

## A Study Based on Plant Disease Prediction System Using Machine Learning

Arvind Kumar Shukla<sup>1\*</sup>, Rajdeep Singh<sup>2</sup>, C.K Dixit<sup>3</sup>

<sup>1,2</sup>School of Computer Science and Applications, IFTM University, Moradabad, U.P. India

<sup>3</sup>Dr. Shakuntala Misra National Rehabilitation University, Lucknow, U.P. (U.P.), India.

E-Mail: [\\*arvindshukla.india@gmail.com](mailto:*arvindshukla.india@gmail.com)

**Abstract**—The Agricultural field plays an important role for the growth any country. Most of Indian populations are depend for their livelihood on the agriculture/crops. Presently, the Indian agriculture is facing a number of hurdles because of the change of climate, water pollution/shortages, lack of fertilizers, old methods/technologies, different plant's diseases and many more. These factors are not good for better food production to fulfill the public demand on time. The crops may be ruined by the infections in the plants, it may harm to our food security. To detect the diseases within the plants is not easy. DFS (Disease forecasting System) for Potato crops using ML (machine learning) is the best method to predict the plant's diseases for necessary solution to prevent it, timely.

**Keywords**—Machine Learning, Disease forecasting System, Image processing, SVM, CNN, K-means.

### 1. Introduction

Quality agriculture relies on gathering comprehensive information from farmland, which includes not only the environmental information but also the plant's information. For example, the environmental condition, such as temperature, soil moisture, humidity, soil composition, solar radiation, wind speed and rainfall, are considered to reveal the weather change and soil pollution, and can help to improve management of fertilizer usage and other inputs. The plant information, such as plant growth, plant disease and insect pest, are useful to predict the production and make decision about pesticide or organic applications.

This Disease Prediction System will give a better prediction of disease on Potato and it will recommend suitable pesticide/ organic materials applications using Machine Learning. Python Language is most widely language for Machine Learning, AI, Data Science, Deep Learning and Image Processing because of its simplicity and power full library. Open source python language is used to implement this system with some libraries like OpenCV, pytesseract, NumPy, SciPy, TensorFlow etc.

In this Disease Prediction System, Machine Learning and Clustering Algorithms can be applicable for plant growth monitoring. In this solution aggregation of data helps to convert raw data coming from plant into exact accurate data which will be understood by farmers.

Finally exact accurate information processed by Disease Prediction system is conveyed to farmers via messaging / App service on their mobiles, which will help to predict the application of pesticide to the plants

## 2. Literature Survey on Disease Prediction Techniques

Andreas Kamilaris et. al.[2018], author discussed and perform a survey of 40 researches efforts that employ deep learning techniques, applied to various agricultural and food production challenges. To study the agricultural problems stated under each work, the specific models and frameworks employed the sources, nature and used dataset and overall performance achieved with help of used methods, comparison of deep learning with other techniques also done. It is stated that deep learning provides high accuracy by using image processing techniques [1].

Konstantinos P. Ferentinos et. al.[2018], author discussed about convolutional neural network models were developed to perform plant disease detection and diagnosis through deep learning methodologies using plants leaves images. Training of proposed models was done with the use of an open database of 87,848 images, containing 25 different plants in a set of 58 distinct classes of [plant, disease] combinations. Some models gets success 99.53% in detecting the corresponding combination [plant, disease]. High rate of success model is very usefull to early detection tool and also have possibilities of further expanded by researchers [2].

Shima et.al. (2018) proposed that plant disease cause decrease in food production. For detection purpose machine learning techniques are used by many researchers like RF processes, SVM processes, K-means processes, CNN processes. The random forest algorithm does the classification. The aim of author is to detect the disease with random forest classifier. We have to convert RGB type images to an HSV type's image[3].

Prem et.al. (2018) proposed that some symptoms are visible from the eyes are wilting, spot, powdery mildew, galls, and dryness. Different attributes are taken in dataset, different techniques are used and different plots like box plot, bar plot are performed. With the help of statistical tests the prediction is done on inbuilt dataset. Many techniques are compared and the accuracy is different from each sample dataset [4].

Budiarianto et.al. (2018) proposes Machine Learning techniques for recognition of disease in corn plant which is a main source of carbohydrate. CNN technique is used to improve plant disease. Researcher used different algorithms and use support vector machines (SVM), Decision Tree (DT), Random Forest (RF), and Naive Bayes (NB) to compare the results. By normal seeing of plant we can understand the problem like color difference. Different parameters are used for dataset attribute [5].

Sherly et.al. (2019) proposed there are different type's bacteria or fungus is responsible for many different plant diseases. It can be predict using algorithm of Machine Learning. Many researchers try many algorithms and get differ results. The classification of diseases is hard to done by algorithms. By CNN technique we can identify the mulberry plant disease [6].

Balwant J Gorad et al.(2019) gives a better disease prediction system for potato plant. K-means clustering used to split the data that is provided by the farmers. Farmers collect images from their phone, tablet, camera and other sources that is forwarded to the system and then system create

dataset from that and periodically it is done and hence plants diseases predicted by the system [7].

MonalisaSaha et al. (2020) takes the tomato and potato plants leaves to predict plant diseases. They collect both plant leaves images from internet sites and some images they collect with their digital camera from farming places. In proposed system after clustering if we give it (cluster) in multiple SVM classes then it gives better results and the performance analysis is 99% and the individual algorithm efficiency like k-means gives 88.6% and SVM gives 91%. Hence the proposed system is better than k-means and SVM [8].

SrideviSakhamuri et al. (2020) describe that there are three types of plant leaves diseases so they collect the leaves and maintain dataset according disease type. They collect different plants leaves like jasmine, grape, apple, beans, rose etc. and used different methods to detect leaves disease and get different accuracy like with k-means algorithm the accuracy was 88.8%, through SVM 95% accuracy was achieved and through ANN it was 70% to 95% for different diseases [9].

Krishnaswamy Rangarajan et al. (2020) proposed an automated disease diagnostic system for ten diseases of four crops (eggplant, hyacinth beans, lime, lady finger). Author used six pre-trained deep learning models for training and validation of created dataset [10].

Pranesh Kulkarni et al. (2021) had taken public dataset for their research in which healthy and unhealthy images of apple, corn, grapes, potato and tomato plants were included. For the classification they used Random Forest Classifier. They got 93% accuracy through the system they developed [11].

Sahil Thakur et al. (2021) authors develop the model with CNN to identify the plant diseases with image processing. They used plants like potato, grape, corn, apple etc [12].

GianniFenu et al. (2021) review the researches from past 10 years in which different plants and crops were used like cherry, coffee, barley, grape etc. and used methods were SVM, SVR, KNN, ANN and many. Author observed that researchers need high quality labeled data for their research work [13].

Kowshiket al. (2021) used Convolutional Neural Network and Deep Neural Network to detect plant diseases. Author detect similar diseases from different plants like banana, beans, jackfruit, lemon, mango etc. The proposed method with CNN and DNN is feasible for early plant disease detection [14].

Jayashri et al. (2021) reviews the existing image processing techniques for disease prediction of pomegranate. They used SVM, ANN, KNN and PNN classifier to detect bacterial, fungal and viral diseases in fruit. K-means clustering for image segmentation, Fuzzy c means gives highest accuracy. According to them very few diseases were covered in the existing system [15].

Punithast al. (2021) reviewed many research papers on detection of plants disease using image processing techniques. For image processing they follow the procedure image acquisition, image pre-processing, image segmentation, feature extraction and disease classification. After comparison of many different models (which used SVM, ANN, KNN and other approaches on different plants) they found that SVM is most accurate method followed by ANN [16].

### 3. The challenges in disease forecasting system for PotatoCrops

Potato is a popular vegetable and widely used for the various food productions in all around the world. Many species of potato are cultivated with different size and colors/shapes by the Indian farmers for better profit and result. The main factors/issues are necessary to point out and observe for good production of potato crops which are-

- To monitor the overall growth of the Potato plant, time to time.
- To check and forecast the plant's infection through its leaf's conditions.
- To observe and suggest suitable/required resolution for forecasted plant's disease in reliable way.
- To fulfill the public need of good and secure /healthy food items.
- To maintain the overall cost and production of potatoes in suitable manners for the people.

### 4. The Research Objectives

With the help of machine learning (ML) and image processing techniques, we can easily solve a number of issues regarding to protect potato crops from the various infections. This research work will be helpful to predict the potato plant's diseases and can suggest the required pesticides/fertilizers for the prevention of infection in potato plants. The objectives of this study/work are-

- To provide a logical forecast for Potato crops through the changing of leaf conditions.
- To recommend the proper pesticides or fertilizers for the predicted diseases of potato crops.
- To design a disease forecasting system, this predicts the growth of plants using ML techniques.
- To publish a new approach/work for agricultural research and various industries.

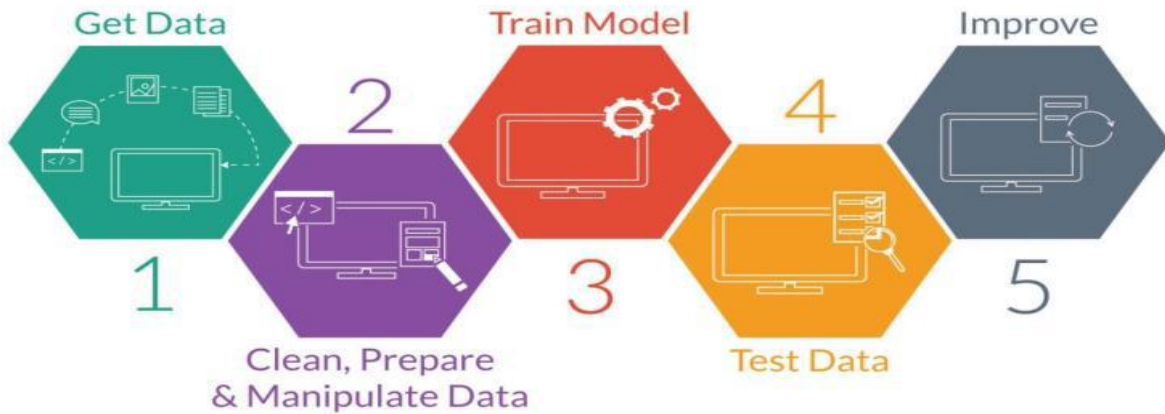
### 5. The Proposed Model

The architecture of DPS is human friendly system which contains separate back-end server for plants image processing and wireless devices for farmers such as mobile, tablet, PDA, laptop etc.



**Fig. 1 “Dataset of disease affected and healthy potato plant images which are used to create trained model”**

The camera of farmer's mobile / PDA / Laptop of DPS system is responsible for capturing image of plant leaf / actual Potato to test with earlier created model. Once image captured will be send to back-end for processing with trained model.



**Fig.2“Working Process of Machine Learning Classifiers”**

K-means Clustering algorithm is used to insert new image into appropriate cluster of disease categories. Using a desktop GUI form / web browser, a user can request and view historical data and along with this the last gathered “almost real-time” data from the back-end server. A Historical data as well as the last gathered “almost real-time” data, plant disease and insect pests of plant is sent to the farmers on their mobile using messaging /App service which can help to improve management of fertilizer usage & other inputs such as pesticide application.

## 6. Research Methodology

The Research will be carried out using following methodologies for development of Intelligent Monitoring Solution for Indian Agriculture.

- Literature study/review is to be carried out on Disease Prediction System for various plants along with strength and drawback in Agricultural Applications.
- Issues and challenges related to DPS for Agricultural Application Systems is to be studied from available literature & Books.
- Periodically Visits are arranged to Potato farm with farmers and agricultural professor / researcher which will help to monitor the growth of plants
- The probabilistic models for DPS are developed to resolve the issues such as prediction of disease on Potato leaf and on actual Potato for Indian Potato crop.
- Mobile Network and traditional network will be interconnected so that IP addressing, Client-Server architecture will be utilized.
- The developed schemes are being tested through application designed / simulation for it.
- Plant growth monitoring system is to be designed which will send information about management inputs such as pesticide application by messaging /App service to farmers.

- Application is to be made available to public on various platforms so that people will get it easily (ex. Google Play Store)
- Performance comparison with standard existing models.
- At every stage, results obtained will be published in suitable conference and journal.

## 6.1 Proposed Methodology

### A. Dataset

In this paper character recognition processes an appropriate and perfect dataset is required. First step of the proposed system is to train the system. After data is given to the system, evaluate the performance. Machine learning algorithm dataset is required. For this training data we take images and specially scanned data. A total of 1000 images are collected from different writing styles for train system and 2000 images are collected for test the system operation.

### B. Augmentation

In this paper, after the collection of dataset, convolutional neural network (CNN) is used for feature extraction which is one of the deep machine learning technique. For feature extraction on the above data set done using a supervised learning technique convolutional neural network (CNN). For this we take fundamental considerable information. Large amount of data can give large and accurate amount of feature attribute in CNN. All the data set are divided into their different data category. Take a data image and rotate the image as 90, 180, 270 respectively in mirroring each rotated images expanding the dataset using CNN. The process of CNN is different in different images. It is divided into 3 different convolutional layer. Each layer have in between a max-building blocks. First take an image as an input. Give the input to the initial convolutional block and modifies the entered image with 36 kernels of size 3.5x3.5. And give the result of primary convolution building block to the max block as an result. After first max building block gives the output to the second convolution building block as an input data. In the second convolutional block image will be filter with 64 kernels of size 4x4. Give this output to the second max pooling layer as an input data. After that max building block gives the result to final convolutional block. It filters images / data as 128 kernels of size 1x1. And give the output as fully connected 512 neurons layers. The result given to soft max function. The soft max function gives a prospected circulation of the four result categories. The last layer is connected to MLP. All convolution layer output have the activated ReLu function. And it fully connected to the layers. The system is trained using Adam. The batch size is size of 100 for 1000 epochs. In that way, we collect the features of the image dataset using the CNN algorithm.

### C. Classification Principle

In the data analysis and database management technique clustering is one of the data structure management technique. Lots of data can be sub divided into subgroup. Same type of data can be placed in same group. Using this method we can define task of identification. Basically find homogeneous subpart inside the data point. Euclidean-based distance or

correlation-based distance is used to identify these methods. It is an application-specific. Based on the features subgrouping clustering analysis is used.

Clustering is an unsupervised machine learning method. Clustering can be done in different ways. Partition the dataset features which are taken from Convolutional neural network (CNN). Each partition of dataset is non-overlapping clusters where every point of features belongs only to one group. Decide the total number of clusters. First centroid the random data point and iterating data. If centroids are not change then iterating repeatedly. Data points assign in same cluster. After that calculate the sum of the squared distance between data points and all data centroids.

$$J = \sum_{i=1}^m \sum_{k=1}^K w_{ik} \|x^i - \mu_k\|^2$$

**Fig.3 “ SVM Calculation”**

After that call SVM algorithm to evaluate k number of clusters. Sort number is denoted as a T. create a condition where every value can be evaluate as a newly generated solution. Then it will give kSVM solution.

`ksvm-model = {(c1,Lsvm1),(c2,Lsvm2).....(ck,Lsvmk)};`

Where, k= local model= no of cluster;

y=it is presented as parameter which is hyper of kernel function of RBF;

c= the error rate of SVM.

And in the very last return the global best solution. Repeat till all such cluster is pruned. And it gives final classification. Then we can classify and identify the characters and numeric values very efficiently. And calculate the accuracy of the characters recognition.

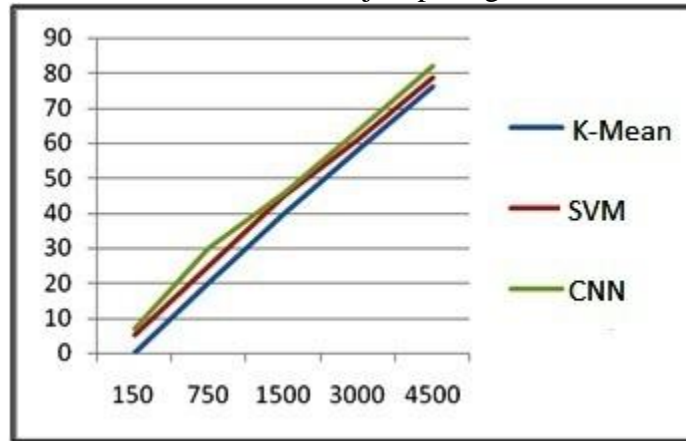
In the fields of Machine Learning, perhaps K-Means is the most known and studied method for clustering analysis.

K-means is a method of clustering which helps to feed up new scanned data or images into required form of blocks for corresponding scanned images. Using a desktop GUI form / web portal, a client can demand and visualize the historical images with the past gathered data from the server.

## 7. The Results and Discussions

The performance, related to Disease Prediction System is calculated, here. The overall performance is calculated as how much time is taken for the recognition process. In this way performance is calculated. For this experiment Anaconda navigator software system is used. Instanced of python programming using jupyter notebook is more efficient. After the proposed system total loss will be 0.02313. Using multilevel k-means and SVM algorithm efficiency of performance will be calculated. After clustering if the data will go through the SVM multi-level classification will be better result. 3500.00 potato data images are used for train the proposed system. Test accuracy with best parameter set is 0.9790.





**Fig.4 “Experimental Work”**

S No.	Classifier	Average Rate (%)
1	K-Mean	77.09
2	SVM	79.95
3	CNN	83.80

In this fig 4 we derived that individual algorithm like k-means process give efficiency 77.09% and SVM give efficiency 79.95%. But in the proposed methodology it gives better result. The performance analysis is 83.80%. The accuracy is better than individual algorithm performance. Table 1 shows the accuracy rate of different classifiers for character recognition.

### 8. Research Outcomes

The CNN approach is proving a valuable performance among three classifiers used here. K-mean algorithm didn't work well in universal cluster and it does not work well with cluster of different data size and different data mass. So that after clustering if we give the clusters in the multiple SVM class then it is give better classification .In this method it is found that big amount of dataset can be easily trained and tested to recognize the different data images. This proposed system gives a better result than other classifiers. Now in daily life, this kind of approached is very useful. Future work can be developing the algorithm better seg-mented techniques. So there is a scope of improvement in the techniques.

Following are the research objectives to be carried out in upcoming period for Disease Prediction System. Following are the research objectives to be carried out in upcoming period for topic of development of Intelligent Monitoring Solution for Indian Agriculture

- Design of Disease Prediction System which will predict disease on Potato crop based on symptoms of leaf and Potato in reliable way
- The probabilistic models is to be developed which will recommend appropriate solution in terms of pesticides / organic material for earlier predicted disease
- Designing of human friendly system (Computer / Mobile) which monitors plants growth by using machine learning techniques.
- Publication/patent of the work in suitable journals and agencies and Organization of conferences.



## References

- [1]. Andreas Kamilaris, Francesc X. Prenafeta-Boldu “**Deep learning in agriculture: A survey**”, Computers and Electronics in Agriculture 147, 70–90, 2018.
- [2]. Konstantinos P. Ferentinos, “**Deep learning models for plant disease detection and diagnosis**”, Computers and Electronics in Agriculture 145, 311–318, 2018.
- [3]. Shima Ramesh Maniyath, PV Vinod, M Niveditha, R Pooja, N Shashank, Ramachandra Hebbar., “**Plant disease detection using machine learning**” ICDI3C, IEEE, 2018.
- [4]. G. Prem Rishi Kranth, M. HemaLalitha, LaharikaBasava, Anjali Mathur., “**Plant Disease prediction using Machine learning algorithms**”. IJCA (09758887), volume 182-No. 25. IEEE, 2018.
- [5]. BudiariantoSuryoKusumo, Ana Heryana, Oka Mahendra, and Hilman F. ardede,”**Machine Learning Based for automatic detection of plant diseases using image processing**”, International Conference on Computer, Control, Informatics and its Applications, 2018.
- [6]. SherlyPuspha Annabel, T Annapoorani, and P Deepalakshmi.,” **Machine learning for plant leaf disease detection and classification—a review**”, ICCSP, IEEE, 2019.
- [7]. Balwant J Gorad, Dr. S. kotrappa, “**Comparativ Study and Review for Development of Disease Prediction System of Indian Crops**”, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 21, pages 68-73, 2019.
- [8]. MonalisaSaha, E. Sasikala, “**Identification of Plants leaf Diseases using Machine Learning Algorithms**”, International Journal of Advanced Science and Technology, Vol. 29, pages 2900-2910, 2020.
- [9]. SrideviSakhamuri, Dr. Vijaya Sri Kompalli, “**An Overview on Prediction of Plant Leaves Disease Using Image Processing Techniques**”, IOP Conf. Series:Materials Science and Engineering, 2020.
- [10]. KrishnaswamyRangarajanAravind and Purushothaman Raja, “**Automated Disease classification in (selected) agricultural crops using transfer learning**” ,Automatika, vol. 61, pages 260-272, 2020.
- [11]. Pranesh Kulkarni et al., “**Plant Disease Detection Using Image Processing and Machine Learning**”, 2021.
- [12]. Sahil Thakur, Darshan Patil et al., “**Plant Disease Detection and Solution Using Image Classification**”, International Journal of Scientific Research & Engineering Trends, Volume 7, 2021.
- [13]. Gianni Fenu, Francesca, “**Forecasting Plant and Crop Disease: An Explorative Study onCurrent Algorithms**”, Big Data and Cognitive Computing, 2021.
- [14]. Kowshik B, Savitha V et al., “**Plant Disease Detection Using Deep Learning**”, ICARD, Vol. 03, 2021.
- [15] JayashriPatil,SachinNaik“**Pomgranate fruit diseases detection using image processing techniques: A Review**”, IT in Industry, Vol. 9, 2021.
- [16] PunithaKartikeyan, Gyanesh Srivastava, “**Review on Emerging Trends in Detection of Plant Diseases using Image Processing with Machine Learning**”, International Journal of Computer Applications, Vol 174, January 2021.
- [17] Shukla A.K. (2020) **Patient Diabetes Forecasting Based on Machine Learning Approach**. In: Pant M., Kumar Sharma T., Arya R., Sahana B., Zolfagharinia H. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 1154. Springer, Singapore. [https://doi.org/10.1007/978-981-15-4032-5\\_91](https://doi.org/10.1007/978-981-15-4032-5_91).