International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 08, August-2018

A SURVEY OF MACHINE LEARNING ALGORITHMS FOR DISEASE CLASSIFICATION IN CROPS AND PLANTS

Mini Rani[1], Rupinder Kaur[2]

[1]*M.tech, CSE Dept. RIMT University, Mandi Gobindgarh, Punjab, India.*
[2]*Assitant Professor, CSE Dept. RIMT University, Mandi Gobindgarh, Punjab, India.*

Abstract- In the plants, fruits, and vegetables, the problem of infections can lead to the reduction in quantity and quality of the products. This infection can further affect the human health when they consume those infected products in their food intakes. Thus, it is important to detect the infections/diseases in the early stages so that the agriculture products can maintain the quality of the products. Previously, the farmers were using the approach of monitoring the crops continuously for any disease identification. This method was too time consuming and requires a lot of efforts from the farmer's side. So, in recent years, researchers focused on this area and tried to improve the work using the machine learning. A number of algorithms and techniques were developed for machine learning to detect the disease on crops, fruits, and vegetables. The techniques used image processing, segmentation, and classification as their backend to achieve the desired goal. This paper presents the detailed review of recent machine learning algorithms that were applied to agriculture products to detect the diseases and infections and achieves great success. This paper also presents the techniques based on detection of diseases, their methodology, and achieved efficiency by previous researches till now.

Index Terms—Machine Learning, Agricultural disease detection, Classification, Agricultural science, Algorithms.

I. INTRODUCTION

Agriculture plays a significant part in the economy of every country. In the countries like India, where more than 50% of the workforce is associated with agriculture but only have the 17.35 % of the GDP (Gross Domestic Product) that is too less as compared to manpower linked [1] [2]. There are many factors that affect the agricultural products and degrades their quality including weather conditions, low-quality seeds and unavailability of resources at some particular time. These problems convert into serious problems if the diagnosis is not made properly and in early stages. Previously, the farmers were performing the diagnosis manually and by checking the crops at some intervals. With this, some areas of crops remain unchecked that was the issue of damage to more crops in that area. The diagnosis process requires the complete checking of the soil, condition of crop and effect of weather on the soil. With the growing technology, it becomes easier to detect the disease of the crops in early stages using the image processing, segmentation, pattern recognition, and texture analysis. Among all these techniques, the image processing technique is commonly used for the detection and classification of disease among the crops.

In image processing systems, the images of the affected areas of crops are given as input. After that, the various image processing techniques are applied on that to get the affected area separate like cropping and segmentation. This process requires complete can deep analyzation of the images first and then selecting the appropriate algorithm accordingly. Here, the selection of algorithm plays an important role in the final results because the way of working of each algorithm is different on a variety of images. The input images are required to be of good quality so that images can be analyzed properly and correct classification/detection can be made.

The researchers and scientists developed some techniques to take good quality images with professional cameras. The images of the affected areas of the crop can wither be taken manually, or robots can be used to capture the images from different angles of the crop. Fig. 1 shows the normal process of detecting the diseases from the plants using the image processing approach. The results of such disease detection system are measured in terms of accuracy of the overall research in detecting the correct disease classification. It was found that the disease in the crops differs in color, size, shape, and texture of the outer surface. This can be analyzed by the image processing system efficiently with the use of some algorithms and image processing techniques.

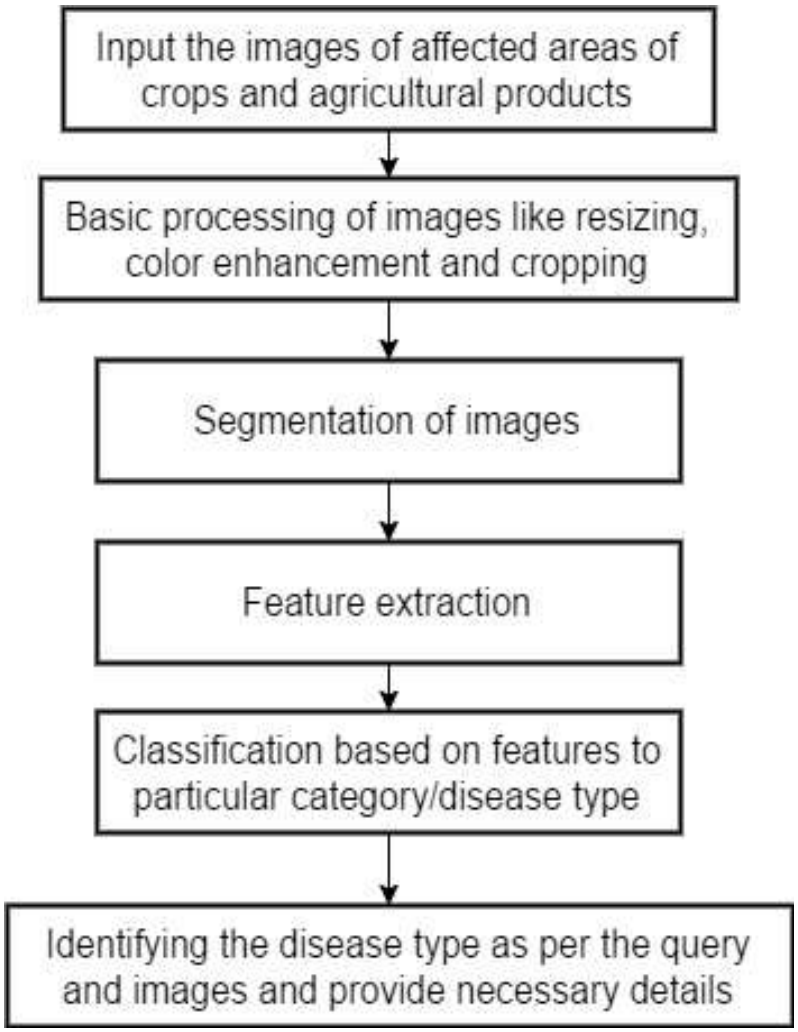

Figure 1. The process of Image processing to detect disease

Most researches have shown their researches using the online available images of affected areas of crops. The internet plays a major role as a large source of the database in this kind of researches [3].

II. Machine Learning Algorithms

Machine learning is becoming more popular due to its techniques and practical applications provided in several areas. Basically, machine learning is used for 4 purposes, detection or identification, Quantification, classification, and prediction [4]. In identification/detection, the machine algorithms are used to detect the object or particular required part in images or objects. For example, identifying the affected area on plants or finding water area from satellite images. In quantification, the machine learning algorithms are sued to measure the quantity of the objects that are detected. The classification works with classifying the images or objects into different categories based on the dataset properties and features. The last prediction is normally used in researches to predict the future outcomes based on the current dataset and facts collected.

However, the Quantification and prediction are more important than classification and identification, and less work is done in this in past researches [4]. The prediction algorithms of machine learning can be proved useful in predicting the early notifications for crop disease based on current environment settings and features of the crops. There are many machine learning algorithms that are used for disease detection in plants, fruits, and vegetables. These include Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Ensemble techniques, Clustering Algorithms, K-nearest neighbor (kNN), etc.

The rest of the paper is organized as Section II, which describes the information related to disease classification researches that were held in the past. Section III describes the results achieved using various techniques in the past by different researchers in the field of disease detection in plants and crops. At last, section IV explains the conclusion and future scope of this area after analyzing the previous researches.

III. LITERATURE REVIEW

Pinto L. S. et al. [1] proposed an approach to detect and classify the disease in the sunflower crop using image processing. The research was carried out using the leaf images of the crop that were taken using a high-resolution digital camera. The prepared dataset is first pre-processed, and then k-means clustering was applied to detect the affected part of the leaf from the disease. After that various machine learning algorithms were used to classify the images based on their texture and color using the MATLAB programming language. The research also presented the results comparison of these algorithms based on the accuracy of these machine learning algorithms namely Multi-class Support vector machine (MC-SVM), K-nearest Neighbour (kNN), Naive Bayes (NB), and Multinominal Logistic regression (MLR). The results of the study showed that MLR is the best classifier algorithm having an average accuracy of 92.57%. Although other algorithms as well presented very close results to this.

Ferentinos K. P. [5] presented a model using the convolutional neural network to detect the disease from plant images. The images were considered of both diseased and healthy plants for preparing the model. The dataset containing 87,848 images were related to 25 different types of plants and contained 58 unique set of the combinations of plants. The proposed model achieved a success rate of 99.53% in identifying the combinations. The results made the proposed model more useful in providing early warning notification, and the approach can further be extended for integrating the disease identification system for plants. As an improvement to this work, the dataset can be extended to more images that are related to different environmental conditions, and areas. The research work can be enhanced by modifying and train the images using some other model.

Moghadam et al. [6] proposed a technique using machine learning and hyperspectral imaging to detect the Tomato Spotted Wilt Virus (TSWV) in the capsicum plants. First, the features are extracted based on probabilistic models, vegetarian indices, and full spectrum. Based on these features, state vector machine (SVM) classifier was trained for differentiating between the inoculated and healthy plants. The results show about 90% accuracy using the proposed technique. This work can be extended further based on the dynamic topic models to find the dynamic nature of plants over the time. New classification techniques can be applied to enhance the accuracy of the work. The 90% of accuracy may or may not detect the test image accurately.

Fuentes A. et al. [7] presented the deep learning based mechanism, which helps in detecting the pests as well as diseases in the tomato plant. This was done with the help of captured images with different resolutions. For the purpose of the deep learning Meta architectures, they have considered three major categories of the detectors. These are Faster R-CNN (Convolutional Neural Network), SSD (Single Shot Multi-box Detector) and R-FCN (Region-based Fully Convolutional Network). For the research, they have collected various images as the dataset under several conditions. These conditions are dependent on time, season and the place. After collecting the dataset, they have manually annotated areas of some images containing the disease. The main aim of the data annotation process is to label the class as well as the location of the infected regions in the image. After this, they have proceeded with data augmentation. This was done as the number of images in the dataset are not enough. For this, they have resized the image, cropped the image and rotated the image. After this, they have used the R-CNN method for the recognition of objects. They have applied SSD and R-FCN techniques. Their resultant AP (average precision) is more than 80%. The future work can includes the methods to improve the present results and applying this technique to other crops as well.

Mohanty S. P. et al. [8] have detected the plant diseases with the help of deep learning from images. They have utilized the total dataset of 54306 images, which has both the healthy and the diseased plant leaves. They have then resized the images to (256*256) pixels. Both the model optimization as well as the predictions were done on these downscaled images. They have performed their work on three different versions of the dataset. These versions are a color plant, Grayscale plant, and the segmented plant. Segmentation of the plant was done by removing the background, which contains the extra information. They have then trained the deep convolutional neural network. Their main goal was to classify the crop species and to identify the diseases present in the images. As a result, they were getting 99.35% accuracy. Their model correctly classifies the diseases and crop. They have also checked the performances of AlexNet and GoogLeNet. In future, the research needs to be more focused on identifying the disease not from the leaves but from the other part of plants too.

Mwebaze, E., & Owomugisha, G. [10] have presented the system for solving the problem of diagnosis of crop disease, which is based on smartphones. For this, they have collected the dataset of around 7386 images of cassava plant leaves. These images are divided into five categories. These include healthy class and other four of diseased class. The diseased class includes the CMD, CBSD, CBB, and CGM. They have extracted the features, which represents the color, shape,

HOG (Histograms of Oriented Gradient), ORB (Oriented FAST and Rotated BRIEF), Hue histograms, and SIFT (Scale Invariant Feature Transforms). They have then used the Scikitlearn machine learning toolbox, which helps in training of the suitable classifiers. The three classifiers, which were trained and used, are LinearSVC, KNN, and Extra trees. They have found that the ORB seems to be fast as well as reliable which can replace the SURF and the SIFT. For ORB, the accuracy is 99.98% in case of LinearSVM, it is 99.88 in case of Extra trees, and is 100% in KNN. The overall accuracy of the research work was around 99%.

Sladojevic, S. et al. [11] proposed the new approach of automatically classifying and detecting the plant diseases with the help of leaf images. The work was implemented using deep learning methods. They have collected the dataset of the images by searching on the internet with the disease name and plant name. They have then enriched their dataset by applying the data augmentation. They have total 30880 images as the dataset. For the feature extraction, they have cropped the images manually and extracted the required region. After this, they have trained the deep convolutional neural network which helps in classification of the images. The CaffeNet architecture was considered for their work. They have included Rectified Linear Units (ReLU). The CNN with the ReLU helps in training in the better way. The overall accuracy of their work is around 96.3% after fine-tuning. The accuracy result was 95.8% without the use of fine-tuning. In future, they are aiming to develop the system, which consists of server-side components and will help in recognizing the diseases in other plants, fruits, and vegetables.

Sarangdhar A. A. and Pawar V. R. [12] have presented the system, which helps in detecting and controlling the diseases especially on the cotton leaf. They have proposed the SVM (Support Vector Machine) system, which is used to identify and classify the cotton leaf diseases. After detecting the disease, they have provided the name of the disease to the farmers along with the prevention remedies. They have provided an Android app, which helps in displaying the soil parameters like moisture, humidity and its temperature. This system of disease detection is interfaced with the Raspberry Pi. The overall accuracy of their proposed system was around 83.26%.

Wu N. et al. [13] have proposed the approach, which is used in classifying the pixels in the images of the diseased crops. Their proposed method is related to the machine-learning algorithm, which is known as LDA (Linear Discriminant Analysis) and color transformation. They have extracted, and grouped various colored features of the pixel and the put all into the LDA model. In their work, they have used four kinds of cucumber images of diseases. They have calculated the accuracy percentage, precision value, and recall value, False positive Rate, and F-score for their values. The best results were obtained by LDASM-RGB with the accuracy of 91.27%, the Precision value of 88.11%, Recall value as 86.48%, False Positive Rate as 6.19%, and F-Score as 87.29%.

AlSuwaidi A. et al. [14] have proposed the system, which is used to analyze and classify the HSI (Hyperspectral Imaging) datasets. Their main focus is to differentiate the diseased plants by detecting and monitoring the condition of the crops. They have integrated the adaptive feature selection, ND (Novelty Detection), and assemble learning in their work. The six features selection mechanisms were used in their work. They are Gini index, ReliefF, information gain, chi-square, CFS, and FCBF.

Joshi A. A. and Jadhav B.D. [15] have proposed the new technique, which is used to diagnose and to classify the diseases related to rice. The four disease related to rice includes rice blast, rice sheath rot, rice bacterial blight, and brown rice spot. They have identified and classified these diseases. Various features were extracted based on the shape of the rice. These extracted features were then combined and classified with the help of MDC (Minimum Distance Classifier), and KNN (k-Nearest Neighbor classifier) approaches. The overall accuracy of their work is around 89.23% in case of MDC and is 87.02% in case of KNN.

Sandika B. et al. [16] have proposed the system, which is utilized in classifying the diseases affecting the grapes. The three diseases, which they have classified in their work, include Downy Mildew, Anthracnose, and Powdery Mildew. The detection is done with the help of image processing and machine learning algorithms. They have used the Random forest approach, and the GLCM features method in their work. The performance of their work is compared with other machine learning approaches like PNN, BPNN, and SVM. The overall accuracy achieved was around 86%.

IV. RESULTS AND DISCUSSIONS

Table 1 shows the methodologies used till now in the field of disease detection from plants using various techniques.

Table 1: Techniques used for disease detection in plants and their Results

Year	Researchers	Techniques/Method/Algorithm	Results/Solution/Conclusion
2016	Pinto L. S. et al. [1]	MC-SVM, kNN, NB, and MLR	Results showed that Multinominal Logistic regression (MLR) is best classifier algorithm with 92.57% accuracy.
2018	Ferentinos K. P. [5]	Convolutional neural network	Research achieved 99.53 % success rate in differentiating and identifying the combinations of datasets.
2017	Moghadam et al. [6]	SVM, Machine learning, and hyperspectral imaging	Achieved 90% accuracy in differentiating between healthy and inoculated plants.
2017	Alvaro Fuentes et al., [7]	R-CNN, SSD, and R-FCN	Their results show that resultant AP (Average Precision) is more than 80%.
2016	Mohanty S. P. et al., [8]	Deep convolutional neural network, AlexNet, and GoogLeNet.	As a result, the research achieved 99.35% accuracy.
2016	Mwebaze, E., & Owomugisha, G [10]	Color and ORB (Oriented FAST and Rotated BRIEF). LinearSVC, KNN, and Extra trees.	The overall accuracy achieved was around 99%.
2016	Sladojevic, S. et al. [11]	CNN with the ReLU (Rectified Linear Units), and deep convolutional neural network.	The overall accuracy of proposed work was around 96.3% after the use of fine-tuning. The accuracy result was 95.8% without the use of fine-tuning.
2017	Adhao Asmita Sarangdhar, and Prof. Dr. V. R. Pawar [12]	SVM (Support Vector Machine) Android App Raspberry Pi	The overall accuracy of their proposed system was around 83.26%.
2017	N. Wu, M. Li, L. Chen, Y. Yuan and S. Song [13]	LDA (Linear Discriminant Analysis) and color transformation. LDASM-RGB (LDA based segmentation model)	The system has an accuracy of 91.27%, the Precision value as 88.11%, Recall value as 86.48%, False Positive Rate as 6.19%, and F-Score as 87.29%.
2018	A. AlSuwaidi, B. Grieve and H. Yin [14]	Features selection mechanisms like Gini index, ReliefF, information gain, chi-square, CFS, and FCBF.	The results are improved with the new approach.
2016	A. A. Joshi and B. D. Jadhav [15	MDC (Minimum Distance Classifier) and KNN (k-Nearest Neighbor classifier)	The overall accuracy of research work was around 89.23% in case of MDC and 87.02% in case of KNN.
2016	B. Sandika, S. Avil, S. Sanat and P. Srinivasu [16]	Random forest approach, and the GLCM features method	The overall accuracy achieved was around 86%.

V. CONCLUSION AND FUTURE SCOPE

Machine learning algorithms are becoming more popular in detecting and classifying the disease in plants and crops due to their significant results and outcomes. This paper reviewed the previous researches based on disease detection in plants and crops using machine learning algorithms based on the techniques used and results achieved till now. This will help in conducting any further research in this area for the advancements. Also, more machine learning algorithms can be tested in this field to get more improved and accurate results. From the review, it has been observed that less work is done using the hybrid of two or more algorithms, so future work can be carried out using a hybrid approach of machine learning so that more accurate results can be obtained and disease can be detected at the early stages. Other than a hybrid, more

algorithms can also be used for detecting the disease from the images dataset using image processing techniques and algorithms of machine learning.

REFERENCES

1. L. S. Pinto, A. Ray, M. U. Reddy, P. Perumal, and P. Aishwarya, Crop disease classification using texture analysis, 2016 *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pp. 825-828, 2016.
2. https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=IN last Accessed 18 April, 2018
3. M. Tripathi, and D. Maktedar, Recent machine learning based approaches for disease detection and classification of agricultural products, *in Conf. 2016 International Conference on Computing Communication Control and automation (ICCUBEA).*
4. X. Yang and T. Guo, Machine learning in plant disease research, *European Journal of BioMedical Research*, vol. 3, no. 1, pp. 6–9, Mar. 2017.
5. K. P. Ferentinos, Deep learning models for plant disease detection and diagnosis, *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
6. P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka and E. Hernandez, Plant Disease Detection Using Hyperspectral Imaging, *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Sydney, NSW, pp. 1-8, 2017.
7. A. Fuentes, S. Yoon, S. Kim, and D. Park, A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition, *Sensors*, vol. 17, no. 9, p. 2022, 2017.
8. S. P. Mohanty, D. P. Hughes, and M. Salathé, Using Deep Learning for Image-Based Plant Disease Detection, *Frontiers in Plant Science*, vol. 7, 2016.
9. H. ur Rahman, N. J. Ch, S. U. Manzoor, F. Najeeb, M. Y. Siddique, and R. A. Khan, A comparative analysis of machine learning approaches for plant disease identification, *Advancements in Life Sciences*, vol. 4, no. 4, pp. 120–126, 2017.
10. E. Mwebaze and G. Owomugisha, Machine Learning for Plant Disease Incidence and Severity Measurements from Leaf Images, 2016 *15th IEEE International Conference on Machine Learning and Applications (ICMLA), 2016*.
11. S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–11, 2016.
12. A. A. Sarangdhar and V. R. Pawar, Machine learning regression technique for cotton leaf disease detection and controlling using IoT, 2017 International Conference of Electronics, Communication, and Aerospace Technology (ICECA), 2017.
13. N. Wu, M. Li, L. Chen, Y. Yuan and S. Song, A LDA-based segmentation model for classifying pixels in crop diseased images, 2017 36th Chinese Control Conference (CCC), 2017.
14. A. AlSuwaidi, B. Grieve, and H. Yin, Feature-Ensemble-Based Novelty Detection for Analyzing Plant Hyperspectral Datasets, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 4, pp. 1041-1055, 2018.
15. A. A. Joshi and B. D. Jadhav, Monitoring and controlling rice diseases using Image processing techniques, 2016 International Conference on Computing, Analytics and Security Trends (CAST), 2016.
16. B. Sandika, S. Avil, S. Sanat and P. Srinivasu, Random forest-based classification of diseases in grapes from images captured in uncontrolled environments, 2016 IEEE 13th International Conference on Signal Processing (ICSP), 2016.