeled leaves from 38 classes and attempted to predict the disease class the proposed system in this research can effectively identify different types of diseases and have the ability to deal with complex situations. The verification result shows 94.6% accuracy indicating the feasibility of the Convolution Neural Network and introduces an AI-based AI approach based on this Complex Problem.

[6] (Bijaya Kumar Hatuwal,2021) In this research work it identifies and predicts the disease given in plant images using different machine learning models such as Vector Support Machine (SVM), Nearest Neighbors (KNN), Random Forest Classifier (RFC), Convolutional Neural Network and compares the results. Image elements such as brightness, correlation, entropy, and contrast times are extracted using an algorithm of Haralic texture elements provided by SVM, KNN and Random Forest Algorithms while CNN feeds directly into images as input. Among the models used CNN produced a high accuracy rate of 97.89% and RFC, SVM and KNN had 87.43%, 78.61% and 76.96% respectively in sixteen different categories of images used.

[7] (Neelakantan. P, 2021) The purpose of this study is to analyze and find the best-supervised machine learning classification algorithm among RF, SVM,DT, KNN, NB, and KNN to classify plant disease. The dataset contains data like, various resolutions, medium, last, and early infection status of disease leaves, complex background, and obscured, halfway images. RF algorithm achieved 89 percent accuracy compared with other algorithms and was found to be the best algorithm in order to classify plant diseases.

[8] (S. Chakraborty, 2021) this paper tried to identify two types of apple black rot and rust of apple cedar. He tried to see again

 DOI : 10.35291/2454-9150.2022.0088 © 2022, IJREAM All Rights Reserved.

Healthy apple leaves where image separation is used as a processing stage. Multiclass SVM is used to train and test our data. The infected image segment circuit is used, and Multiclass SVM detects the type of disease in the actual leaf image among 500 images with 96% accuracy. It also shows the percentage of total infected area of that diseased apple leaf image.

[9] (A. KP, 2021) this research helps to identify plant diseases early to prevent crop losses and disease spread. The CNN model is used to accurately predict plant diseases. The performance of previously trained CNN models such as VGG, ResNet, and DenseNet is recognized, and based on performance metrics, the DenseNet model is found to be more accurate. Model testing was performed using performance analysis metrics such as accuracy, precision, memory, and F1 score. DenseNet model achieved high accuracy of 98.27%.

[10]. this research has analyzed and classified plant and plant disease predictors over the past 10 years (2010-2020). Forty-six research projects were identified and reviewed and evaluated of accepted methods and strategies for prior analysis and data used. Examples of predictions of plant and plant diseases were evaluated using a variety of metrics and algorithms with their tools. As revealed in the study of in the current sub-field, it is confirmed that the average accuracy is achieved to 72% from the analysis of the research papers evaluated.

TOP 5 MACHINE LEARNING ALGORITHMS USED FOR DISEASES PREDICTION

TABLE I

S.No	Algorithm used	Frequently used	Implementation tool	Performance & accuracy
1.	ANN	Crop based diseases detection	Keras & python API	65%
2.	CNN deep learning	Image based diseases detection	EfficientNetB0	99%
3.	SVM	Plant diseases detection	Image processing methods using SVM	88%
4.	CNN	Image perception diseases detection	R-CNN, SSD(Single-shot Multibook Detector)	94%
5.	RF	Image sgmentation ,diseased and non-diseasesd leafs of plant	Image processing methods using RF	73%

Table 1.0 : Summery Table for ML algorithm with their disease prediction accuray.

From the above research table we examines the number of machine learning algorithms that are used for the plant diseases prediction such as SVM, CNN,ANN, RF, and NB etc. The things such as accuracy and memory, accuracy are considered to measure the performance of a trained model using test data. The above Various monitored classification and regression algorithms are used for analysis. SVM (Support Vector Machine) is a straightforward model that divides data into classes by drawing a large plane. At a high level space, SVM acquired, a very advanced aircraft that maximizes the margin space between classes and data points. Random Forest, or RF algo rithm produces decision-making trees with data samples, and receives predictions from each for the best solution for multiple votes.In Bayes's theory can be used to calculate the positives of class label data easily using the conditional probability product of each element. CNN's, are the majority promising way to automatically learn the decisive and discriminatory aspects. Deep learning (DL) consists of different layers representing the learning features from data set.

III. CONCLUSION

There are many advanced methods for diagnosing and classifying plant diseases using diseased plant leaves. However, there is no effective and efficient solution that can be used to diagnose diseases. This research focuses on crop or plant diseases prediction and calculation of its yield with the help of various machine learning techniques. Several ML teaching methods and tools are used to calculate the accuracy. The purpose of this study was to analyze and identify the most widely used machine learning algorithm between RF, SVM, ANN, DT, KNN, NB, and CNN to differentiate and identifying the plant diseases. And the state of the leaves of the original disease contains with healthy and unhealthy the complex background, the hidden images, the middle. The CNN i.e. Convolution Neural Network algorithm achieved approximate 95% percent accuracy as compared to the other algorithms and also found to be the best algorithm for classifying and prediction the plant & crop diseases detection.

IV. REFERENCES

[1] N. Kanaka Durga, G. Anuradha "Plant Disease Identification Using SVM and ANN Algorithms" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.

[2] M. Hussein and A. H. Abbas, "Plant Leaf Disease Detection Using Support Vector Machine," Al-Mustansiriyah Journal of Science, vol. 30, no. 1. Al-Mustansiriyah Journal of Science, pp. 105–110, Aug. 15, 2019 [Online]. Available: http://dx.doi.org/10.23851/mjs.v30i1.487

[3] Hassan, S.M.; Maji, A.K.; Jasi ´nski, M.; Leonowicz, Z.; Jasi ´nska, E. Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. Electronics 2021, 10, 1388. https://doi.org/10.3390/electronics10121388.

 DOI : 10.35291/2454-9150.2022.0088 © 2022, IJREAM All Rights Reserved.

[4] Kamalalochana. S, Nirmala Guptha "Optimizing Random Forest to Detect Disease in Apple Leaf" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958,Volume-8, Issue-5S, May 2019.

[5] Sumit Kumar, Veerendra Chaudharyand Ms. Supriya Khaitan Chandra "Plant Disease Detection Using CNN" Turkish Journal of Computer and Mathematics Education Vol.12 No.12 (2021), 2106-2112.

[6] Hatuwal, Bijaya & Shakya, Aman & Joshi, Basanta. (2021). Plant Leaf Disease Recognition Using Random Forest, KNN, SVM and CNN. Polibits. 62. 13-19. 10.17562/PB-62-2.

[7] Neelakantan . P, "Analyzing the best machine learning algorithm for plant disease classification," Materials Today: Proceedings. Elsevier BV, Aug. 2021 [Online]. Available: http://dx.doi.org/10.1016/j.matpr.2021.07.358.

[8] S. Chakraborty, S. Paul, and Md. Rahat-uz-Zaman, "Prediction of Apple Leaf Diseases Using Multiclass Support Vector Machine," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). IEEE, Jan. 05, 2021 [Online]. Available: http://dx.doi.org/10.1109/ICREST51555.2021.9331132 .

[9] A. KP and J. Anitha, "Plant disease classification using deep learning," 2021 3rd International Conference on Signal Processing and Communication (ICPSC). IEEE, May 13, 2021 [Online]. Available: http://dx.doi.org/10.1109/ICSPC51351.2021.9451696.

[10] G. Fenu and F. M. Malloci, "Forecasting Plant and Crop Disease: An Explorative Study on Current Algorithms," Big Data and Cognitive Computing, vol. 5, no. 1. MDPI AG, p. 2, Jan. 12, 2021. doi: 10.3390/bdcc5010002.

[11] https://techvidvan.com/tutorials/types-of-machine-learning.

[12] https://www.javatpoint.com/machine-learning-algorithms.

[13] https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network.

 DOI : 10.35291/2454-9150.2022.0088 © 2022, IJREAM All Rights Reserved.