Machine Learning based Image Classification of Papaya Disease Recognition

Conference Paper · November 2020

DOI: 10.1109/ICECA49313.2020.9297570

Md. Ashiqul Islam
Daffodil International University

Md. Shahriar Islam
Daffodil International University

Md. Sagar Hossen
Institut Teknologi Sepuluh Nopember

Minhaz Uddin Emon
Jahangirnagar University

Machine Learning Based Image Classification of Papaya Disease Recognition

Md. Ashiqul Islam [1]
Student, Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
ashiqul15-951@diu.edu.bd

Md. Shahriar Islam [2]
Student, Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
shahriar15-1341@diu.edu.bd

Md. Sagar Hossen [3]
Student, Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
sagar15-1504@diu.edu.bd

Minhaz Uddin Emon [4]
Student, Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
minhaz15-1294@diu.edu.bd

Maria Sultana Keya [5]
Student, Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
maria15-1215@diu.edu.bd

Ahsan Habib [6]
Student, Dept. of Computer Science and Engineering
Eastern University
Dhaka, Bangladesh
ahredoan@gmail.com

***Abstract*— To help distant farmer and rural people of Bangladesh, many research have been performed in recent year for recognize the papaya diseases that takes a great deal of advantage into machine learning fields. That research is mainly required to support for agriculture. This advantage is highly effective and helpful for papaya cultivation. The primary objective of this paper is to compare some algorithms into papaya disease recognition and identify ailment by capturing the image and classify them based on their diseases using an intelligent system. To overcome this advantage the recognition of papaya disease are mainly involves two challenges those are detecting the disease and another is classifying the diseases based on their symptoms. In our proposed system we are presenting an online machine learning based papaya disease in which a person captures an image via mobile app and sends it to the system for detect the disease and also compare some algorithms accuracy those are random forest, k-means clustering, SVC and CNN. The system process the images and will give feedback. This intelligent system can easily detect the diseases and we are getting high accuracy 98.4% to predict the papaya diseases.**

Keywords— Machine Learning, Papaya Cultivation, Intelligent System, Disease Recognication.

I. INTRODUCTION

Bangladesh is a hugely populated country and its population is around 166.59 million (2019), where the remarkable population depends on agriculture. This agricultural has a remarkable contribution to GDP, which comparison 13.07% (2019) of the whole GDP [2]. When this agricultural product satisfies the country's demands it has exported to another country. To maintain this status, it is important to produce fruits with fresh and disease-free. In Bangladesh, the literate rate is 73.9% (2019) and others are illiterate, but most of the farmers are illiterate [3]. They don't have proper knowledge about cultivation and their diseases. Even many farmers are hardly trained in their profession but sometimes a few problems come out to the cultivation. Many agricultural areas have a huge distance from their agricultural center, and many times they could not go to the agricultural center for transportation problems. Even if the farmer needs any suggestion they can't go to the proper place to support their work, and their ignorance cause hampers the crops. They are all right with the antiquated horticultural devices and systems. If they can produce a better product, it will be enormously profitable for them and will make a better place in our GDP and export the crops. In our process, images through an expert system and the system determine the diseases of the crops. Among numerous rural items, we have picked papaya for our work which is a popular and enormous cultivated fruits in Bangladesh. During the cultivation of papaya, it's attracted by many kinds of diseases causing unusual damage [1]. Papaya fruits are mainly damaged by a fungal disease which is one of the major problems to cultivate this product. For this contamination, farmers face a large amount of economic loss.

In this examination, we performed sickness acknowledgment of papaya organic products with a machine learning-based framework which takes pictures and perceives the pictures with the rottenness of any sort of illness of papaya organic products. In this system, firstly we describe some algorithms (Random forest, K-means clustering, SVC, and CNN) and their accuracy level. Based on this accuracy level we will use to detect and recognize the diseases of the papaya fruits via intelligent mobile apps. The framework should work continuously.

II. Related Work

Veeraballi et al [1] build up a framework for deep learning the way to deal with papaya leaves disease characterization and classification. Deep learning and image processing average result of accuracy 85.1%, sensitivity 0.90%, and specification 0.61%. Behera et al [4] build up an article to classification papaya fruits which utilize machine learning and transfer learning approaches. Asraf et al [5] classified their study as taking an oil palm leaves sample. They applying the Polynomial kernel process with a soft margin and it gave an average of 95% accurate classification. Islam et al [6] introduced a methodology their examination that coordinates picture preparing and AI to permit diagnosing ailments from leaf images. Ventura et al [7] perform a study to collect all the papaya disease and the integrated control of this disease. Din et al [8] perform a study about the online papaya disease diagnostic system (OPDDS). They use artificial intelligence concepts to feature an expert system. Pothen et al [9] perform a study to detect rice leaf disease, they use image processing techniques. They use the SVM classifier and gain 94.6% accuracy with polynomial Kernel SVM and HOG. Samajpati et al [10] perform a hybrid approach to detect apple fruit disease. They use a random forest classifier to detect the symptoms. Nadeem et al [11] perform a study to find out papaya leaf curl disease fast report in Pakistan. They determine a relationship between cotton leaf curl and papaya geminivirus infection. Porter et al [12] perform a study to develop and coordinate the use of papaya disease resistance and provided valuable sources for nutrition. Hewajulige et al [13] perform a study on chitosan coating on anthracnose disease control in papaya. Ming et al [14] perform a study about Transgenic Tropical Fruit Papaya's draft genome. Camargo et al [15] build up a system that machine vision system to implement crop disease and used some color image to detect symptoms. Yu et al [16] perform a physical papaya BAC map that was built with the genetic map and genome sequence. Habib et al [17] presented a system of agro medical experts based on machine vision. They presented a two-character set here Composed of 10 characteristics to solve the recognition of a papaya disease. Vijayakumar et al [18] developed a method to live pictures collected at the various stages of the dragon fruit was used to recognize the mellowness of the dragon fruit and the training and testing process. Chen et al [19] presented a study that provides a detailed analysis of the inner tissue of strawberries and exterior tissue due to external mechanical forces, which it is subjected from the moment and harvested from the plant to the moment then it hits the customer. Kelly et al [20] perform a study to maintaining a steady optimum temperature in the supply chain is necessary to reduce the efficiency of strawberry losses and ultimately waste. Liang et al [21] presented that the covering of the edible layer would have a positive impact in order to prolong and preserve the firmness of the fruit. As a result, approximately 1% of chitosan is coated on strawberry, which increases the visco-elastic and gas permeability properties and enhances the shelf life of the fruit. Chandy et al [22] analyzed that studying the growth of fruits after harvesting, the researchers focused on the improvements in firmness that were detected. It also relied on the method of fresh-keeping. Manoharan et al [23] focusing on the benefits of using the Hermitian wavelet transformation implemented in the graph wavelet detection function.

III. System Architecture

In this section, we will present an outworking architecture for our proposed model. Using an intelligent mobile app, capture an image of the inflected fruits and send it to the system that recognizes the fruit based on their diseases. To recognize the papaya disease at first we will load all the training and testing images into our python IDE. After complete the training and testing process, we need to extract the feature using color features (Local Binary Pattern, Complete Local Binary Pattern, Global Color Histogram) [9].

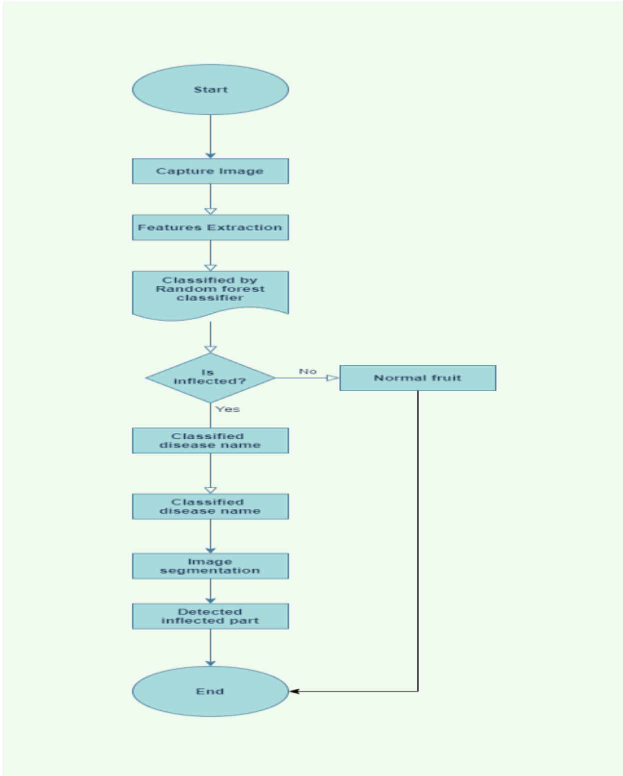

Fig.1. System architecture of the proposed approach

Apply random forest classifier to the segmented image which is already processed. To classify the image random forest algorithm show higher accuracy and it also uses for regression task. Random forest algorithm selects some random sample from the given dataset [7]. Draw a tree sample from the original data and get a prediction for the generating tree. Each of the prediction results performed by voting and select the best result for the final prediction. Now apply K-means clustering for the segment of the image. To complete the segmentation part, load the image which converted RGB color space to grayscale and make all the images that segment the image by color. Make sure the segment part is containing the disease. If the fruit is disease free then recognize it as a normal fruit otherwise classified the papaya fruit disease detects the inflect part.

Image segmentation is a process where images are segmented out to locate objects, and their boundaries such as lines, curves, etc. This process also helps with assigning every pixel of the images and using that pixel and the same layer snack a certain character. To recognize an image segmentation is a part of it, these techniques isolate the specific part of the object from the images for a better analysis [11]. So using this image segmentation part our system gives a much more accurate result which capable of significant value for a farmer, and the farmer will identify easily which portion of papaya is infected and which diseases are mainly responsible for this. It can help them to grow super papaya, make changes in the social and economic revenue of our country.

IV. System Methodology

We have proposed some texture and color features in this study that have been extracted from the images where the images sent by mobile apps. Using the random forest classifier we classified the papaya images. Random forest classifier are using for classification and regression. It offers higher accuracy. If any of the diseases infect the classified images, then k-means clustering techniques used for the disease area segment.

Segmentation of images refers to the region of the area of inflecting from the images and it was carried to an image to find the region of the disease. Features such as color, texture, and shape are extracted from the images in this article [1].

We are choosing a random forest classifier to segment the infected area of papaya fruits. Random forest classifier easy to use and its accuracy is good and satisfactory. It can easily find out the problems and helps the farmer to cultivate good fruits.

In our dataset, we have five types of disease and each of those diseases recognizes by some diagnostic. Our five diseases are anthracnose, black spot, phytophthora, powdery mildew, and ring spot. When a papaya fruit inflected by the pathogen Colletotrichum gloeosporioides fungal then our system detects it as anthracnose. The black spot is detected by a fungus Aspersorium caricae and on the papaya fruit shows some black color spot. When papaya usually shows water-soaked lesions radiating smooth latex then it is recognized as a phytophthora and while a papaya fruit's surface is inflected by whitish power then the fruit is recognized as a powdery mildew disease. The last one ring spot is detected by round concentric rings causes up to 56-60 % yield misfortune into the papaya fruit [17]. When a papaya fruit full fills with those conditions then recognize the fruit other it's unable to detect the papaya fruit.

A. Random Forest

Random Forest Classifiers are a group learning method which are using for a regression of classification to the construction of a decision tree during the training period and t he mode and the prediction of a tree [10]. Random forest estimates the uncertainty of the prediction which can lead to standard deviations of all regression trees on p', l_m use for regression tree and m also use for regression, M is a free parameter which represent number of tree, depending upon the size and function of the training collection, a few hundred to several thousand trees are used. $\hat{l}$ means average predicting value. And α classify the value.

$$\alpha = \sqrt{\frac{\Sigma_{m=1}^{M}(l_m(p') - \hat{l})^2}{M-1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

B. K-Means Clustering

k-means clustering is a technique for quantifying the vector, originally from signal processing, intended to divide n observations into k clusters where the clusters are each of the nearest means (cluster centers or cluster centroid), which serves as the cluster's prototype. The distortion of a cluster with K-means depends on the data population and the distance between the objects and the cluster center [6]. Where p_j is the j cluster distortion, v_j is the j cluster center, m_j is the number of cluster j objects, y_{jq} is the q^{th} object of the j cluster, and $b(y_{jq}, v_j)$ is the differentiation from y_{jq} to v_j core of the j cluster.

$$p_j = \sum_{q=1}^{m_j} [b(y_{jq}, v_j)]^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (2)$$

C. SUPPORT VECTOR CLUSTERING (SVC)

If data are not labeled it will not be possible to learn supervised and an unsupervised approach will be required to find a natural grouping of the data and to map new data into these groups. The support vector clustering algorithm applies the support vector statistics developed in the algorithm for support vector machines to categorize unlabeled data and is one of the most commonly used clustering algorithms in industrial applications.

D. CONVOLUTIONAL NEURAL NETWORK (CNN)

A subfield of deep learning that deals with images on all scales are computer vision. It enables the computer, through an automatic process, to process and understand the content of a large number of images. Computer vision's primary architecture is the conventional neural network that is a derivative of feed forward neural networks. Its applications are very different, such as image classification, object detection, transfer of neural styles, and identification of the face [11]. A series of convolutional and pooling layers is a convolutional neural network that allows the main features to be extracted from the images that respond best to the final objective. Convolutional layer defend on convolutional operation between feature images. Here λth means the layer of λ, p_{ji} means Convolutional kernel, a_j means bias, N_i means input feature image.

$$y_i^{\lambda} = \sum_{j \varepsilon N_i} y_i^{\lambda-1} \times p_{ji}^{\lambda} + a_i^{\lambda} \quad \text{.........................(3)}$$

V. DATASET DESCRIPTION

For this research, we use five papaya disease to recognize our model. Those five diseases are very common in our country. Most of the farmer is facing this disease in their plants and this disease are Anthracnose, Blackspot, Phytophthora, Powdery mildew, Ring spot. Those five diseases contain 214 images where 60% images are used for training and 40% images are used as test data. We collect our dataset from Mendeley (https://data.mendeley.com/datasets/7dxg9n2t6w/1) which title is Papaya Fruit Diseases.

In our paper we work with some algorithm and their accuracy, compare all of the accuracies we chose a random forest classifier for segmentation of the images, and for classifying the image we use k- means clustering algorithm.

In our system, we take a picture using our handheld devices and send them to our system. Now the system extracts the color of the given image and extracts the features using feature extraction. After completing the extraction part using a random forest classifier if the image inflected then classified with the disease name and segment the image using k- means clustering and detect the inflected part and send it to the system. On the other hand, if the given fruit is inflected free then notify the system with normal fruit.

A. ANTHRACNOSE

Fig.2. Anthracnose

Papaya anthracnose is a genuine fungal infection caused by the pathogen Colletotrichum gloeosporioides. This disease is spread in rainy, splash black, humid periods, plant to plant by unsanitized tools. When the temperature between 18-25 C or 64-77 F then spread are common [17]. The fungus can infiltrate the plant tissue legitimately and symptoms may appear up to 5-7 days after contamination has occurred.

B. BLACK SPOT

Fig.3. Black spot

The reason for the black spot of papaya is affected by fungus Aspersorium caricae. This disease is generally serious during a rainy session. Leaves and fruits of papaya infected in black spot disease. Initial symptoms found in leaves upper side and it is small water-soaked lesions [1]. In disease progress, we saw small black sport undersides of leaves. When papaya leaves infected rapidly then leaves color goes to brown and die. When leaves die then tree growth and fruits are infected.

C. PHYTOPYTHORA

Fig.4. Phytopythora

It basically infected young fruit on plants and usually show water-soaked lesions radiating smooth latex. Roots may get dull and spoiled, causing hindering of plant development and yellow, fallen leaves. Seriously tainted plants may pass on. Plants that have a heavy load of fruits may fail [4]. Papaya plants with spoiled roots are susceptible to dry season pressure.

D. POWDERY MILDEW

Fig.5. Powdery mildew

In this disease under the surface of infection leaves are discovered patches of whitish powder growth. On upper surfaces, leaves at the contamination site show blotches of yellow or light green for the most part close to the vein, encompassed by ordinarily shaded tissue. Fungus attacks the young seed ingrown under light conditions [7]. The fungus grows seemingly on the undersurface of the leaves with drawing supplements from the cell of leaf surface by a particular engrossing structure known as haustoria.

E. RING SOPT

Fig.6. Ring spot

When this disease infected plant at first shows chlorosis on youngest leaves followed by vein clearing, rugosity, and unmistakable mottling of laminate. As the plants mature the rings and spots become bigger and get more obscure in shading, and as the natural product ages they can change in shading to yellow and brown colored. Affected papaya fruits show round concentric rings causes up to 56-60 % yield misfortune [17].

VI. DISEASE FEARUTE DESCRIPTION

A. DISEASE DESCRIPTION

This is a very important part of our research approach for this section we can analysis disease properly and understand this disease flaw and we gain more knowledge about the suitable features [22]. For this research, we use five papaya disease to recognize our model. Those five diseases are very common in our country Bangladesh. Most of the farmer is facing this disease in their plants and this disease are Anthracnose, Blackspot, Phytophthora, Powdery mildew, Ring spot.

B. FEATURES EXTRACTION

In view of the investigation of infections, a list of capabilities, which comprises of factual and gray-level co-occurrence matrix (GLCM) highlights, is chosen for perceiving the disease. Tarek Habib and Anup Majumder(2018) shown that Machine vision based papaya disease recognition we also take some disease to detect the papaya disease we have picked a few biased factual feature introduced in Tarek Habib and Anup Majumder(2018). These feature are examined here:

Mean (μ): If there are T pixels in defective locales, M pixels in shortcoming free locales, and gray scale shading force of a pixel in defective areas is spoken to by MS, at that point

$$\mu = \frac{\sum_{i=1}^{T} MS_i}{T} \quad \text{……………………} (4)$$

Standard deviation (σ): If there are T pixels in flawed locales, where MS and μ speak to gray scale shading power of a pixel furthermore, mean gray scale shading power of all pixels separately, at that point.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{T} (MS_i - \mu)^2}{T}} \quad \text{………………} (5)$$

Variance(σ^2): If there are T pixels in flawed locales, where MS also, μ speak to gray scale shading force of a pixel and mean gray scale shading force of all pixels individually, at that point.

$$\sigma^2 = \frac{\sum_{i=1}^{T} (MS_i - \mu)^2}{T} \quad \text{…………………..} (6)$$

Kurtosis (k): If there are T pixels in flawed areas, where MS and μ speak to gray scale shading power of a pixel and mean gray scale shading force of all pixels individually, at that point.

$$k = \frac{\frac{1}{T}\sum_{i=1}^{T}(MS_i - \mu)^4}{(\frac{1}{T}\sum_{i-1}^{T}(MS_i - \mu)^2)^2} - 3 \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

Skewness (γ): If the mean, mode, and standard deviation of gray scale shading force of all pixels in flawed districts are represented by separately μ, σ and M_0 at that point.

$$\gamma = \frac{\mu - M_0}{\sigma} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

Now adding some statistical feature, various GLCM features is utilized, which are first proposed by Haralick et al. (1973). These feature have been end up being a valuable technique for extricating textural feature from pictures, since they give a measure of the variety in force at the pixel of enthusiasm by considering the connection between two pixels one after another. The feature are examined here.

Let us accept that $g(c,d)$ be a two-dimensional advanced picture of size $M \times K$ pixels with number of gray levels L. We more accept that (c_1, d_1) and (c_2, d_2) are two pixels in $g(c,d)$, the separation is d and the edge between the two and the ordinate is h. At that point a GLCM P(i, j, d, h) becomes according to (Bicubic interpolation, xxxx):

$$p(i,j,d,\theta) = |\{(c_1,d_1),(c_2,d_2)\varepsilon \mathrm{M} \times \mathrm{K} : d, \theta, g(c_1,d_1)$$

$$= \mathrm{I}, g(c_2,d_2) = j\}| \ldots\ldots\ldots\ldots \quad (9)$$

In this work, five GLCM highlights have been utilized, to be specific contrast(C),correlation(ρ), energy(E), entropy (S) and homogeneity (H). Now describe this by equation.

$$\text{Contrast: } C = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2 \mathrm{P}(i,j). \ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

$$\text{Correlation: } \rho = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} i.j.p(i,j) - \mu_c.\mu_d}{\sigma_c.\sigma_d} \ldots\ldots \quad (11)$$

$$\text{Energy: } \mathrm{E} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\mathrm{P}(i,j) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

$$\text{Entropy: } S = -\sum_{i=0}^{L-1}\sum_{i=0}^{L-1}\mathrm{P}(i,j)\log\mathrm{P}(i,j) \ldots\ldots\ldots\ldots\ldots \quad (13)$$

$$\text{Homogeneity: } \mathrm{H} = \sum_{i=0}^{L-1}\sum_{i=0}^{L-1}\frac{\mathrm{P}(i,j)}{1+(i-j)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (14)$$

Here, μ_c, μ_d, σ_c and σ_d are total of expected and variance for the row and column matrix.

VII. RESULT ANALYSIS

Papaya is the most common fruit all over the world and people are really like this fruit. To increase the papaya fruits we can approach a system that can easily detect the diseases and infected area. It can help the farmer to earn more money and increase our countries GDP day by day. We are analyzing the dataset to predict the diseases easily. Here some of the result is shown:

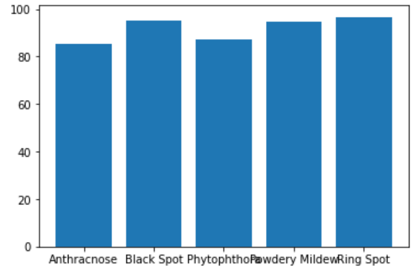

Fig.7. Bar chart of papaya diseases accuracy

In our dataset we have five diseases those are Anthracnose, Black Spot, Phytophthora, Powdery Mildew, Ring Spot and we have the accuracy for all the diseases. From all of the diseases, the Ring Spot disease shows the higher accuracy and the Anthracnose shows the lowest accuracy to our dataset.

TABLE.1. Multiclass confusion matrix

Actual Class	**Predicted Class**				
	Anthracnose	Black Spot	Phytophthora	Powdery Mildew	Ring Spot
Anthracnose	6.5	0.5	0	0	0.5
Black Spot	1	6.5	0	0	0
Phytophthora	0	0.5	6	0.5	0.5
Powdery Mildew	0	0	1.5	6	0
Ring Spot	0	0	0	0.5	7

Confusion matrix show their predicted class value for all the five diseases also shows the visualization of the performance of our dataset.

TABLE.2. Algorithms accuracy with sensitivity, specificity, precision, false positive rate, false negative rate

Algorithms	Accuracy	Sensitivity	Specificity	Precision	False positive rate	False negative rate
Random forest	96.3 %	84.12 %	97%	87.5 %	2.5%	12.4%
K-means clustering	93.5 %	83.87 %	92.2%	84.33%	3.2%	11.96%
SVC	95%	85.5%	97.15	85.6 %	2.88 %	14.4%

CNN	98.04 %	64%	89.76 %	79.4 9%	2.03 %	16.27%

In this table hatted with four algorithms with their accuracy, sensitivity, specification, precision, false-positive rate, and false-negative rate. All of those algorithms the random forest and CNN show the best accuracy while it burns into our IDE.

TABLE.3. Algorithms with accuracy

Algorithms	Accuracy
Random forest	96.3%
K-means clustering	93.5%
SVC	95%
CNN	98.04%

The main part of an algorithm is accuracy. While we run our algorithms it gives some accuracy where the random forest shows the accuracy is 96.3%, the k-means clustering algorithm shows 93.5%, Support Vector Machine (SVM) shows 95% accuracy, and CNN show 98.04% accuracy. From all of the accuracy, the CNN show the best for our dataset.

TABLE.4. Accuracy per disease

Algorithms	Papaya Anthracnose	Papaya Black spot	Papaya Phytophthora	Papaya Powdery mildew	Papaya Ring spot
Random Forest	88.97%	94.5%	87.37%	98.02%	97.32%
K- means clustering	84.7%	90.63%	89.98%	88.76%	91.33%
SVM	76.2%	98.31%	84.57%	93.34%	98.11%
CNN	87.31%	98.67%	88.34%	97.3%	96.53%

As our dataset contains five diseases so for all the diseases we try to show the accuracy using some algorithms (Random forest, K-means clustering, SVM and CNN).

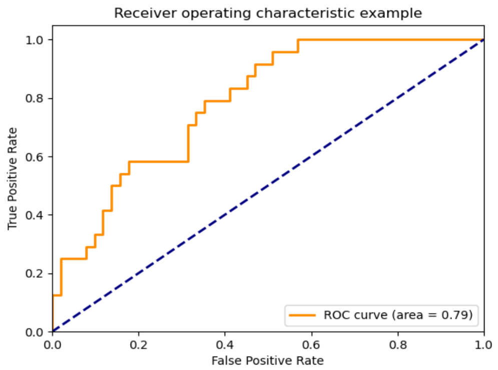

Fig.8. ROC curve of random forest

ROC curve used to predict the actual curve of the false positive rate and the false-negative rate and in this graph, we can see the threshold value is varied where the area curve value is 0.79.

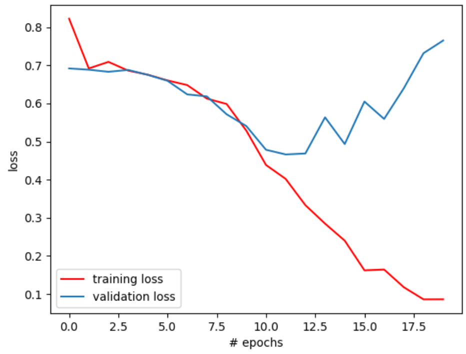

Fig.9. Loss and validation

After processing our model it generates a loss and validation graph for our algorithm in which loss is 0.1179, validation loss is 0.64, and validation accuracy is 0.7949.

VIII. Comparative Study

TABLE.5. Comparative study

Existing Model	Our Model
Most of the study classify with SVM, CNN, HOG, KNN and LR.	We are using K-means clustering for segmentation and then use Random Forest for Classifier, CNN, and SVC.
Existing model find out only accuracy.	We proposed an intelligent system which find out infected area and give feedback.
The majority of the previous study used only machine learning approach to find out some relation of papaya disease and how it occur.	In our study we use filed image to find out disease area to help the farmer.
Usually existing model only detect the disease.	Our approach detect the disease and classify the symptoms.
Most of the existing model accuracy 85-95%	Our accuracy is 98.04%

IX. Conclusion

In recent years, many scientists are performing to invent super papaya using genome sequencing, Dr. Maqsudul Alam one of them [14]. In this paper, we have presented a machine learning-based intelligent system which can detect papaya diseases. We are using random forest, k-means clustering, SVC, CNN and we are getting good accuracy in CNN (98.4%). It can help the farmer to find out the problem easily and take proper steps to reduce the diseases [22]. In the future, we are working on a large dataset to predict the factor which is mainly responsible for the papaya diseases.

REFERENCES

1. Veeraballi, R. K., Nagugari, M. S., Annavarapu, C. S. R., & Gownipuram, E. V. (2018, December). Deep Learning Based Approach for Classification and Detection of Papaya Leaf Diseases. In *International Conference on Intelligent Systems Design and Applications* (pp. 291-302). Springer, Cham.

2. Agriculture in Bangladesh, xxxx

Agriculture in Bangladesh, Available online: https://en.wikipedia.org/wiki/Agriculture_in_Bangladesh. [Last accessed on May 18, 2020]

3. List of countries by literacy rate, xxxx

List of countries by literacy rate, Available at https://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate [Last accessed on june 2, 2020].

4. Behera, S. K., Rath, A. K., & Sethy, P. K. (2020). Maturity Status Classification of Papaya Fruits based on Machine Learning and Transfer Learning Approach. *Information Processing in Agriculture*.

5. Asraf, H. M., Nooritawati, M. T., & Rizam, M. S. (2012). A comparative study in kernel-based support vector machine of oil palm leaves nutrient disease. *Procedia Engineering*, *41*, 1353-1359.

6. Islam, M., Dinh, A., Wahid, K., & Bhowmik, P. (2017, April). Detection of potato diseases using image segmentation and multiclass support vector machine. In *2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE)* (pp. 1-4). IEEE.

7. Ventura, J. A., Costa, H., & da Silva Tatagiba, J. (2004). Papaya diseases and integrated control. In *Diseases of Fruits and Vegetables: Volume II* (pp. 201-268). Springer, Dordrecht.

8. Din, N. M., & PAHANG, U. M. (2011). *Online Papaya Disease Diagnostic System (OPPDS)* (Doctoral dissertation, UMP).

9. Pothen, M. E., & Pai, M. L. (2020, March). Detection of Rice Leaf Diseases Using Image Processing. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 424-430). IEEE.

10. Samajpati, B. J., & Degadwala, S. D. (2016, April). Hybrid approach for apple fruit diseases detection and classification using random forest classifier. In *2016 International Conference on Communication and Signal Processing (ICCSP)* (pp. 1015-1019). IEEE.

11. Nadeem, A., Mehmood, T., Tahir, M., Khalid, S., & Xiong, Z. (1997). First report of papaya leaf curl disease in Pakistan. *Plant Disease*, *81*(11), 1333-1333.

12. Porter, B. W., Christopher, D. A., & Zhu, Y. J. (2014). Genomics of Papaya Disease Resistance. In *Genetics and Genomics of Papaya* (pp. 277-307). Springer, New York, NY.

13. Hewajulige, I. G. N., Sultanbawa, Y., Wijeratnam, R. S. W., & Wijesundara, R. L. (2009). Mode of action of chitosan coating on anthracnose disease control in papaya. *Phytoparasitica*, *37*(5), 437-444.

14. Ming, R., Hou, S., Feng, Y. *et al.* The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452,** 991–996 (2008). https://doi.org/10.1038/nature06856

15. Camargo, A., & Smith, J. S. (2009). Image pattern classification for the identification of disease causing agents in plants. *Computers and Electronics in Agriculture*, *66*(2), 121-125.

16. Yu, Q., Tong, E., Skelton, R.L. *et al.* A physical map of the papaya genome with integrated genetic map and genome sequence. *BMC Genomics* **10,** 371 (2009). https://doi.org/10.1186/1471-2164-10-371

17. Tarek Habib, Anup Majumdar, A.Z.M. Zakaria, Morium Akter, Mohammad Shorif Uddin, Farruq Ahmed. “Machme vision Based Papaya Diseases Recognition” Journal of King Saud University- Computer and information Sciences. Volume 32, issue 3, March 2020.

18. Vijayakumar, T., and Mr R. Vinothkanna. "Mellowness Detection of Dragon Fruit Using Deep Learning Strategy." Journal of Innovative Image Processing (JIIP) 2, no. 01 (2020): 35-43.

19. Chen, Joy Iong Zong, and Lu-Tsou Yeh. "Analysis of the Impact of Mechanical Deformation on Strawberries Harvested from the Farm." Journal: Journal of ISMAC September 2020, no. 3 (2020): 166-172.

20. Kelly, K., Madden, R., Emond, J. P., & do Nascimento Nunes, M. C. (2019). A novel approach to determine the impact level of each step along the supply chain on strawberry quality. Postharvest Biology and Technology, 147, 78-88.

21. Liang, Z., Xu, L., De Baerdemaeker, J., Li, Y., & Saeys, W. (2020). Optimisation of a multi-duct cleaning device for rice combine harvesters utilising CFD and experiments. Biosystems Engineering, 190, 25-40.

22. Chandy, A. (2019). RGBD Analysis for Finding the Different Stages of Maturity of Fruits in Farming. Journal of Innovative Image Processing (JIIP), 1(02), 111-121

23. Manoharan, Samuel. "Study On Hermitian Graph Wavelets in Feature Detection." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 24-32