**Precision Agriculture based Intelligent system for early**

**stage Cherry Leaf Disease Identification**

Supriya 1 and Ashutosh Shukla2

1 *Computer Science and Engineering Department,*, College of Engineering Roorkee,

Roorkee,Uttarakhand,INDIA

[supriya@coer.ac.in](mailto:supriya@coer.ac.in)

2 Electrical Engineering Department, UPES University,

Dehradun,Uttarakhand,, ,INDIA  
[2ashutiet@gmail.com](mailto:2ashutiet@gmail.com)

Abstract: Precision agriculture is expected to play a progressively important space in the future of farming, as it allows for more efficient use of resources and helps farmers to meet the growing demand for food in a sustainable way by implementing smart IT techniques for several activities like Site-specific planting and fertilization, Crop monitoring, pest management, variable rate irrigation. Autonomous vehicles, etc. keeping this in view a research work on machine learning have been proposed for classification and detection of healthy and unhealthy leaf ,this will show-cast a comparative analysis of three paramount machine learning methods: Support vector machine, K-Nearest Neighbor, Artificial Neural Network to identify and classify Powdery mildew disease caused by pedosphere pannosa fungus and by considering various evaluation metrics like accuracy, recall, precision, ROC & AUC.it is found that ANN provides best results of 91% on 1,200 images of the standard open source repository.

**Keywords:** Precision Agriculture, SVM, KNN, ANN,ROC,F1-SCOREFirst Section

1. **INTRODUCTION**

### Agriculture is of vital importance to India today as it is the strength of the country's economy. It accounts for approximately 18% of India's GDP and employs around 55% of the country's workforce. Agriculture is also a major source of food for the country's large population, and it plays a critical role in feeding the nation.

### India is facing the challenges of climate change and population growth, which are affecting agriculture. The Indian government has implemented several policies and programs to increase agricultural productivity and to promote sustainable farming practices. FurthermoreTraditional agricultural

### practices that heavily rely on local expertise, customary implements, raw materials, organic fertilizer, and the farmers' cultural beliefs. It is interesting that approximately 50% of people on Earth still use it. Major challenges include population growth, soil degradation, increasing instances of extreme weather events, falling water table, climate change, decline agriculture land, etc.

### The Solutions can be assured by providing timely and regular monitory of agricultural land by surveying the Crop protection, yield projection, and land appraisal measures which are essential elements of world food production and security. These measures can be implemented by the use of Information Technology, this smart way of handing the agriculture task by the use of Information Technology is termed as precision agriculture. [1-2]

### 

### Fig. 1. Challenges affecting agriculture

### Precision agriculture is a farming method that uses advanced technology to optimize crop yields and reduce waste. This method utilizes a combination of precision equipment, sensors, and software to collect and analyse data about soil conditions, weather, crop growth, and other factors. This information is then used to make precise and targeted decisions about planting, fertilizing, and harvesting crops. Precision agriculture can help farmers to increase yields, reduce costs, and improve the sustainability of their operations.

### Precision agriculture relies on a variety of technologies to collect, analyse, and utilize data to optimize crop yields and reduce waste. These technologies include:

### Sensors: These devices are used to measure various factors such as soil moisture, temperature, and pH levels. They can also be used to track crop growth and monitor pests and diseases.

### GPS and mapping equipment: These tools are used to create in depth maps of fields, which can be used to identify areas that need more attention or to pinpoint areas of high yield.

### Drones and unmanned aerial vehicles (UAVs): These devices are used to capture images of fields, In addition to maximizing resource utilization, lowering costs, reducing environmental degradation, and maximizing social and economic efficiency, precision farming also standardizes output.

### D:\AI_ML_AGRI_NPTEL\Smart-farming-architecture.png

### Fig. 2. Smart agriculture components

### It is clear that India has a huge potential for precision farming. It is crucial to create a database of agricultural resources that will serve as a decision support system at the farm. This is a huge effort and a daunting challenge for both agricultural and space scientists who are currently removed from the reality of Indian farming. [3]Sample Heading (Third Level). Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

1. **LITERATURE SURVEY**

Lot of researchers are using advance image processing and machine learning technologies for detecting and classifying plant leaf disease at early stage. Some of the work have been discussed their;

K. Singh et al. in 2019 [14] applied SVM classifier to diagnose the disease of pea leaves. The model successfully diagnose the disease with 89.60% accuracy. Chaitali G. et.al implemented a color and cluster-based segmentation technique for plant leaf disease detection and classification sun burn, yellow mosaic, and grasshopper leaf disease using support vector machine (SVM) classifier of was done by using [4].Kishore G. et.al applied open-source algorithms for the Tomato Plant Leaf disease using image processing techniques built by Image segmentation, clustering. Yogesh Kumar Rana et.al. used methodology to train deep convolutional neural network for extract features using Particle Swarm Optimization to classifying 23 different classes with classified with an accuracy of 97.39%. Khirade et al. in 2015[16] implemented a decision tree model for variety of plant disease detection using GLCM for feature extraction and gives the accuracy of 93%. Xion et al. in 2021[17] a model for tomato plant leaves using the Plant-Village dataset and produced better results of 84.94% as compared with svm and decision tree. Geetha et al. in 2020[18] pre-process the image by segmentation, feature extraction and further classify by implementing the k-nearest neighbors (KNN) algorithm by provides 96.76% accuracy. D. Gupta et al. in 2019[19] implement an ANN model for Citrus plant leaves classifying leaf disease gives 96% accuracy. C. U. Kumari et al.in 2019[20] proposed an ANN model for detection of disease in Cotton and Tomato with accuracy of classification with 92.5 % using ANN.

1. PROPOSED METHODOLOGY

In this work, the dataset of cherry leaves is considered, this dataset was Introduced by Hughes et al. in an open access repository kaggle, it consists of 854 Healthy images & 1052 unhealthy (Powdery mildew) images. The cherry leaves are mainly infected by Podosphaera pannosa fungus.

Podosphaera pannosa is a type of fungal pathogen that affects cherry leaves. It causes a disease known as powdery mildew, which results in a powdery white coating on the leaves, stems, and fruit of infected trees. The disease can weaken the tree and reduce fruit yield. It can be controlled through the use of fungicides, as well as cultural practices such as pruning and proper watering.This fungus causes a serious disease - powdery mildew. Powdery mildew is a kind of fungal virus that affects a huge range of plants, including fruits, vegetables, and ornamental plants. The fungus grows on the surface of leaves, stems, and fruits, forming a powdery white or gray coating. The affected plant parts may become distorted, and the leaves may yellow and drop prematurely. In severe cases, the disease can weaken the plant and reduce crop yields. The fungus thrives in warm and humid conditions, and can be spread by wind and insects. Powdery mildew can be controlled by maintaining proper plant spacing to improve air circulation and reduce humidity, avoiding overhead watering, and using fungicides specifically labelled for powdery mildew control. Some resistant cultivars are also available for some plants. Cultural control methods, such as sanitation and pruning, may also help to reduce the spread of the disease. [4].

An early stage identification technique based on machine learning is proposed to classify diseased in cherry leaves. In this study, a support vector machine (SVM), k-nearest neighbour (KNN) and artificial neural network are three traditional machine learning methods which are evaluated. The ANN achieves the best precise performance in identifying diseased cherry leaves, with the testing accuracy of 91%, as per the computable estimations carried out on a data set of 1,200 images collected from the standard dataset. As a result, ANN can be used to detect unhealthy cherry leaves.

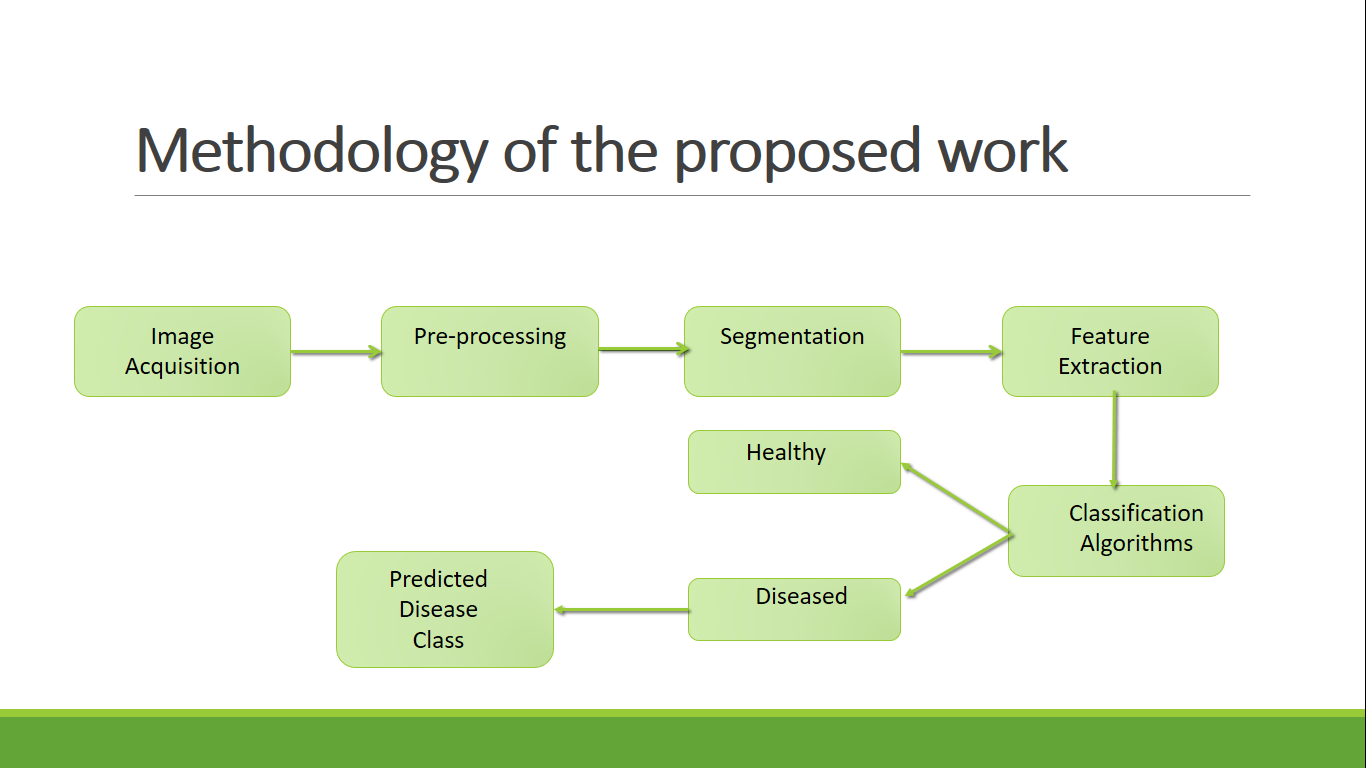


Fig. 4. Flowchart of proposed model

The labelled images of the dataset are segmented and essential feature are extracted by GLCM technique. GLCM (Gray-Level Co-Occurrence Matrix) is a technique used in image processing and computer vision to extract texture features from an image. It Create a co-occurrence matrix by counting the number of times different gray level values occur in pairs at a certain distance and direction from each other. Compute statistics, such as contrast, correlation, energy, and homogeneity, from the co-occurrence matrix. The resulting feature vector will be used to train the model. Then a model architecture is chosen and its weights are initialized. Architecture used in this classification are SVM, KNN and ANN .[5].

The labelled images of the dataset are segmented and essential feature are extracted by GLCM technique. GLCM (Gray-Level Co-Occurrence Matrix) is a technique used in image processing and computer vision to extract texture features from an image. It Create a co-occurrence matrix by counting the number of times different gray level values occur in pairs at a certain distance and direction from each other. Compute statistics, such as contrast, correlation, energy, and homogeneity, from the co-occurrence matrix. The resulting feature vector will be used to train the model. Then a model architecture is chosen and its weights are initialized. Architecture used in this classification are SVM, KNN and ANN .[5]

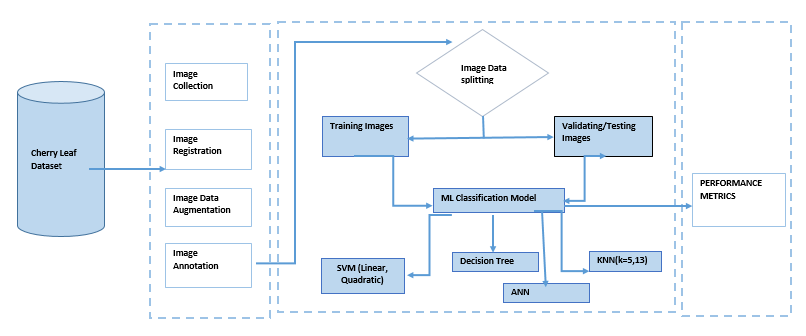


Fig. 5. Several phases of proposed model

The **Supervised Machine Learning** methodology known as SVMs can be useful for both classification or regression tasks. The objective is to find a best fit line (or "hyperplane") that divides the data into various classes. This hyperplane can be selected to minimize the separation between the boundary and the nearest data points for each class termed as support vectors. New data can be categorised by determining which side of the boundary it falls on once the boundary has been established. SVMs can be used with a kernel trick to transform the data into a higher-dimensional space where the data ara particularly effective. [6-7]

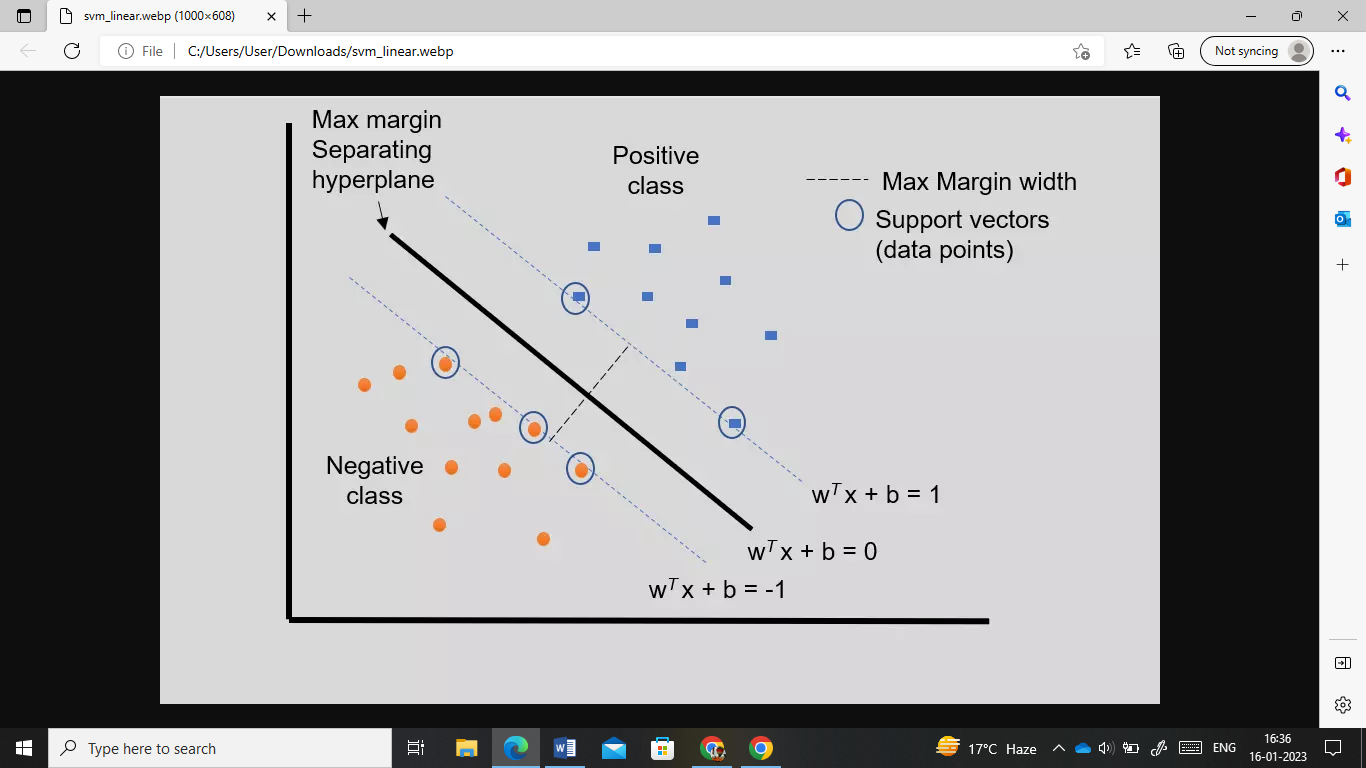


Fig. 6. SVM Hyperplane

Another supervised learning method **K-Nearest Neighbors** (KNN) is used for classification and regression tasks. It is a

Non-parametrized and lazy learning technique for finding the K training data points that are most similar to a new data point and using that group's most prevalent class to categorise the new data point are the fundamental principles of KNN. The value of K, which establishes the number of nearest neighbours utilised to produce the prediction, is a crucial element in the KNN method. Cross-validation is a popular method for determining the best value of K because it can have a big impact on the model's performance. Although the KNN approach is straightforward and simple to use, it can be computationally expensive when the training dataset is large.

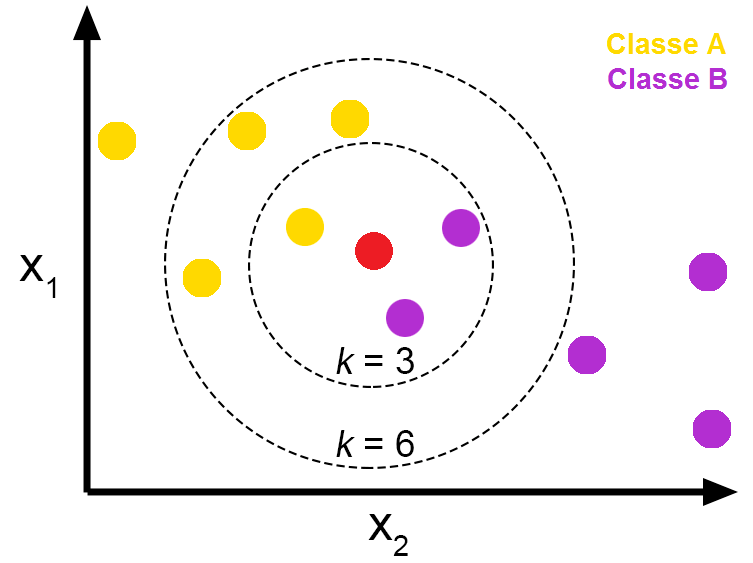


Fig. 7. KNN model at different values of k

In this work, another machine learning algorithm called an Artificial neural networks (ANN) are demonstrated following the design and development of the human brain. Its structure is made up of layers of interconnected "neurons" that transmit information. ANNs are capable of learning from data and generating predictions or decisions without being explicitly trained. They have proven helpful in a variety of settings, such as image processing, NLP and gaming.

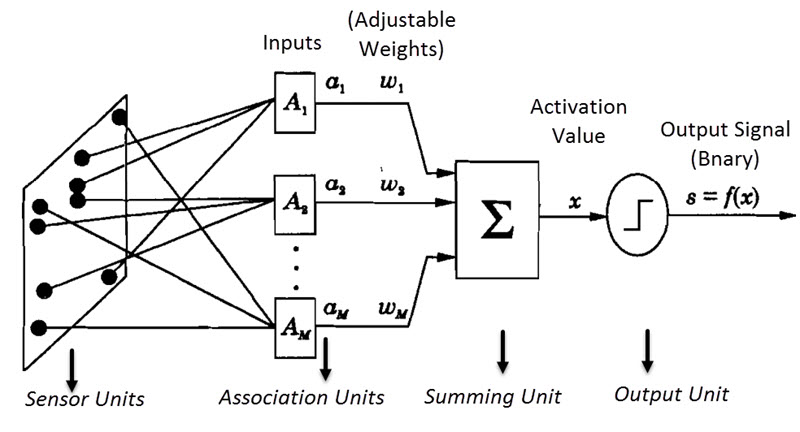


Fig. 7. ANN layers

1. Results and Discussion

TABLE I

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | Algorithm | Accuracy (%) | Precision (%) | **Recall(%)** | **F-Score** |
|  | Decision Tree | 82.6 | 83.66 | 67.77 | 0.56 |
|  | Linear-SVM | 88.0 | 88.2 | 90.2 | 0.89 |
|  | KNN (k=13) | 86.1 | 89.4 | 84.81 | 0.87 |
|  | KNN (k=5) | 85.9 | 88.95 | 81.17 | 0.84 |
|  | SVM (Kernel Quadratic) | 89.7 | 91.39 | 89.82 | 0.90 |
|  | ANN | 90.1 | 91.9 | 91.1 | 0.91 |

Above table, shows the details of the results obtained by the algorithms discussed, by consider these standard metrics: -

(1)

𝑅𝑒𝑐𝑎𝑙𝑙 (2)

𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 (3)

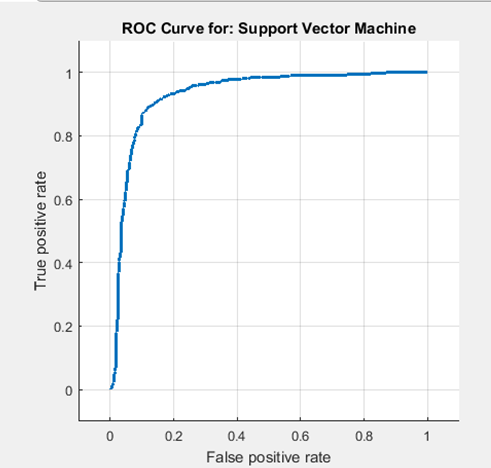
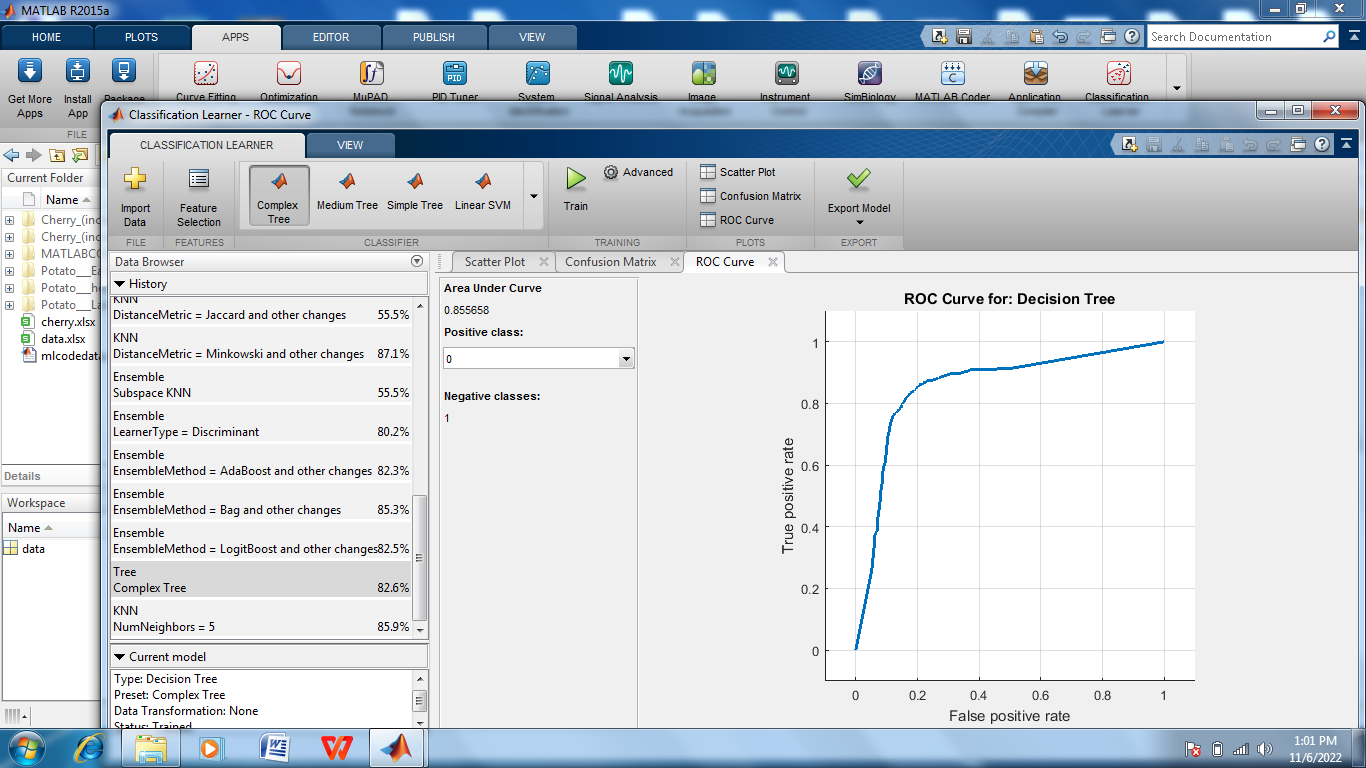
𝐹1 − 𝑠𝑐𝑜𝑟e (4)

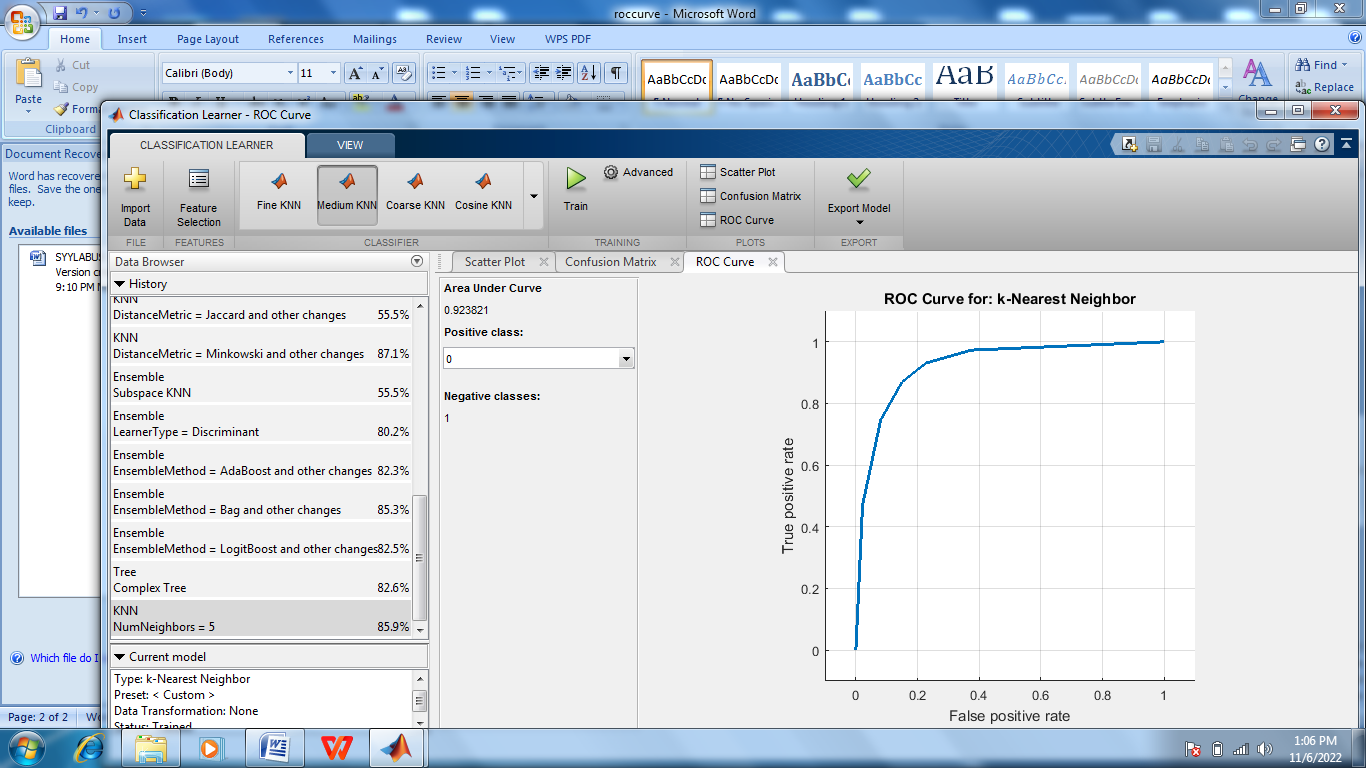
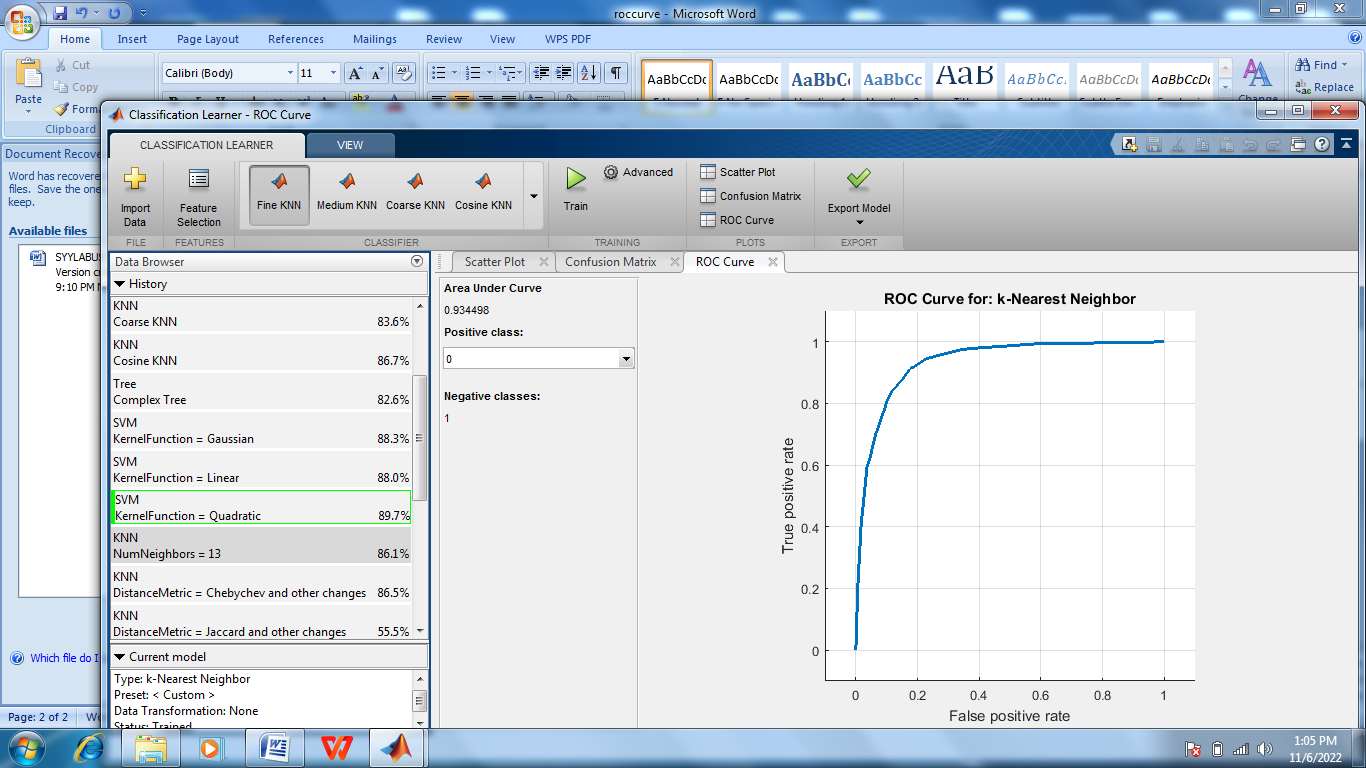
Where, FN=No. of False Negative, FP= No. of False Positive, TN= No. of True Negative and TP= No. of True Positive, in equations 1,2,3 & 4

The proportion of successfully predicted observations to all observations obtained by our model is a spontaneous performance parameter called accuracy. Precision, defined as the proportion of correctly predicted positive observations to all anticipated positive observations, is an acceptable statistic where the overheads of False Positive are noteworthy. Recall is the ratio of correctly predicted positive observations to all actual observations in the class, and is a useful metric to consider for determining the best model when False Negatives have a high cost. The weighted average of Precision and Recall is known as the F1-Score. Therefore, when calculating this score, both false positives and false negatives were taken into account

The Receiver Operating Characteristic (ROC) of each technique is shown below; this graphical plot generated by plotting the true positive rate (TPR) vs. the false positive rate (FPR) at various threshold levels. The number of true positives (actually recognised as positive) divided by the total of true positives and false negatives yields the true positive rate (incorrectly identified as negative). The false positive rate is defined as the proportion of false positives (positive results misinterpreted as negative) to all other erroneous positives and genuine negatives (correctly identified as negative).Area Under the Curve, or AUC, is frequently used to evaluate a system's overall performance.

AUC (Area Under the Curve) of 1 indicates that the classifier is able to effortlessly discriminate between the positive and negative classes. It is generally considered as a perfect classifier.





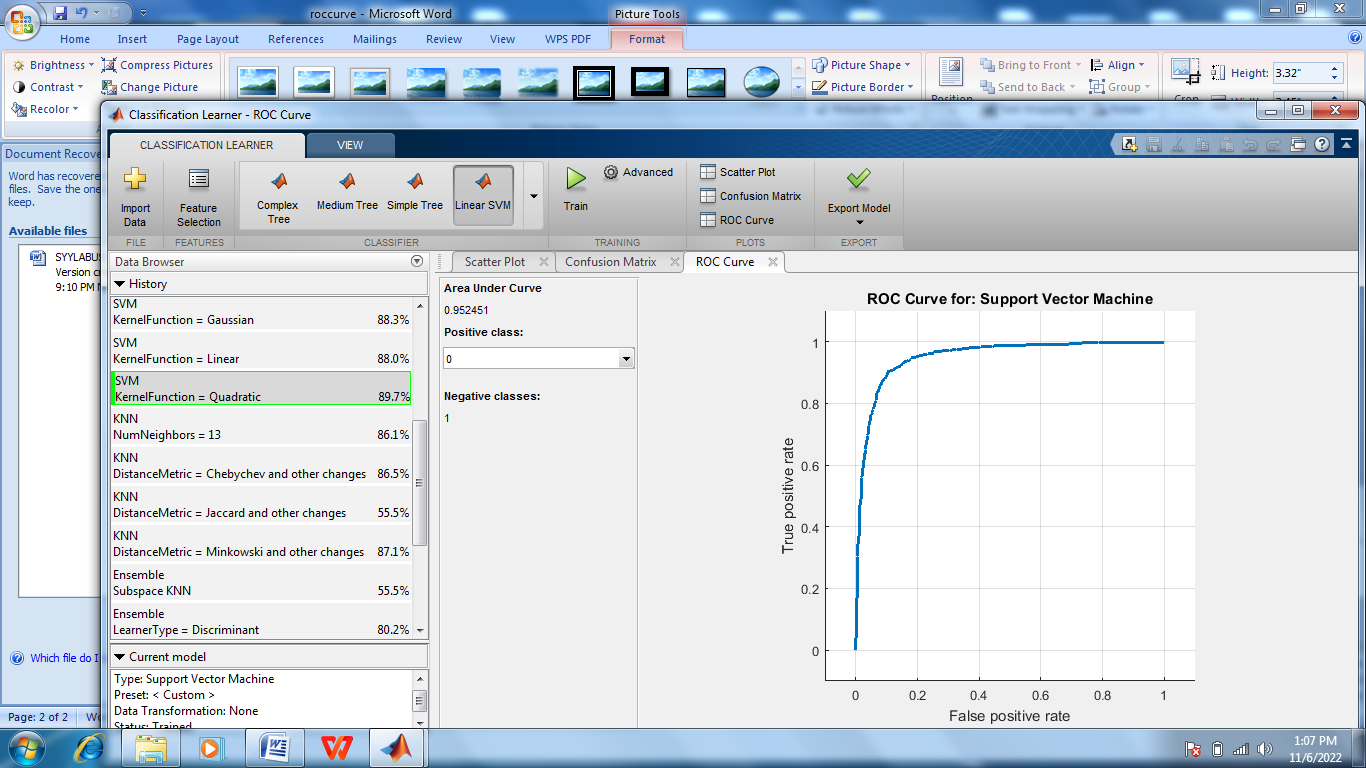


Fig. 8. ROC for decision tree(a.),SVM(Linear)(b.),KNN-3(c.),KNN-13(d.),SVM(Quadric)

From ROC graphs of different algorithms, we can depict that artificial neural network give the highest accuracy during training and validation phase.

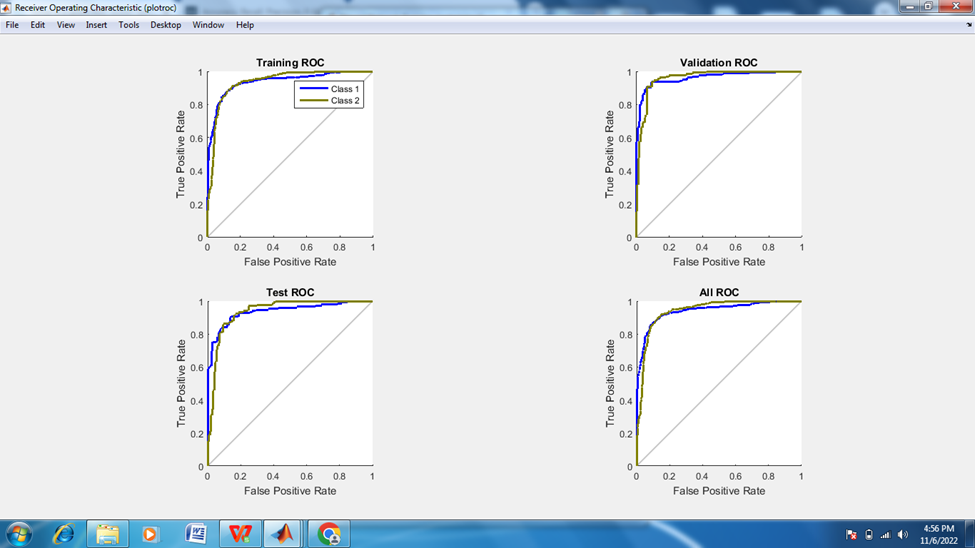


Fig 9. ROC for Training, Validation & Test phases of ANN showing best results.

1. CONCLUSIONS

In conclusion, agriculture is of paramount importance to India today as it is a major contributor to the country's economy, provides food security, and plays a critical role in rural development. it is essential to continuously support the agricultural sector to ensure its continued growth and success by the use of smart agriculture component using machine learning .This work is comparative analysis of three paramount machine learning techniques: Support vector machine, K-Nearest Neighbor, Artificial Neural Network for identify and classify Powdery mildew disease, the accuracy of algorithms like Decision Tree, Linear-SVM ,KNN (k=13) ,KNN (k=5),SVM (Kernel Quadratic),ANN are 82.6% ,88.0%, 86.1%,85.9% ,89.7%, 90.1%, respectively .It can be seen that performance of ANN is highest among others in term of accuracy,precison,recall,F-score and ROC graphs Future work will be carried out the classification for many more diseases in different plant and crops. Besides for the improvement of classification accuracy convolutional neural network will be deployed..

References

1. Neupane, K.; Baysal-Gurel, F. “Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review”, Remote Sens. 2021, 13, 3841.
2. Bauriegel, Elke, and Werner B. Herppich. "Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to Fusarium spec. infections on wheat." Agriculture 4.1 (2014): 32-57.
3. Pothen, Minu Eliz, and Maya L. Pai. "Detection of rice leaf diseases using image processing." 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.
4. Pothen, Minu Eliz, and Maya L. Pai. "Detection of rice leaf diseases using image processing." 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.
5. Al Bashish, Dheeb, Malik Braik, and Sulieman Bani-Ahmad. "A framework for detection and classification of plant leaf and stem diseases." 2010 international conference on signal and image processing. IEEE, 2010.
6. Sinha, Aditya, and Rajveer Singh Shekhawat. "Olive spot disease detection and classification using analysis of leaf image textures." Procedia Computer Science 167 (2020): 2328-2336.
7. Haque, Intisar Rizwan I., and Jeremiah Neubert. "Deep learning approaches to biomedical image segmentation." Informatics in Medicine Unlocked 18 (2020): 100297.
8. Barbedo, Jayme Garcia Arnal. "Plant disease identification from individual lesions and spots using deep learning." Biosystems Engineering 180 (2019): 96-107.
9. Nanehkaran, Y. A., et al. "Recognition of plant leaf diseases based on computer vision." Journal of Ambient Intelligence and Humanized Computing (2020): 1-18.
10. Ebrahimi MA, Khoshtaghaza MH, Minaei S, Jamshidi B (2017) “Vision-based pest detection based on SVM classification method”,Comput Electron Agric 137:52–58.
11. Gaura, Jan, Eduard Sojka, and Michal Krumnikl. "Image segmentation based on k-means clustering and energy-transfer proximity." International Symposium on Visual Computing. Springer, Berlin, Heidelberg, 2011.
12. K. Singh, S. Kumar, P. Kaur, “Support vector machine classifier based detection of fungal rust disease in Pea Plant (Pisam sativam),” Int. j. inf. tecnol. 11, 485–492 (2019).
13. Disease Detection Using Deep Learning Techniques," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 979-983, doi: 10.1109/ICCES48766.2020.9137986.
14. Yadav, Rishabh, Yogesh Kumar Rana, and Sushama Nagpal. "Plant leaf disease detection and classification using particle swarm optimization." International Conference on Machine Learning for Networking. Springer, Cham, 2019.
15. Singh, Vijai. "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization." Artificial Intelligence in Agriculture 3 (2019): 62-68.
16. Xian, Tan Soo, and Ruzelita Ngadiran. "Plant Diseases Classification using Machine Learning." Journal of Physics: Conference Series. Vol. 1962. No. 1. IOP Publishing, 2021..
17. A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, D.P. Hughes, (2017), “Deep learning for image-based cassava disease detection” in Frontiers in plant science, 8, 1852.
18. Geetha, G., et al. "Plant leaf disease classification and detection system using machine learning." Journal of Physics: Conference Series. Vol. 1712. No. 1. IOP Publishing, 2020.D. Gupta, R. Singh, (2019), “Disease Detection in Plant using Artificial Neural Network,” in International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 7 Issue VI, June 2019. Page No: 2560-2565.
19. E. Hossain, M. F. Hossain and M. A. Rahaman, "A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679247.
20. Trivedi, Jay, Yash Shamnani, and Ruchi Gajjar. "Plant leaf disease detection using machine learning." International conference on emerging technology trends in electronics communication and networking. Springer, Singapore, 2020.
21. N. Prashar, "A Review on Plant Disease Detection Techniques," 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), Jalandhar, India, 2021, pp. 501-506, doi: 10.1109/ICSCCC51823.2021.9478132.
22. Gupta, Varun, et al. "Automated segmentation of powdery mildew disease from cherry leaves using image processing." 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI). IEEE, 2017.
23. Chaschatzis, Christos, et al. "Detection and characterization of stressed sweet cherry tissues using machine learning." Drones 6.1 (2022): 3.