❒ 917

Fruits Disease Classification using Machine Learning Techniques

Yassine Benlachmi[1], Aymane El Airej[2], Moulay Lahcen Hasnaoui[3]
[1,2,3]ENSAM Moulay-Ismail University, LMMI Laboratory, Institute, Meknes, 50000, Morocco

Article Info	ABSTRACT
	Due to increased population, there is a high demand for agricultural products these days and therefore, effective growth and increased fruit production have become critical. Consequently, for better fruit yield cultivators employ traditional methods for monitoring fruit yield from harvest till ripening of fruit. However, manual monitoring and visual inspection doesn't always bring the actual identification of fruit disease due to variety of reasons, such as less knowledge about pathogens, requiring more time for disease analysis and that too with less accuracy and so on, consequently, leaving for the need of a professional assistance and expertise. Moreover, the task also becomes difficult as various fruits demonstrate their gesticulation by changing the colour of their skin which can come from nature and resulting in various black or dark brown spots on the fruit skin indicating various diseases. As a result, it is necessary to propose an efficient smart farming strategy that will aid in increased productivity while at the same time involving less human effort. The proposed research work attempts to classify the fruit disease at its early stage by using machine learning techniques. For this purpose, fruit's texture, and skin colour have been utilized. The approach fundamentally employs three machine learning classifier algorithms - KNN, Decision Tree, and Random Forest. Whereas the features have been determined by using three prominent feature extractors - Haralick, Hu Moments and colour histogram. Finally, the system has been evaluated by utilizing the k-fold cross validation method. Specimen dataset was divided into two groups — the training subset and the test subset. As a rule, four-fold cross-validation, three-fourths of the images were used for training the models whereas, the remaining one-fourth were used for testing purposes. Assessment results for suggested methodology after conducting experimentation on publicly available dataset and drawn confusion matrix and learning cure shows that Random Forest classifiers achieves accuracy about 99% while for K-Means accuracy statistics stands at 98.67% and for Decision trees it is about 97.75% - for colour histogram features.

Corresponding Author:

Yassine BENLACHMI
ENSAM Moulay-Ismail University, LMMI Laboratory, Institute
Meknes, 50000, Morocco
Email: yassin040@gmail.com

1. INTRODUCTION

Fruit diseases can result in significant losses in productivity and quality when they manifest themselves during the harvesting process. Therefore, premature fruit disease identification can help in the lessening of such losses as well as can help in the prevention of the disease's subsequent spread to the other parts of tree or other fruits as well. The traditional approach for identification and detection of fruit diseases is formed on professional observation with the visual inspection to determine the presence or absence of

disease. The physical recognition of the fruit infection is completed by watching and detecting pathogens, which takes more time and is more expensive than other methods and has lesser accuracy [1]. Adding to that every disease that affects fruits causes a distinct surface or a specific shaded area to appear which makes the task more complicated. Therefore, fruit disease identification necessitates a tremendous amount of effort and knowledge of the fruit disease itself. Apple rot, apple blotch, and apple scab are just a few diseases that affect apple trees and fruits. Moreover, because of the large distance between specialists' locations in a few emerging nations, consulting with them can be a time-consuming and expensive endeavour in some of these countries.

Upon the detection of fruit diseases in the preliminary phases, it is possible to control them by correct management measures such as fungicide-based vector control, disease-specific chemical and pesticide applications and other measures. However, excess use of chemical in some cases could lead to groundwater contamination [2]. Currently, some methods exist to identify oligonucleotides or pathogen-specific antigens. These methods include immunoassays, RNA (ribonucleic acid) sequencing as well as PCR (polymerase chain reaction). However, these methods are costly, timewasting, and call for expert human resources [3]. Furthermore, because they are carried out on specimen samples, they can't ensure the standard of the full batch, therefore there are chances of inaccurate quality assessment.

To overcome this deficiency, researchers in the contemporary times have developed image processing and computer vision methods to automate visual evaluation of fruits for color, texture, size and for various other diseases during pre to post harvesting of fruit crops. Electronic perception of an image makes interpretation and character recognition in image possible which in turn allows image processing and computer vision to mimic the effect of human visual inspection in assessing the grade of vegetables and fruits [4]. However, because of different skin texture and color of different fruits, variance of fault types, and presence of stems and calyxes, utilizing images to find fruit flaws continues to be a challenge. Additionally, most of agriculture's information and data are primarily derived from difital images, however estimating, or processing photographic data is difficult analytically [4]. Digital images on the other hand overcome these deficiencies of photographs.

The objective of this research is to make it simple to identify the diseases of fruits via the combination of various machine learning approaches, and by doing so, we can prevent farmers from experiencing a financial setback. Undertaken work focuses on the experimental evaluation of images used to detect and classify fruit disease. The remainder of the research is organized as follows: Section 2 contains a review of the literature. Section 3 describes the various types of fruit disease. Section 4 illustrates Dataset; the proposed methodology is presented in Section 5. Section 6 describes the Experimental environment. Section 7 describes the Evaluation Measures, Section 8 illustrated Results and Section while Section 9 concludes the paper.

2.   LITERATURE REVIEW

Traditionally, fruit disease detection is done by visual inspection however, this trend is changing in contemporary world as detection is being automated now a days. In this section we will discuss prior research work conducted by numerous researchers in the fields of fruit and crop disease detection and image classification. To diagnose fruit illnesses, clustering and fruit image segmentation algorithms have recently been gaining momentum [5]. The former approach divides pixels into multiple classes by employing various classification algorithms [6].

One of the most famous deep learning approaches is the use of ANN (artificial neural networks) for detecting disease within images. In this connection, system at [7] makes use of the artificial neural network concept. Two apple and three grape diseases were selected for the research and subsequently, two image datasets were employed, one for training on disease images while, the other for testing purposes. Additionally, a radial basis function was employed to aid in the composition of training set images. Three carefully chosen criteria were employed to categories and link the images to the appropriate disease groups. These parameters include color, texture, and morphology. The proposed system achieves 90% accuracy.

Similarly, for improving the pixel quality, a concept of intent search has been chalked out [8]. Images of leaves are acquired from agricultural fields using digital cameras, then they are preprocessed for noise removal and image enhancement. Afterword's, K-means clustering approach is employed for color-based image segmentation, which can detect disease-affected leaves. The segmented image features are collected via GLCM (Gray-Level Co-Occurrence Matrix) and then SVM classifier is applied on extracted features for classification of leaf diseases.

Adding to that authors at [9] suggest an alternative technique to molecular or serological tests for orange disease detection. This study shows that from around 62 plants their leaves were collected and put under examination for viral disease infection by utilizing polymerase chain reaction (PCR) test. For

hypothesis testing, a commercialized E-Nose technology, Alpha MOS Fox 3000, was employed, gaining a precision of about 95.30% utilizing only Random Forest algorithm.

In one of the study [10] MIB classifier has been used. The suggested classifier is intended to distinguish between normal and diseased citrus fruits as well as leaves which manifest diseases such as greening, scab, canker, melanoses, and blackspots among others; most of them are common citrus fruit diseases. To remove the layers from various images, this model uses gaussian feature extraction method. In addition, a pre-trained deep learning model was tested against a variety of citrus fruits and leaves datasets. Conclusion shows that the MIB model outperforms opponents on several computational data counts. The model achieves 98% accuracy.

Moreover, a relatively new concept of disease detection is instance segmentation [11] that is employed to distinguish distinct items and identify each pixel that is a part of an instance of an object. This method employs stills of several fruit diseases, including bitter rot, sooty blotch, and powdery mildew. To identify impacted sections of the apple fruit, instance segmentation is conducted using the CNN algorithm. The results are better than RNN approach as well.

Furthermore, researchers at [12] propose a single stream CNN (convolutional neural network) approach for analyzing diseases in citrus fruits. Four contrast enhancement processes are used to perform data augmentation in the first step: shadow was removed, pixel intensity was adjusted, brightness was improved, and local contrast was also enhanced. In the second stage, the MobileNet-V2 CNN model is chosen as well as fine-tuned. Enlarged citrus dataset was used for training the fine-tuned model via transfer learning method. Deep feature extraction is performed using the freshly trained model; nevertheless, examination reveals that there is some presence of redundant information in recovered deep features. Consequently, in the third stage, a corrected Whale Optimization Algorithm (IWOA) is applied. Optimal features are shortlisted in the final stage via ML technique. The upgraded leaves, citrus fruits, and hybrid datasets were employed in trial, and their accuracy was 99.4, 99.5, and 99.7%, respectively.

Another slightly different though relevant work is at [13]. By employing a few ML (Machine learning) algorithms, including support vector machines (SVM), logistic regression, decision trees, and naive bayes, etc., the visible markings on rice plants are processed. Thereafter, plants are then divided into normal and diseased based on classification results. Numerous fields in Kashmore Pakistan were visited to collect a dataset of about 1000 rice seed images and only the rice seed is subjected to the full analysis of acquisition, pre-processing, and feature extraction. With the aid of a knowledgeable expert in plant diseases dataset is labelled for specimen that are normal and diseased. Testing accuracy and F1-score are utilized to assess data processing techniques. Results from conventional classifiers are included in this paper, and transfer learning has also been utilized to correlate the outcomes. The output of deep learning networks and conventional classifiers are compared in the final analysis.

Similarly, [14] employs region-based convolutional neural network (Faster R-CNN) for real-time detection of disease affecting rice leaves. Faster R-CNN uses improved RPN structure to develop candidate regions. This scheme addresses an object's location in a very accurate and precise manner. It was established that this deep learning model was very precise in mechanized diagnosis of 3 selective diseases of rice leaves - Hispa, brown spot, and blast -
with 98.09%, 98.85%, and 99.17% precision results, respectively. Furthermore, while in the identification of healthy rice leaves model yielded 99.25% accuracy.

There are some other notable works also detecting disease within the fruit and crop images. For instance, Naïve Bayes Classifiers have been used for detecting disease within apple fruit [15], an improved Densenet Fusion Defogging Algorithm have also been used to detect for fruit disease identification [16] and so on. According to the findings in literature review, plant nomenclature and fruit family classification pose a significant problem because of their varied intrinsically complicated feature set. This study uses color histogram, Hu moments, and Haralick texture as feature extraction and classification using three different classifiers.

3. CLASSIFICATION OF VARIOUS FRUIT DISEASES
Some of very common apple fruit disease has been classified as under:

3.1. Apple Scab
This apple illness is the most lethal of all infections because it causes the most harm. Moreover, it is also quite common in apple-growing areas. During the blooming season, apple scab is more severe in cool and rainy environments, although it is not as severe in dry or warm climates. Apple scabs can be spotted on an apple tree's leaves, petals, blooms, new shoots, and bud scales. The most usually diseased and visible sections of the body are the fruits and leaves. Figure 1 depicts an image of an Apple scab.

Figure 1. A sample image of Apple Scab disease [1]

3.2. Bitter Rot

Botryosphaeriaobtusa is a fungus that causes bitter rot in fruit. The key components influencing the apple tree are its leaves, bark, and fruits. After one to three weeks, first signs of apple rot appear on the outer surface of the leaves, followed by a little purple blotch on the petiole, which eventually turns dark tan and yellowish-brown in the center. A few weeks later, the fruit begins to show signs of the second stage of apple rot. When the leaf spots reach this stage, they start to grow back to their previous size.

Figure 2. Specimen image of Apple bitter rot disease [1]

3.3. Black Rot

This issue is acquired from an infection that spreads on many woody plants and trees, and the bacteria that causes it are carried by wind. Many large brown scars on the leaves that begin circular and expand into patches as they grow together. The spots are dry and paper-like. Spots on fruit can emerge at any time of year, and affected fruit rots and hangs on the tree. Cankers on branches are also caused by the fungus.

Figure 3. Specimen image of Apple black rot disease [1]

3.4. Apple Blotch

In the northwest regions, apple blotch is the most frequent "summer disease" of apples. It is caused by two microbes. An apple blotch is characterized by dark greenish-blue patches on the surface of diseased fruit. Individual colonies range from one to many almost circular colonies. These symptoms appear three to four weeks after the leaflet falls.

Figure 4. Specimen image of Apple blotch disease [1]

3.5. Aspergillus Rot

Aspergillus rot or Alternaria fruit rot occurs when fruit's internal parts are affected with infection specially in rainy seasons. Some of the indicators of aspergillus rot include a slight variation in skin color and a loss of weight owing to internal deterioration. However, this illness does not normally manifest until after harvesting or during fruit sorting. Fungus may be growing inside the fruit without any visible indications. Infected fruit usually has little yellowish to brownish-red stains and is off-color, such as a pale red.

Figure 5. Specimen image of Aspergillus rot [17]

3.6. Gray Mold

Botrytis cinetea or gray mold infestation damages the flower section of the fruit and impairs the yield till it ripens. When a gray mold infection occurs, the flower component of the fruit is damaged, and the crop is affected until it is ripe. When the fruit is cleaned or stored in high humidity, condensation or water on the blossom tissues stimulates the growth of the fungal mycelium, which allows the fruit to ripen more quickly. The characteristic grey coating of spores and pathogen sporulats is developed on the flower parts. The crown tissue will become colonized when the fungus has sufficiently expanded throughout the fruit tissue. We must store defective fruit in a high-moisture environment.

Figure 6. Specimen image of Gray Mold [1]

4.   RESEARCH METHOD

A summary in the form of schematic diagram of the research can be viewed in figure 7. The underlying object of this study is to chalk out a robust technique that could help in identifying various diseases among various fruits such as banana, orange, and apple. Specimen images for both health and diseased fruit have been chosen from the dataset. The input RGB image is preprocessed for size adjustment, noise removal and other data cleaning functions such as making all images homogenous in size, that is, reducing them in 227 * 227 pixels as well as adjusting color space. After pre-processing features have been collected from the ROI (Region of interest) via haralick texture, Hu moment, and color histogram techniques. The computed features vector was further used for classification purposes. K-fold cross validation has been used to assess the performance of the system. The data set was divided into two parts - training and testing. Afterword's, 4-fold cross validation was employed where three-fourths of the images were utilized for training while one-fourth of the images were used for testing purposes.
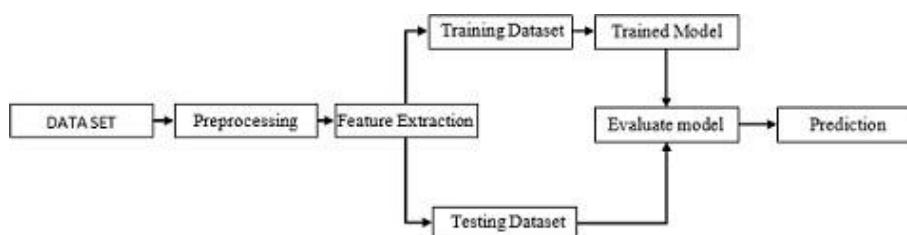

Figure 7. An overview of Proposed Methodology

4.1. Dataset Acquisition

To evaluate the proposed methodology a publicly available data set [18] has been acquired experimentation has been extended on that data set. The said data set includes three types of fruit disease images of apple, banana, and orange. The disease image data has been labelled as healthy and rotten. Moreover, there are a total of 13587 images of fruits considered in this dataset. Additionally, input images from digital cameras have also been considered for experimentation purposes.

4.2. Data Cleaning
4.2.1. Preprocessing

For the training of the chosen machine learning classifiers input images have been initially preprocessed from the given dataset [18]. Preprocessing is one of the fundamental steps in any of the machine learning techniques as it smoothens the image and is instrumental in image noise removal and thereby contributes greatly towards achieving high accuracy results. The preprocessing includes image binarization, skew detection, noise removal, grayscale conversion, and size normalization.

In the proposed research initially, images have been resized in a uniform shape along with increasing contrast for improving quality as well as uniformly illuminating image. Subsequently, and RGB image was converted into BGR and finally transformed into HSV. BGR and HSV conversion has been illustrated in Figure 8.
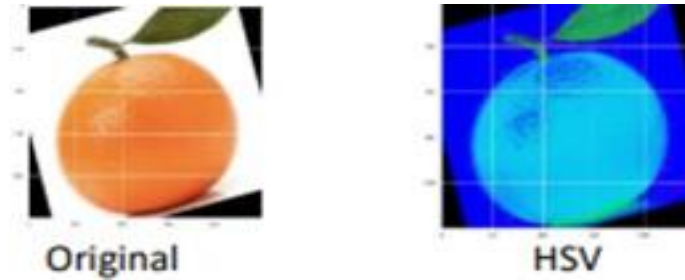
Figure 8. BGR and HSV conversion

4.2.2. Image Segmentation

Segmentation of image followed the preprocessing. It is a technique for splitting an image into sections that are more representative and easier to examine than the individual parts of the image. Segmentation has primarily two approaches one is region based and the other is boundary based. In the former discontinuities are identified and linked to boundaries surrounding areas whereas in the latter commonalities are identified among various groups of pixels.

Here, we divide the image based on color using a technique called color-based segmentation. Subsequently, on this gray scale, black and white conversion is applied. A sample image of image segmentation is presented in figure 9.
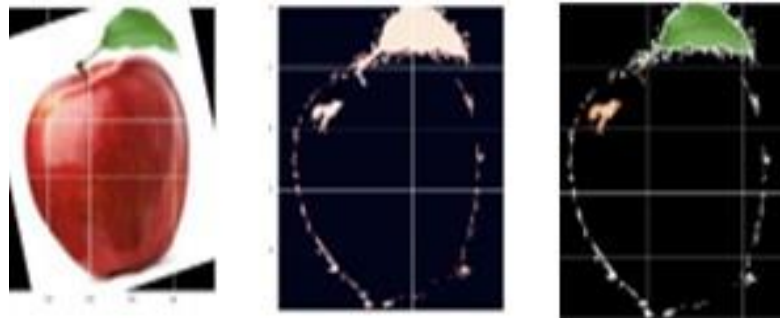

Figure 9. Image segmentation with various flavors

4.3. Feature Extration

Once the fruit image is preprocessed then the features are extracted for classification purposes. These features are supposed to be non-redundant and full of information, thereby easing the following learning and generalization processes and improving human interpretations in some circumstances. Dimensionality reduction is also one of the processes of features extraction as well as huge data from images is sometimes becomes too unwieldy and unable to be processed in an optimized manner.

The proposed study employs global feature descriptor (GFD) to extract an image's color, shape, and texture information. GFD describes the image to generalize the complete item of a particular object. Contour representation, shape descriptor, and texture characteristics are examples of global features. Additionally, Hu-Moments, Haralick, and color histogram have also been exploited for feature extraction. Finally, the computed feature vector was further fed to the undertaken classifier to classify fruit disease.

4.4. Classification Models

Classification is recognizing, comprehending, and categorizing pixels of an image into predetermined groups, often known as subpopulation of pixels [19]. It allows ML methods to take into consideration and apply on pre-categorized training datasets classification so to categorize future data into the most pertinent and applicable classifications. For classification, we have employed three different ML algorithms, i.e., KNN, Random Forest, and Decision Tree. The detail of each classifier is mentioned in a subsequent section.

4.4.1. K-Nearest Neighbor Classification

K-Nearest Neighbor algorithm or KNN, is a simple and straightforward ML algorithm and it falls into the supervised learning technique category. This ML algorithm takes into account similarity factors among various existing and new cases. When it encounters a new case KNN places it in the existing category showing most resemblance to the new case. This approach can be used for classification and regression issues, but it is most frequently used for classification problems. KNN is non-parametric ML algorithm that means it makes no assumption about the data it is given as a starting point. We employed this algorithm for

the classification of fruit disease, where we optimized this algorithm with different numbers of neighbors and realized very effective performance. Figure 10 advocates KNN working principle.
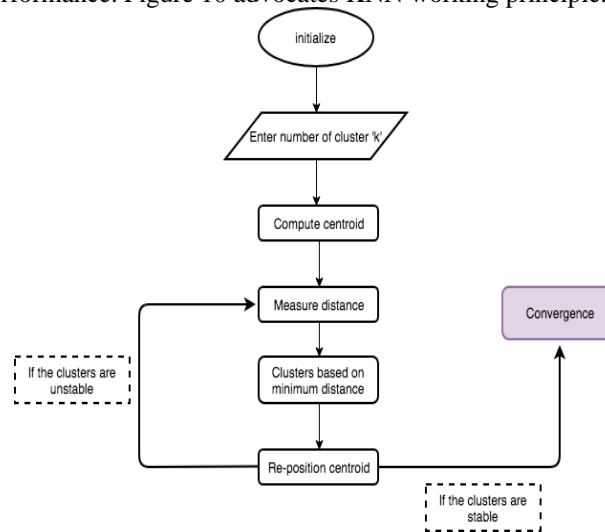

Figure 10. KNN working [20]

4.4.2. Decision Trees

A non-parametric supervised learning technique that can be utilized for regression and classification tasks is the decision tree. To achieve this, it is necessary to develop a model that predicts the value of the target variable by learning straightforward decision rules that can be inferred from the properties of the data. Whereas Tree can be thought of as an approximation to a piecewise constant. Decision trees are simple to understand and interpret and don't necessarily call for gathering a lot of data. Whereas Data normalization, the introduction of dummy variables, and the elimination of blank values from the data are typically required by other methods. It makes use of a white- box model where a particular scenario can be observed in a model, and the explanation for the condition can be easily described using Boolean logic. It performs admirably even when some of its assumptions are partially broken by the underlying model from which the data was derived [21]. A general working diagram of decision tree is presented in figure 11.
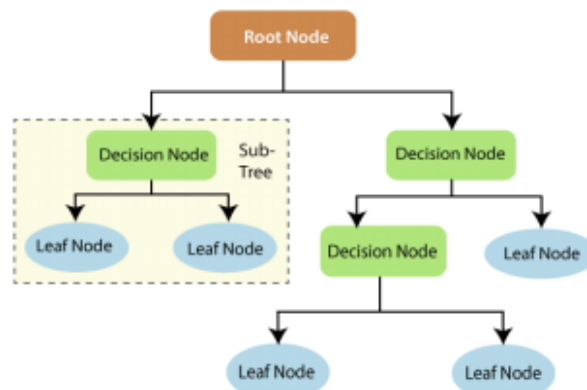

Figure 11. General working diagram of Decision tree [22]

4.4.3. Random Forests

In classification and regression problems, Random Forest is a supervised machine learning algorithm that is commonly utilized. Using different samples, it creates decision trees and uses the majority vote for classification and the average vote for regression, depending upon the situation. It can deal with a dataset that contains both continuous and categorical variables. It produces superior results when dealing with classification challenges. Using sample training data with replacement generates a distinct training subset, and the final output is determined by a simple majority vote. To turn weak learners into strong learners, it builds sequential models in such a way that the final model has the best accuracy possible [23]. A general working diagram of Adaptive Random forests (ARF) is presented in figure 12.
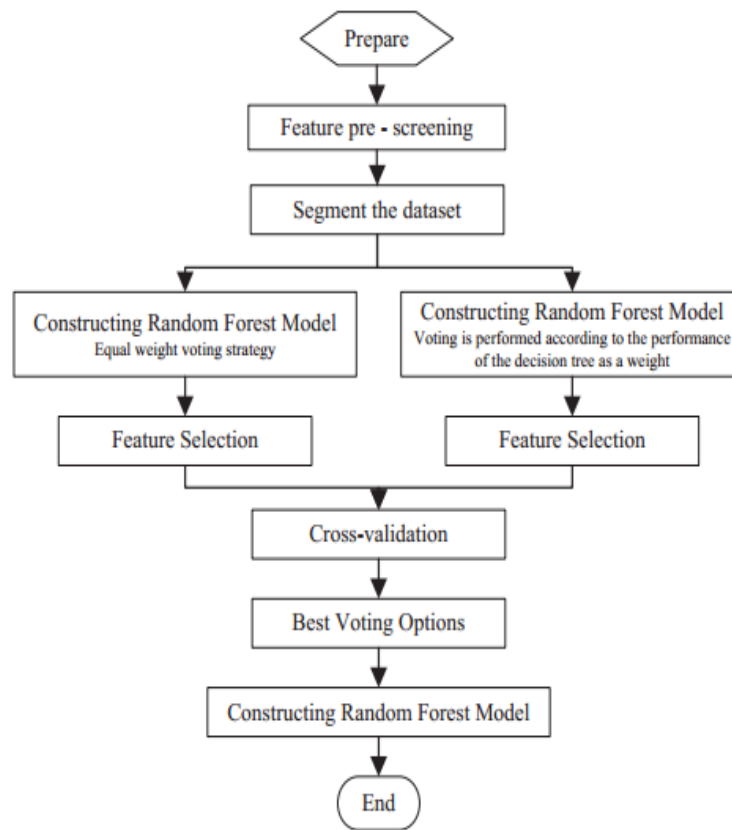
Figure 13. Flow chart of ARF algorithm [24]

4.5. Experimental Settings
4.5.1. Hardware Settings
The research was carried out using a computer machine with 8GB of RAM, 1 TB hard drive, and a GPU with 11G of processing power.

4.5.2. Software Settings
To design and test the models, the Python programming language version 3.10.4 was configured. In addition, following machine learning libraries were utilized:

Scikit-learn: It is a machine learning package that is widely used and incredibly powerful. This library provides a collection of quick tools for machine learning and statistical modeling, including classification, clustering, and data preprocessing. We used this library for classification models

NumPy: In python, NumPy is a numerical programming language that allows for the computation and manipulation of elements in multi-dimensional and one-dimensional arrays. We used this library for data preprocessing.

Matplotlib: It is a Python and its extension NumPy-based cross platform data visualization and graphical charting tool that can be used on a range of different operating systems. Matplotlib can be used to integrate graphs into graphical user interfaces and other applications. We used this library to create charts and graphs.

5.   RESULTS AND DISCUSSION
In this section, we present the performance of each classifier for various features extractors employed. Table 1 shows the different fruit classes that have been considered for experimentation

5.1. K-fold cross validation
The k-fold cross validation method is used to evaluate the suggested methodology. Specimen dataset was divided into two groups — the training subset and the test subset. As a rule, four-fold cross-validation, three-fourths of the images were used for training the models whereas, the remaining one-fourth were used for testing purposes.

Table 1. Various fruit classes

S.no	Types of fruits	No. of fresh fruits	No. of infected fruits
1	Apple	24	24
2	Infected apple	24	24
3	Banana	24	24
4	Infected banana	24	24
5	Orange	24	24
6	Infected orange	24	24

5.2. Models' performance via Hu Moments features

We considered the Hu-moment feature of all fruit images and prepared one feature set that was fed to all different classifiers for the classification purpose. In hindsight, it was evident that it would never work because we couldn't distinguish whether a certain fruit was good or bad regardless of the physical shape. The model predicted 73% accurate responses in the case of rotten banana, while the model did not predict any correct responses in the rotten orange according to the confusion matrix. Classifiers based on a Decision Tree was one that scored the highest in terms of accuracy. The confusion matrix (figure 13), accuracy table (Table 2) and learning curve (figure 14) is illustrated in subsequent figure and table.

Table 2. Results of Model Classification with Hu-Moments features

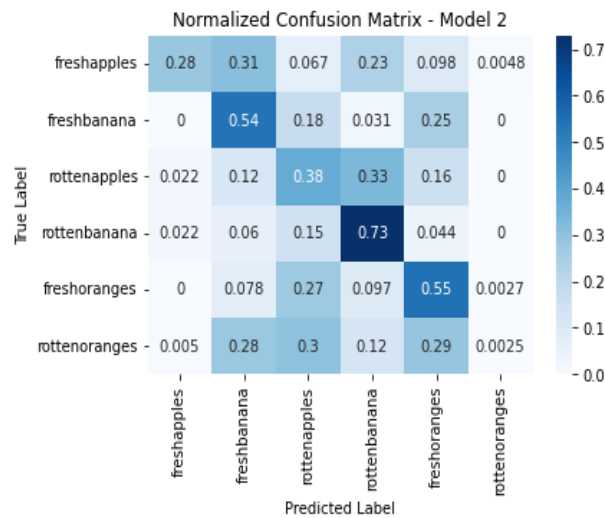| Model	Overall Accuracy
KNN	38.15%
Random Forests	42%
Decision Trees	48.74%

Figure 13. Confusion Matrix for classification through Hu- Moments features (Random Forests)
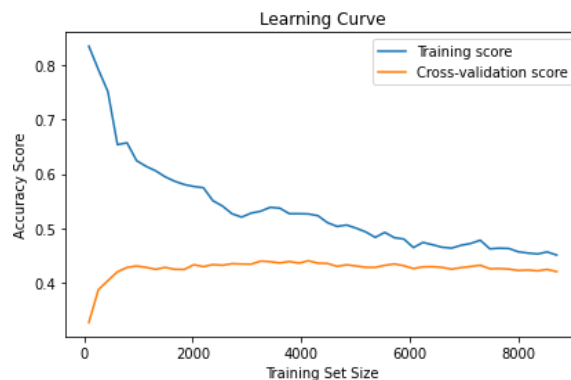

Figure 14. Learning Curve for Model classification through Hu- Moment features (Random Forests)

5.3. Models' performance via Haralick features

Random Forest Classifier performed well on Haralick features achieving 92.85% classification accuracy. The model was able to distinguish between fresh and rotten bananas with an accuracy of 96%, followed by oranges with a 94% percent accuracy, and finally apples with 92% accuracy. Complete statistics have been presented in table 3. Whereas confusion matrix and learning curve for random forests model on Haralick features have been resented in figure 15 and figure 16 respectively.

Table 3. Results of Model Classification with Haralick features

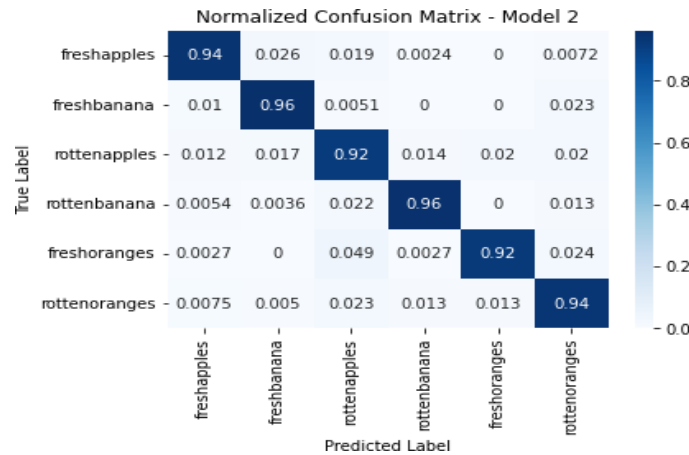| Model	Overall Accuracy
KNN	57.62%
Random Forests	92.85%
Decision Trees	85.33%

Figure 15. Confusion Matrix for classification through Haralick features (Random Forests)
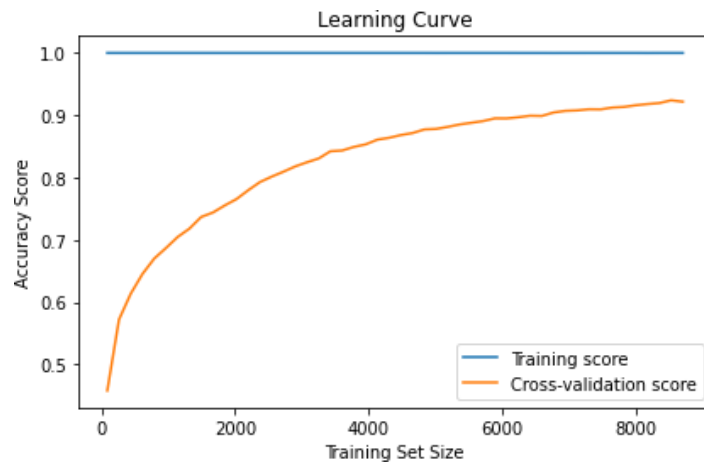

Figure 16. Learning Curve for Model classification through Haralick features (Random Forests)

5.4. Models' performance via color histogram features

In the third analysis, we employed color histogram as features set and integrated with the classifiers. The complete findings are presented in table 4. Whereas confusion matrix and learning curve for random forests model on color histogram features have been resented in figure 17 and figure 18 respectively.

Table 4. Results of Model Classification with color histogram features

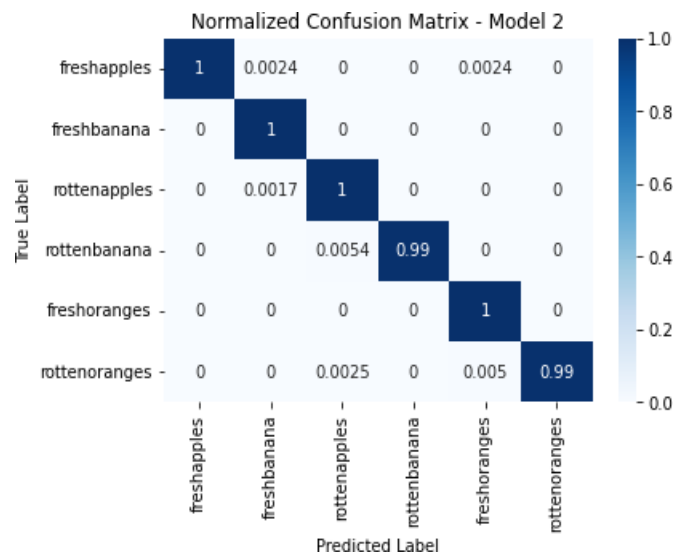| Model	Overall Accuracy
KNN	98.67%
Random Forests	99.54%
Decision Trees	97.75%

Figure 17. Confusion Matrix for classification through color histogram features (Random Forests)
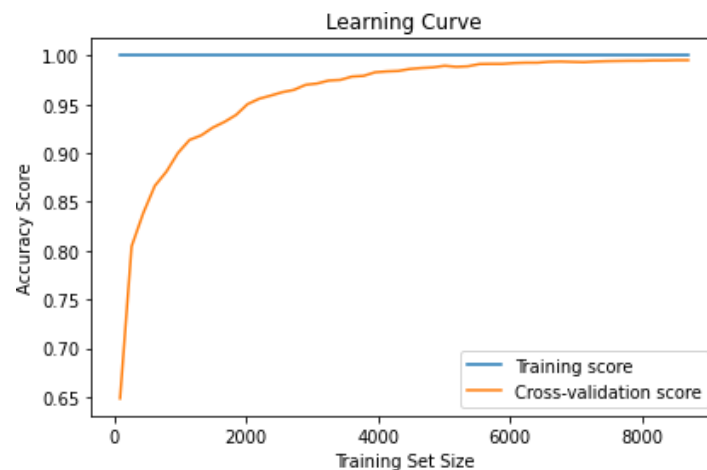

Figure 18. Learning Curve for Model classification through color histogram features (Random Forests)

6. FUTURE WORK AND CONCLUSION

It has been demonstrated in practise that the features extracted from processed images using colour histogram and other techniques have the potential to be used not only for quantitatively categorizing and differentiating between various fruit disease but could also be extended to predict other crop disease as well as could be used in medical imagery.

Overall, this paper attempts to decipher fruit disease in a more effective and meaningful way via machine learning algorithms. Machine learning provides an effective means for detecting fruit disease in a timely and effective manner and thereby providing an edge over traditional detection techniques of fruit disease which could be time and resource consuming. A publicly available data set along with other auxiliary resources for fruit disease images have been analyzed, preprocessed for noise removal and deblurring the image. The preprocessed image is further analyzed for feature extraction through various feature extracting techniques such as color histogram, Haralick, and hu moments. For the predicting and classification purposes three well known machine algorithms have been put into practice i.e., KNN, Decision tree, and random forest. The statistical analysis of the results produced by the various models suggest that random forest achieves the better accuracy of 99% in classification when it has been fed with the features extracted via color histograms.

REFERENCES

[1]   H. Modi, M. Patel, M. Patel, and H. Patel, "Implementation of Algorithm to Detect the Diseases in Fruit Using Image Processing Technique," *International Journal of Applied Engineering Research,* vol. 14, no. 9, pp. 2093-2106, 2019.

[2] K. Ślósarczyk and A. J. Witkowski, "Preliminary Evaluation of the Possible Occurrence of Pesticides in Groundwater Contaminated with Nitrates—A Case Study from Southern Poland," *Water,* vol. 13, no. 21, p. 3091, 2021.

[3] A. Fazari *et al.*, "Application of deep convolutional neural networks for the detection of anthracnose in olives using VIS/NIR hyperspectral images," *Computers and Electronics in Agriculture,* vol. 187, p. 106252, 2021.

[4] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," *Journal of King Saud University-Computer and Information Sciences,* vol. 33, no. 3, pp. 243-257, 2021.

[5] P. K. Devi, "Image Segmentation K-Means Clustering Algorithm for Fruit Disease Detection Image Processing," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020: IEEE, pp. 861-865.

[6] G. Chao, S. Sun, and J. Bi, "A survey on multiview clustering," *IEEE transactions on artificial intelligence,* vol. 2, no. 2, pp. 146-168, 2021.

[7] Z. B. Husin and A. H. B. A. Aziz, "Ali Yeon Bin Md Shakaff Rohani Binti S Mohamed Farook,"Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques"," in *2012 Third International Conference on Intelligent Systems Modelling and Simulation*.

[8] M. K. Kaleem, "A Modern Approach for Detection of Leaf Diseases Using Image Processing and ML Based SVM Classifier," *Turkish Journal of Computer and Mathematics Education (TURCOMAT),* vol. 12, no. 13, pp. 3340-3347, 2021.

[9] S. Hazarika, R. Choudhury, B. Montazer, S. Medhi, M. P. Goswami, and U. Sarma, "Detection of citrus tristeza virus in mandarin orange using a custom-developed electronic nose system," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 11, pp. 9010-9018, 2020.

[10] G. M. Brindha, K. Karishma, J. Nivetha, and B. Vidhya, "Automatic Detection of Citrus Fruit Diseases Using MIB Classifier," in *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2022: IEEE, pp. 1111-1116.

[11] S. Mohanapriya, V. Efshiba, and P. Natesan, "Identification of Fruit Disease Using Instance Segmentation," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021: IEEE, pp. 1779-1787.

[12] M. Hassam *et al.*, "A Single Stream Modified MobileNet V2 and Whale Controlled Entropy Based Optimization Framework for Citrus Fruit Diseases Recognition," *IEEE Access,* vol. 10, pp. 91828-91839, 2022.

[13] R. Kumar, G. Baloch, A. B. Buriro, and J. Bhatti, "Fungal blast disease detection in rice seed using machine learning," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 2, 2021.

[14] B. S. Bari *et al.*, "A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework," *PeerJ Computer Science,* vol. 7, p. e432, 2021.

[15] Y. Sugiarti, A. Supriyatna, I. Carolina, R. Amin, and A. Yani, "Model Naïve Bayes Classifiers For Detection Apple Diseases," in *2021 9th International Conference on Cyber and IT Service Management (CITSM)*, 2021: IEEE, pp. 1-4.

[16] Y. Qi, C. Zhichao, and L. Lan, "Fruit Disease Identification Based on Improved Densenet Fusion Defogging Algorithm," in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2021: IEEE, pp. 313-317.

[17] T. Michailides, "Black heart and tree decline issues of pomegranate," ed, 2011.

[18] unknown. "dataset.zip." https://www.mediafire.com/. (accessed 18, 2022).

[19] P. Vasavi, A. Punitha, and T. V. N. Rao, "Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: A review," *International Journal of Electrical and Computer Engineering,* vol. 12, no. 2, p. 2079, 2022.

[20] M. H. D. S. T. D. S. Memon, "Malignant Melanoma Detection in clinical images using SVM classification," in *International conference on computational sciences and Technology INCSST*, Mehran University Jamshoro Sindh Pakistan, D. M. Javaid, Ed., 10-12 April 2019 2019.

[21] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society,* vol. 18, no. 6, pp. 275-285, 2004.

[22] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends,* vol. 2, no. 01, pp. 20-28, 2021.

[23] Q. Ren, H. Cheng, and H. Han, "Research on machine learning framework based on random forest algorithm," in *AIP conference proceedings*, 2017, vol. 1820, no. 1: AIP Publishing LLC, p. 080020.

[24] X. Li, T. Zhai, Y. Jiao, and G. Wang, "Using Bayesian hierarchical models and random forest algorithm for habitat use studies: a case of nest site selection of the crested ibis at regional scales," *PeerJ PrePrints,* 2167-9843, 2015.