

Apple Leaves Disease Detection Using Machine Learning Approach

Kulbir Kaur Sandhu

Assistant Professor, Department of Computer Science

Baba Farid College, Bathinda

kulbirkaursandhu01@gmail.com

Abstract: Agriculture plays an indispensable role in the development of the country especially in the growing country like India where most of the people's revenue is generated from agriculture. Disease affected crops leads to the loss of crop productivity. Therefore, leaf disease prediction in apple cultivation is of considerable importance to overcome these problems. The proposed work intends to predict different disease in apple leaf like apple scab and marssonina using different algorithms like K nearest neighbor(KNN),support vector machine(SVM), classification decision tree, regression decision tree and Naïve Bayes. From the simulation result, it can be concluded that KNN performs better as compared to other algorithms in terms of accuracy of disease prediction.

Keywords: SVM, KNN, Apple disease, Marssonina, Naïve Bayes, inverse, Apple scab.

1. INTRODUCTION

India is ranked number two in the production of fruits. Sixty percent of the population depends on agriculture for employment. The methodology used for fruit illness recognition is perception through naked eyes by specialists. In some countries, such specialists are costly and tedious because of they are not accessible. Damage can also be seen in areas such as stems, leaves, and branches of the tree . Apple rot,apple scaband apple blotch are some of the most common diseases of apple fruits. Apples with gray or brown corky blotch have apple scab diseases . If the surface of the apple have black specks that may be covered by a red halo or the fruit is faintly hallow, circular brown then, it is referred to as apple rot type diseases. Apple blotch type diseases have surface with dark,irregular or lobed edges .The diagnosis of apple diseases are mostly performed by an expert, on site by inspecting the leaves and fruits visually. In more complicated or new cases,analysis in laboratory can be done. However, it needs experts that demand specialized training and this approach increases the overall cost. The need for experts limit scale. Since experts are often specialized in a few types of disorders the effectiveness may reduce.

There are mainly two type of images and these are Analog and Digital image. Processing of image means to turn an image into its digital form and then different kind of operations will be performed on it to get some kind of results with enhanced type of same image. These results will give us the useful information which can be extracted from it and can be very helpful. There are many type algorithms which can be implemented to perform a particular type of operations or task. Sometimes when some diseases are not visible to naked eye but actually they are present, then it is difficult to detect it with the naked eye. And when it is visible it will be too late to detect disease and can't help anymore. Earlier, this disease was detected with the microscopes but it was very difficult process to observe each and every plant and leaf. So now remote sensing techniques are developed and used in computer science. They can detect the disease from the multilevel and hyper spectral images of plants which can be digitally captured.



Fig.1. Apple Leaf disease

In this paper we report on common disorders of apple tree leaves (*Malus domestica Borkh*), which are automatically classified by training and applying machine learning models to images of apple tree leaves (*Malus domestica Borkh*). Our hypothesis is that, in particular deep belief network and related training algorithms, which are the current state-of-the-art models, are able to attain better Performance comparatively to human experts. There are various classifiers used to detect disease.

2. RELATED WORK

There are a few commercial systems in use in order to diagnose various disorders, although not restricted or applicable to apple trees. In this paper apple rot, apple blotch and apple scab diseases have been covered. K-means Clustering has been used for image segmentation process. In the next step features such as color coherence vector (CCV), Global color Histogram (GCH), local binary pattern (LBP) and complete local binary pattern (CLBP) are extracted. In paper different methods like Local Binary Pattern, Color Invariant, Color histogram are used for the feature extraction from images and further classification is carried out with the help of neural network. Paper aimed to distinguish healthy sugar beet leaves from dangerous ones, to differentiate between three forms of diseases and to discover illnesses even before, early unique signs and symptoms have begun to show up. Support Vector Machines are used by the author and as enter they used 9 spectral plants indexes, as functions, resulting in classification accuracies up to 97% while differentiating healthy from dangerous, 86% while distinguishing between 3 illnesses and among 65% and 90% for per-symptomatic detection of sicknesses. Notably, to gain the spectral images and significant feature selection this technique requires specialized hardware. In the paper five samples are considered. 3 samples of grapes leaves and 2 samples of apples leaves are picked up for detection. As a result 90% accuracy has been achieved using neural network algorithms. In the paper from leaf images six diseases has been categorized. The authors used from each of the six classes of 32 leaves and for classification, performed a Multilayer Perceptron. Features were manually extracted from the image and defined its 10 texture features and showing accuracies between 83% to 94%.

3. APPLE LEAVES DISEASES AND SYMPTOMS

Proposed work includes number of attacks on apple leaves. Following are the some of the apple leaves diseases and its symptoms :

1. Apple Scab: This disease is usually seen on leaves and fruits. Surface beneath leaves, have spots which are velvety and may coalesce to cover the leaf entirely. Leaves usually get crumple or puckered and have black, circular spots on their upper surface. The leaves which are acutely affected may become yellow and drop. Flowers could drop if the flower stems get infected. The scrap later become hallow and turns brown and may have spores around their borders. Secondary organisms could enter if the fruit becomes contort and crack open.



Fig.2 Apple Scab

2. Marssonina leaf blotch:The symptoms start to show up as dark green circular patches on the surface above the leaf, which turns to 5-10 mm brown leaf spots, which later turns darker. On maturing it can also be seen on lower surface of the leaf. Small black acervuli are visible on the surface of leaf. When lesions are innumerable, they coalesce to form large dark brown blotches and areas around become yellow.



Fig.3 Marssonina leaf blotch

3. Black Rot Canker:The symptoms usually can be seen around early spring when the leaves are unfolding. Tiny purple specks on the upper surface of leaves appear which, enlarges into circular lesions of 1/8 to 1/4 inch (3-6 mm) in diameter. The center of the lesions become tan to brown, whereas the boundaries remain purple. Within weeks, secondary enlargement of spots occur. Acutely infected leaves become chlorotic and leaves start to fall. As the affected area expands, a series of brown and black uniform size concentric bands form. Rotted area's flesh stays leathery and solid. On the surface of infected fruit Black pycnidia are often noticed. Wound in the bark and Lesions resulting in canker formation are usually related.

4. Alternaria leaf spot/blight:During late spring and early summer spots occur on the leaves. Initially, they are brown with purple border, round and 1/8 to 1/4 inch in diameter. They often turn tan to ash gray, as they mature. Some spots become irregularly shaped as they go through secondary enlargement. Defoliation happens as the acutely infected leaves often abscise. Small, dark, raised pimple-like lesions related with the lenticels are caused by fruit infections. Round, blackish spots bordered by cracks and hollow stem lesions appear on susceptible cultivators such as Indo but have not been observed on Delicious.



Fig.4(a) Blight

4. METHODOLOGY

In this paper we have presented some of the color and texture features, which are extracted from the test images. Some color and texture features are combined together for better precision after feature extraction. Finally the apple leaf diseases are classified using various classifiers. Further classification is done by KNN which is a better approach, if the leaf is found to be infected by any of the diseases .

Steps of proposed methodology: The main steps for methodology are:

A. Brightness Preserving Dynamic Fuzzy Histogram Equalization for preprocessing

Histogram in broader sense refers to a graph between pixels value or intensity vs pixels probability. In other words we can conclude this as a graph between pixels intensity vs pixels occurrence and frequency. Equalization refers to changing the intensity of pixels to make it more enhanced. It finds a great use in those images which become so light that they cannot be viewed properly. There are many techniques to perform histogram equalization but justification for the choice of Fuzzy histogram lies in the discussion that it does not remap peak of histogram and further does not harm mean brightness of image and preserve it in a better way.

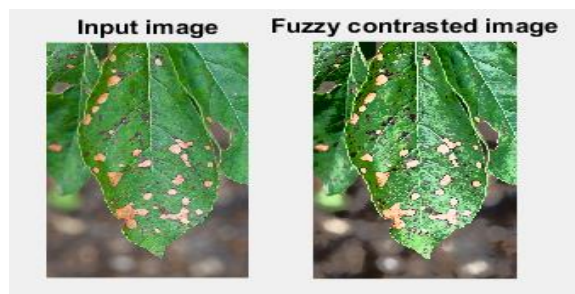


Fig.4(b) Preprocessing using fuzzy function

B. Color based image segmentation using k means clustering

When an image has different colors and we need to segment colors into different clustering an approach of conversion of RGB image into $L a b$ is followed, where L is a luminous layer, a is a chromatic layer and b is another chromatic layer. Once we obtain the above designed image, we found cluster centers and cluster index. According to these results we will be able to separate different parts of image and different pattern. This enables one to examine every part of image and helps in detection of image disease. In the proposed approach, we have used k means clustering to cluster the different parts of the image into three different clusters.

C. Feature selection using Gray Level Co-Occurrence matrix

Feature extraction using gray level co-occurrence matrix (GLCM) is used. It is a statistical method for examining texture and which takes spatial relationship of pixels into consideration. It calculates such pairs of pixels having a specified relationship between the pixels. In an algorithm different features like contrast, co-relation, energy and homogeneity are calculated and other properties like mean, standard deviation, entropy, root mean square variance, kurtosis and skewness are calculated from segmented image. Segmented image is formed using k-means clustering which has been discussed and explained in the previous section.

D. Classification using different machine learning algorithms

Classification is one of the most important task for different applications such as text categorization, image classification, disease prediction, data classification and many more. It is the process which is implemented on features extracted using any feature extraction method. With those features, machine is trained and model gets ready to perform real time prediction. Classification can be performed using different methods such as Support Vector Machine, K nearest neighbors, Naives Bayes, classification decision tree. However decision tree can be used to predict both classification and regression which is called classification and regression tree (CART). In this part we will discuss different algorithms we have implemented to identify disease and their application and we will show their graph and area they are covering.

Let us discuss each of them in detail:

E. Support Vector Machine: It is a statistical technique which is a new promising, non-linear and non-parametric classification. It fund its wide range of application in medical diagnostic, optical character recognition, disease prediction,

disease detection and many more. During training of SVM model, kernel function is passed as a parameter which helps to do certain calculation faster and decreases time. If one has few points in a high dimensional space, linear SVM is a better option.

F. KNN: It is one of the simplest classification algorithms. KNN has some of the features which make it handier like easy to interpret output and calculation is less. It is a non-parametric lazy learning algorithm. From lazy, I mean it does not use any generalization or in simple words we can say that its training phase is pretty fast. It makes use of distance of new data to every cluster and depending upon distance it classifies new data into a particular cluster. KNN in Matlab is implemented using different distance calculating methods like Minkowsk, Euclidean, Mahalanobis. It is very important in KNN to have a good choice of the number of clusters. Validation error curve helps us to choose the optimum value of K. If we want to plot lot of points in a low dimensional space, KNN is a better option.

G. Naïve Bayes: These classify data based on probability and

is called probabilistic classifier and is based on Bayes theorem. One of the wide applications is automatic medical diagnosis. It uses the following formula:

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

If you are working with a dataset having millions of records with some attributes, Naïve Bayes approach is one of the good approaches widely used in Natural Language Processing. It provides probability of a real time data belonging to a particular class. Class with the highest probability is considered as the most likely class. It is fast, scalable algorithm used for both binary and multiclass classification which can be trained on a small dataset.

H. Decision Tree: This approach is used in statistics, data mining and machine learning. Decision tree categorize data points. Data is passed to different nodes of the tree based on some conditions and data is assigned to respective nodes. Top node is called root whereas nodes which cannot be classified are called leaf nodes. Further at each node probability of each data point is calculated and it gives an impression of its occurrence at a particular node.

Decision tree has many advantages and some of them are shared below. Decision tree helps in selecting features automatically as top few nodes are essential main features. Even missing values will not prevent splitting data for building trees. Even non-linear relationships between variables do not affect tree performance and is very easy to interpret and explain.

In the proposed algorithm, we have applied the above machine learning algorithms to given dataset which contains apple disease like apple Scab and Marsonina Coronaria. The detailed version of each step and dataset will be given in the experimental discussion part. Here, I am showing you a graph showing how apple disease is predicted by different machine learning algorithm and KNN performs better as compared to other above mentioned algorithm.

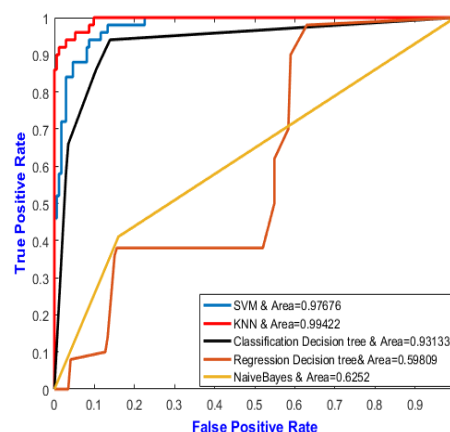


Fig. 5 Graph shows the performance of different algorithms to detect the apple disease.

5. RESULTS AND DISCUSSIONS

Simulation experiments in this article are done on a laptop with Intel(R) Core (TM) i7 CPU Q820@ 1.73 GHZ and 32 GB memory. The proposed KNN-based apple disease detection approach is designed with Matlab running on Windows 7.

The datasets used for experiments were constructed based on real sample images of diseased apple leaves approx. These dataset were collected from different farms in February and March at temperature between 16 and 20 degree.

In this section we are going to evaluate its experimental results. As a result it is concluded that KNN is best classifier among other classifiers. During experimental evaluation we take comparison using two factors i.e. True positive rate and false negative rate for all the algorithms.

True positive rate and false negative rate plays a very important role in evaluating many machine learning algorithms and holds a great significance. True positive refers to all the positive cases which are record correctly whereas false positive means all the positive case which is not recalled correctly. If we look at the following graphs for small false positive rate KNN holds good true positive rate as compared to other methods. In other words we can also conclude that KNN predicts the disease with more accuracy.

The key advantage of KNN lies in its ability to get trained even by noisy data and is quite effective if the training sample is large. Further from the graph it can also see that when distance, distance weight as well as with different NS methods leads to different accuracy but still KNN proved to be a better approach.

Above experiment is performed with apple leaves diseases dataset and accuracy is calculated using KNN and comes out to be more than 96% which can be justifies in following figure 6.

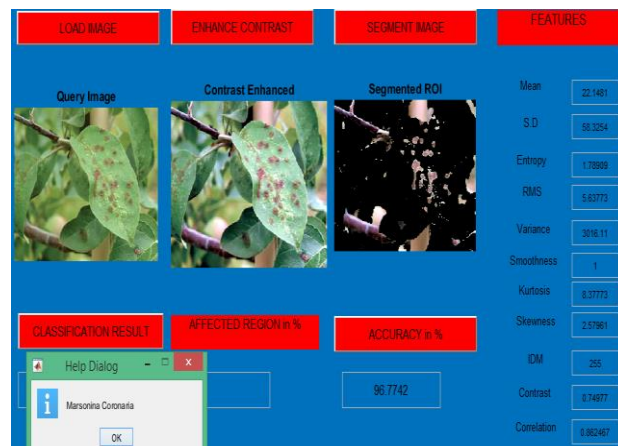


Fig. 6 Calculation of Accuracy using KNN

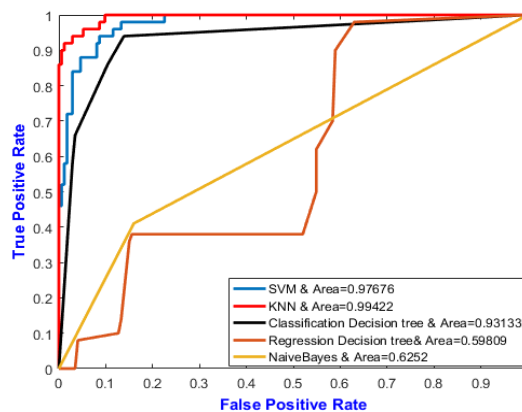


Fig.7 shows the performance of different algorithms to detect apple diseases with Neighbor=4 and distance Minkowsk

Figure 7 depicts the True positive rate and false negative rate of various classifiers. It is concluded that KNN gives better accuracy as compared to the other classifier.

Table 1: Performance comparison of various classifiers with neighbor 4

Name	Neighbors	NS Method	Distance	Area
KNN	4	Exhaustive	Minkowsk	0.99422
SVM	-	-	-	0.97676
Classification detection Tree	-	-	-	0.93133
Regression Decision Tree	-	-	-	0.59809
Naïve Bayes	-	-	-	0.6252

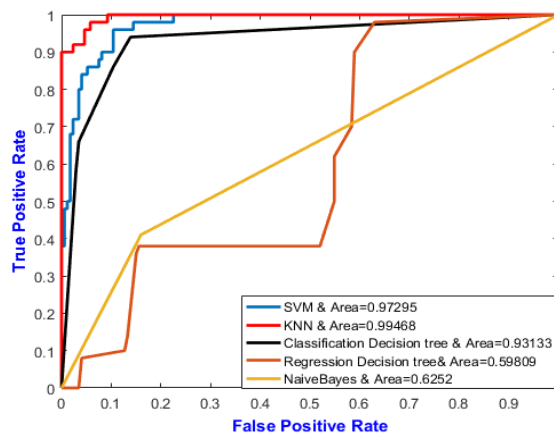


Fig.8 shows the performance of different algorithms to detect apple diseases with neighbor=5 and distance Euclidean

Table 2: Performance comparison of various classifiers with neighbor 5

Name	Neighbors	NS Method	Distance	Area
KNN	5	Exhaustive	Euclidean	0.99468
SVM	-	-	-	0.97295
Classification detection Tree	-	-	-	0.93133
Regression Decision Tree	-	-	-	0.59809
Naïve Bayes	-	-	-	0.6252

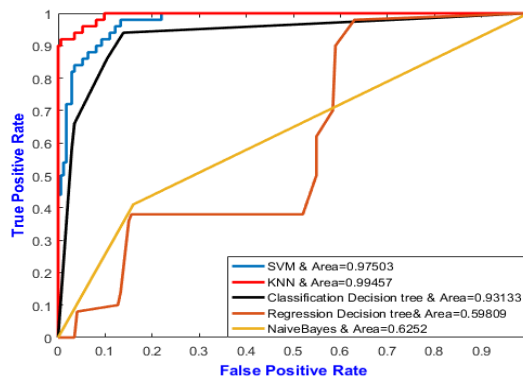


Fig.9 shows the performance of different algorithms to detect apple diseases with neighbor=6 and mahalanobis

Table 3: Performance comparison of various classifiers with neighbor 6

Name	Neighbors	NS Method	Distance	Distance Weight	Area
KNN	6	Exhaustive	mahalanobis	Inverse	0.99457
SVM	-	-	-	-	0.97295
Classification detection Tree	-	-	-	-	0.93133
Regression Decision Tree	-	-	-	-	0.59809
Naïve Bayes	-	-	-	-	0.6252

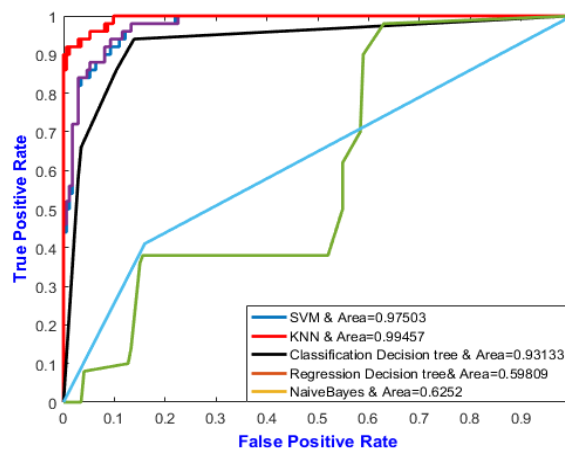


Fig. 10 shows the performance of different algorithms to detect apple diseases with neighbor=6 and otherwise

Table 4: Performance comparison of various classifiers with neighbor 7

Name	Neighbors	NS Method	Distance Weight	Area
KNN	7	Exhaustive	SuperInverse	0.99457
SVM	-	-	-	0.97295
Classification detection Tree	-	-	-	0.93133
Regression Decision Tree	-	-	-	0.59809
Naïve Bayes	-	-	-	0.6252

6. CONCLUSION

As we know, we are living in the era of digital technology, so we should make the maximum use of it in all the sectors. Agriculture covers major area of India’s geography. More than 70% people depend directly on agriculture. Various crops and fruits suffer from different disease and this produces fewer yields. So investing huge amount of money in cultivating a crop selected by just intuition or judgment is not a wise option today. Therefore, machine learning came into play. In the proposed work, we primly focused on apple diseases (apple scab and marssonina) and produces different results on disease prediction using artificial intelligence approach. Further as the part of experiment we proved priority of KNN over other machine learning algorithms in terms of accuracy. We have plotted the graphs in Matlab showing the variation of true positive rate vs false positive rate and using these values accuracy is calculated. From the graphs it is clear that with KNN, we can achieve accuracy more than **96%**.

REFERENCES

- [1] Shiv Ram Dubey, Anand Singh Jalal," Detection and classification of Apple Fruit Diseases Using Complete Local Binary Patterns", 978-0- 7695-4872-2/12 IEEE 2012.
- [2] R. Sivamoorthi, 2Dr. N. Sujatha," A Novel Approach of Detection and Classification of Apple Fruit Based on Complete Local Binary Patterns", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, April 2015.
- [3] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977. Gilmar Ribeiro Nachtigall
- [4] Monika Jhuria, Ashwini kumar,Rushikesh Borse, "Image processing for smart farming: detection of diseases and fruit grading ", 978-1-4673- 6101-9/13/2013 *IEEE* .
- [5] H Al-Hiary, S Bani-Ahmad, M Reyalat, M Braik, and Z ALRahamneh.Fast and accurate detection and classification of plant diseases. *Machine learning*, 14:5, 2011.
- [6] Revathi and M Hemalatha. Identification of cotton diseases based on cross information gain deep forward neural network classifier with pso feature selection. *International Journal of Engineering and Technology (IJET) ISSN*, 2014.
- [7] Dennis W Ruck, Steven K Rogers, and Matthew Kabrisky. Feature selection using a multilayer perceptron. *Journal of Neural Network Computing*, 2(2):40–48, 1990.
- [8] [8] T Rumpf, A K Mahlein, U Steiner, E C Oerke, H W Dehne, and L Pl`umer. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74(1):91–99, 2010.
- [9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXivpreprint arXiv:1409.1556*, 2014.
- [10] Shiv Ram Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta, "Infected Fruit Part Detection using K-Means Clustering Segmentation Technique", International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 2, No 2,2013.
- [11] Haloi, Mrinal. "Improved microaneurysm detection using deep neural networks." *arXiv preprint arXiv:1505.04424* (2015).
- [12] Srivastava, Nitish, et al. "Dropout: A simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15.1 (2014): 1929-1958
- [13] M. C. S. Geetha, "Application of classification technique in data mining for agricultural land," IJARCCCE, pp. 352–355, Mar. 2015.
- [14] Bangladesh Agricultural Research Council: Climate Information Management System. [Online] Available: <http://climate.barcapps.gov.bd>. Accessed: Jan. 28, 2017.
- [15] M. Somvanshi and P. Chavan, "A review of machine learning techniques using decision tree and support vector machine," 2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, India, 2016, pp. 1-7. doi: 10.1109/ICCUBEA.2016.7860040
- [16] K-Nearest Neighbors Algorithm [Online]. Available at: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#kNN_regression Accessed: 27 Nov. 2016.
- [17] P. Tan, M. Steinbach and V. Kumar, Introduction to data mining, 1st ed. Dorling Kindersley: Pearson, 2015, p. 149.
- [14] Yearbook of Agricultural Statistics - (2008-2015) [Online] <http://bbs.gov.bd> Accessed: 27 Nov. 2016.
- [18] F. Baheri, F. Davardoost and V. Ahmadzadeh, "Data mining with learning decision tree and Bayesian network for data replication in Data Grid", Proceedings of the Third International Conference on Contemporary Issues in Computer and Information Sciences, pp. 51- 51, 2012