Harumanis mango leaf disease recognition system using image processing technique

R. A. JM. Gining¹, S. S. M. Fauzi², N. M. Yusoff³, T. R. Razak⁴, M. H. Ismail⁵, N. A. Zaki⁶, F. Abdullah⁷

¹,²,³,⁴Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis, Malaysia
⁵Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Perlis, Malaysia
⁶Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Perlis, Malaysia
⁷Pusat Penyelidikan, Hortikultur Mardi Sintok, Kedah, Malaysia

ABSTRACT

Current Harumanis mango farming technique in Malaysia still mostly depends on the farmers' own expertise to monitor the crops from the attack of pests and insects. This approach is susceptible to human errors, and those who do not possess this skill may not be able to detect the disease at the right time. As leaf diseases seriously affect the crop's growth and the quality of the yield, this study aims to develop a recognition system that detects the presence of disease in the mango leaf using image processing technique. First, the image is acquired through a smartphone camera; once it has been pre-processed, it is then segmented in which the RGB image is converted to an HSI image, then the features are extracted. Lastly, the classification of disease is done to determine the type of leaf disease. The proposed system effectively detects and classify the disease with an accuracy of 68.89%. The findings of this project will contribute to farmers and society's benefit, and researchers can use the approach to address similar issues in future works.

Keywords: Harumanis, Image processing, Leaf disease, Matlab

This is an open access article under the CC BY-SA license.

Corresponding Author:
S. S. M. Fauzi
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Perlis Branch, 02600, Arau, Perlis, Malaysia
Email: shukorsanim@uitm.edu.my

1. INTRODUCTION

One of Malaysia's famous mango varieties is Harumanis mango, which has a phenomenal commercial demand and thrives well, particularly in Perlis, Malaysia's northern state. Yielding good and ripe Harumanis is not an easy task; thus, farmers always must take extra care of them, mainly to prevent them from the diseases that may infect and ruin the crop. Usually, farmers will make regular monitoring through their naked eyes to detect potential diseases. For instance, the federal agriculture marketing authority (FAMA) in Malaysia uses human experts for the fruits grading process using perception and hand method.

However, this approach seems less efficient as it consumes more time and farmers may overlook some leaves that are already infected and missed the right time to prevent and treat them [1]. Besides, this approach is also not applicable to every farmer as they hugely rely on human experts. Additionally, human experts are prone to human error; hence this made the method less reliable.

Therefore, it is necessary for redefining the operation of the farming industry, and the key to this is smart farming. This approach's research effort has been dramatically expanded significantly with the rise of various methods, such as machine learning and image processing. However, most of the reported works on smart farming were conducted on rice and tomato crops, whereas very little on the Harumanis mangoes. The
previous solutions converge on the fruits rather than its leaves; which also play a vital role in crop productivity. Thus, this study intends to develop a recognition system that can quickly help farmers detect mango leaf diseases easily without any human interference.

Finding from this study will contribute to society's benefit as Malaysia's economic backbones is the agriculture sector. The progressively competitive and volatile marketing and challenges such as insects and pests ruining the crops, exhaustive farming methods and farmers attempting to produce food in the most sustainable manner justify the need for more efficient farming methods. Thus, by implementing the approach derived from this project's results, farmers can also take precautions or treat the infected leaves at the right time. The beginner farmers can also detect the diseases even without prior knowledge about it. A few major leaf diseases were detected in the Harumanis mango family such as Anthracnose, mango scab, and powdery mildew. Moreover, some farmers may also use pesticide to avoid a disease outbreak and pest growth before diagnosing the diseases. Hence, the project can reduce the use of hazardous pesticides that pose a significant threat to human health.

2. RELATED STUDIES
2.1. Harumanis mango

Many mango (Mangifera indica Linn.) is a tropical fruit that belongs morphologically to the deliquescent drupe sub-type. It possesses one large seed which is surrounded by fleshy mesocarp and has many varieties [2]. In Malaysia, 209 clones of mango have been registered by the Department of Agriculture. Any clones that possess sufficiently high qualities may be recommended for general planting from these registered clones. Malgoa and Apple Mango, Harumanis, MA 162, MA 165, MA 204, and MA 205 are among the promising ones [3]. Harumanis mango is honoured as the “King of Mangoes” in this country due to its deliciousness, unforgettable sweet taste and aromatic smell [4]. The fruit is also very rich in vitamin A, C, and medicinal qualities. Its leaves are mostly used in rituals as they are anti-bacterial against gram-positive bacteria [5]. Within the species, two types are recognized. The first type produces a seed with a single zygotic embryo. It is called monoembryonic mangoes which originated in sub-tropical India, and the other type is polyembryonic mangoes which create a seed with a few embryos [6].

2.2. Harumanis mango leaf diseases

The Harumanis leaves are usually exposed to specific mango leaf type diseases. The anthracnose (colletotrichum gloeosporioides) is one of the well-known mango leaf diseases. The anthracnose exhibits severe fungal bulb, seeds, and leaves fungus. In young shoots, flowers and fruits, the disease causes severe losses [7]. Other than that, Elsinoë mangiferae. It is also known as Denticularia maniferae causes mango scab. Only live plant tissue can survive this fungus. There are no records of this disease affecting other plants other than mangoes. Small black spots developed on newly set fruit, and the fruit falls off when multiple black lesions infect it. The remaining affected fruit on the tree develops scar tissue that makes it unmarketable or downgrades it [8].

There are other leaf diseases like Xanthomonas campestris pv. Mangiferaeindicicae, a bacterium that causes black spot capable of attacking leaves, branches and fruit [8]. There is a substantial seasonal variation in the disease severity. It may be spread through wind-driven rain from tree to tree in the field or through tools used to control tasks such as pruning [9].

The fungus, Oidium, causes powdery mildew on the leaf surface. Although a moderately sporadic disease, due to flower and panicle infection and subsequent fruit set failure, it may cause severe crop loss. The critical key to differentiating this disease is the presence of panicles and young fruit of a whitish, powdery growth of the fungus. The infected young fruit changes colour and falls. The white growth of young infected leaves can also be seen on the lower layer [10]. This study is focusing on two types of diseases which are anthracnose and bacterial black spot. These two diseases are selected as these are the most prevalent and prominent disease, and the data for it is extensively obtainable.

2.3. Related works

There are several techniques of detection and identification of crop diseases. One of those techniques, the molecular technique can be used for high-throughput research when analyzing vast numbers of samples [11]. Molecular technique refers to the minimum amount of microorganism that can be found in the sample, and ELISA is one of the widely used molecular methods for disease detection. With ELISA, the microbial protein (antigen) associated with a plant disease is injected into an animal which produces antibodies against the antigen in the detection of disease. These antibodies are collected with a fluorescent dye and enzymes from the animal's body and used for antigen detection. The sample will fluoresce in the presence of the disease-causing microorganism (antigen), thus confirming the existence of specific plant disease.
On the other hand, the hyperspectral technique can be used to collect valuable plant health information over a wide range of wavelengths ranging from 350 to 2500 nm [11]. For large-scale cultivation, hyperspectral imaging is primarily being used for plant phenotyping and for detecting crop disease. The methodology is very stable and offers a brief analysis of the image data. It is widely used for the detection of plant disease by measuring the changes in reflection resulting from changes in biophysical and biochemical characteristics at infection.

The image processing technique is another technique used for crop disease detection. Study in [12] have decided to pursue this technique to figure out specific diseases and to provide the treatment for the farmers for their sugarcane crops. The technique starts with the leaf image input taken in the form of RGB using the digital camera and then transformed into grayscale to eliminate the hue and saturation information. Next, through image pre-processing, unneeded part of the image is removed by filtering the noise and then the appropriate portion used for extraction is segmented. After that, texture statistics will be completed, and disease prevention for the sugarcane leaves is given according to the study.

Similarly, image processing technique is also used to detect the most frequently occurring diseases on cotton leaves such as bacterial blight, alternaria, grey mildew, cercospora, and fusarium wilt [13]-[15]. In these three papers, image processing also starts with image acquisition where hundreds of leaf images are captured to build a database. The images are later pre-processed to improve certain image features which are critical for further processing. The pre-processing steps include resizing the image into 250 x 250 pixel and applying the filter. Next, segmentation of the images is performed so that the infected region can be separated from the healthy region. Extraction of the features is the next critical step to do after segmentation so that a set of features representing each character can be obtained. This increases the acknowledgement rate with a minimum number of components. Following extraction of the feature, a classifier is used to identify the disease based on the features extracted.

The study in [16] have opted for an image processing technique in order to get useful critical features for the analysis of various rice blast diseases. The leaf images are first acquired through a web camera and processed by Raspberry Pi. Then, the image processing is done by converting the RGB images into grayscale, and the diseased contours are identified by applying edge detection (sobel). Next, several analytics techniques are conducted to identify the images according to the specific problem at hand. By using optimization techniques, the images will then be sent to the cloud storage for comparison.

Other researchers have also used image processing technique which involves similar steps such as image acquisition, image pre-processing, image segmentation, feature extraction and classification [5], [17]-[19]. In these papers, during the image segmentation, Otsu’s method and K-means clustering are used to classify the object based on a set of the features into K number of classes. K-means clustering is more superficial than other clustering methods and works with a vast number of variables as well. However, a different number of cluster numbers and different initial centroid values produces a different cluster result. Hence, initializing the appropriate number of cluster k and the correct initial centroid number is necessary. Some of the drawbacks of the molecular and hyperspectral technique are time-consuming, labour-intensive and require an elaborate process. The drawback occurred, especially during sample preparation (collection and extraction) in order to obtain reliable and accurate results on the identification of plant disability [20].

Image processing technique, on the other hand, minimizes the subjectivity of traditional classification methods and mistakes committed by human beings [17]. Thus, the reliability of the estimation is enhanced, and reliable data for disease studies are given. The approach is also simple, in which it only needs merely computers, digital cameras, as well as the software programs. The image processing approach also minimizes the subjectivity of traditional classification methods and is more straightforward. Other image processing techniques that uses hybrid approach [21], image segmentation [22] and cluster-based feature [23] also proven to be useful in the process of image processing technique. Hence, according to the stated benefits, the proposed recognition system implements the image processing technique to detect the Harumanis mango leaf diseases.

3. METHODOLOGY

As stated in the previous section, the technique utilized for the development of the prototype is the image processing technique. Image processing technology used for plant disease detection eliminates the subjectivity of traditional classification methods and human-induced error. Thus, the estimation reliability is improved, and accurate data acquired for disease studies. The technique is also convenient, which needs computers, digital cameras with the combination of necessary software programs to realize for disease detection system [24].

The necessary steps for disease detection using image processing are: image acquisition, pre-processing, feature extraction, segmentation and finally, classification [5], [17]-[19]. In this study, the
MATLAB's Image Processing Toolbox was utilized. It provides a comprehensive collection of standard reference algorithms and workflow frameworks for image processing, analysis, visualization, and algorithms development. This toolbox was used to perform the image pre-processing, image segmentation, extraction of feature and classification process. The MATLAB's Guide also employed to develop the GUI of the prototype.

3.1. Image acquisition

The process of acquiring leaves images by capturing images of different types of leaves using a high-resolution camera to obtain good results and efficiency [25]. The leaves' images were taken manually using a Redmi Note 7 smartphone camera, taken at two harumanis mango farms located at Repoh and Batu Bertangkup, Perlis. There are 103 healthy leaf images, and 43 images of the diseased leaves (anthracnose and bacterial black spot) were obtained. Each type of images sample was divided into two categories; training (±70%) and testing (±30%) images, saved into a file named "dataset". Both categories undergo the same processes throughout the methodology phases. However, the training is done first before the testing. Hence, training the system using training images will be discussed first, followed by the testing, which uses testing images.

For the system training, each training image is uploaded into the system, running through a series of functions implementing the pre-processing, segmentation, and feature extraction phases consecutively - this will be discussed more in the description of each phase. The system will then prompt to label the uploaded images using the following tags: 0 for healthy, 1 for anthracnose and 2 for bacterial black spot. These labelled data are saved in a database named db.mat.

3.2. Image pre-processing

Before extraction and classification, it is standard practice for leaf images to go through pre-processing. Various pre-processing techniques can be taken into account to eliminate the noise from the acquired images; the usage of image clipping to get the area in question by cropping, the smoothing filter for image smoothing, and image enhancement to control the contrast level [17]. In this phase, the noise on the training images was reduced using a filtering technique - image enhancement, which increased the image's contrast. Two MATLAB's filtering technique functions were applied in this phase, imadjust to increase the contrast and stretchlim to adjust the image intensity. The GUI for testing was also developed to illustrate the overall testing processes. The imread function is utilized to enable the testing image upload function. Figure 1 illustrates the two functions' execution with a click on the enhance image contrast button.

![Image enhancement](image1.png)

Figure 1. Image enhancement

3.3. Image segmentation

Image segmentation is a method by which an image is more meaningful and convenient to analyze [25]. Segmentation comprises partitioning the image, or any relation, into separate parts of the same features. Upon enhancing the image, the RGB picture was then transformed to hue, saturation, and intensity (HSI). The transformation was enabled using im2double function that converts true RGB colour image to double precision and, rescale the data and extract the individual component of the image where appropriate. Afterwards, the conversion equation is applied to transform the RGB image into HSI format. Figure 2 depicts the segmented images after clicking on the Segment Image button.
The extraction of the feature plays a vital role in image processing. Feature extraction is used in many applications of image processing. Colour, texture, and morphology can be considered as a feature for leaf disease detection. The grey level co-occurrence matrix (GLCM) for each pixel map for H and S images of infected clusters was created for this system. GLCM statistics such as mean, standard deviation, entropy, RMS, variance, correlation, energy, and homogeneity were extracted.

3.5. Disease classification

After the previous phase's execution, the result could be viewed by clicking the view result button. Figure 3 shows the system successfully processed the image and identified the leaf disease, in which in this demonstration, the leaf was infected by anthracnose.

4. RESULTS AND DISCUSSION

The system's performance was evaluated to test its accuracy, recall, precision, and F1-score in detecting the leaf diseases. The details of this evaluation are discussed in this section. To deliver the system evaluation, an experiment was carried out on 146 images of the three chosen classes which were healthy leaf as well as leaf that was infected by anthracnose and bacterial black spot. Table 1 shows the image dataset that was divided into two subsets where a set of 113 images are used for training and a set of 33 images are used for the test.

After dividing the images into the two subsets, the 33 images from the test data were tested and the results were recorded in a confusion matrix table. A matrix of confusion is a table depicting the values of true negative, true positive, false positive as well as false negative. This allows thorough results rather than only the correct proportion of guesses (accuracy) [3]. Figure 4 shows the general confusion matrix table.
Table 1. Dividing image datasets into training data and test data

| Disease class         | Total image datasets | Training data | Test data |
|-----------------------|----------------------|---------------|-----------|
| Healthy               | 103                  | 83            | 20        |
| Anthracnose           | 30                   | 20            | 10        |
| Bacterial Black Spot  | 13                   | 10            | 3         |
| Total                 | 146                  | 113           | 33        |

The next step was to calculate the accuracy, precision, recall and F1 scoring. Accuracy represents the amount of test data that were accurately identified over the sum of test data. In (1) displays the formula of overall accuracy. Precision is the ratio of correctly expected positive observations to the actual total predicted observation. In (2) illustrates the equation of precision. Recall is the ratio of correctly predicted positives to all actual class observations. In (3) indicates the recall formula. Lastly, F1 is the harmonic mean of precision and recall and is a better measurement than accuracy. In (4) shows the F1-score equation.

\[
\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (2)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (3)
\]

\[
F_1 - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)
\]

Where TP- True Positive, TN- True Negative, FP- False positive, FN- False Negative. The confusion matrix table of Harumanis Mango Leaf Disease Detection System is shown in Table 2. The classes were a healthy leaf, the leaf infected by anthracnose and the leaf infected by the bacterial black spot.

Table 2. Confusion matrix table of the system

| Disease class         | TP  | TN  | FP  | FN  |
|-----------------------|-----|-----|-----|-----|
| Healthy               | 10  | 4   | 6   | 0   |
| Anthracnose           | 7   | 2   | 1   | 0   |
| Bacterial Black Spot  | 2   | 1   |     |     |

Figure 4. General confusion matrix table

From Table 2, TP indicates that the leaf is either healthy or infected and was correctly identified by the system. TN indicates that the leaf is not healthy or not infected and was correctly identified by the system. FN indicates that the leaf is infected, but the system identified it as healthy. FP indicates that the leaf is healthy, but the system identified it as infected.

Few factors might contribute to why the system inaccurately detect the diseases. The first factor is the angle from which the leaf image was taken. If the camera was not positioned well, the captured image might be intricate for the system to process primarily during the segmentation process. The leaf was still considered healthy even though there were a few black or brown spots on the leaf. The system might interpret these spots as an infected leaf when it was not. The lack of leaf images infected by anthracnose and bacterial black spot might also contribute to why the leaf's inaccurate classification.

Table 3 shows the accuracy, precision, recall and F1-score for each class based on the confusion matrix from Table 2. These results were obtained after performing the calculation based on the four equations.
presented. A successful classifier’s precision should preferably be 1, and high recall implies that the class (a small number of FN) was identified correctly. F1-score is only one if both recall and precision are 1. The F1 score only gets high if both recall and precision are high.

After the experiment, the results obtained were then compared and analysed. Table 4 depicts the comparison between the proposed approach and the other two existing approaches based on the classifier used and the three systems’ accuracy rate. The overall accuracy of the proposed system is 68.89%.

Table 3. The accuracy, precision, recall and F1-score of the system

| Disease class     | Accuracy (%) | Precision | Recall | F1-score |
|-------------------|--------------|-----------|--------|----------|
| Healthy           | 70           | 0.71      | 1      | 0.77     |
| Anthracnose       | 70           | 1         | 0.70   | 0.82     |
| Bacterial Black Spot | 66.67      | 1         | 0.67   | 0.80     |

Table 4. Comparison between the proposed system with the existing system

| Paper                  | Accuracy (%) |
|------------------------|--------------|
| Healthy                | 87.80        |
| Anthracnose            | 83.26        |
| Bacterial Black Spot   | 68.89        |

From the table above, it is shown that the proposed system had the lowest accuracy rate. The lack of training images causes this lowest result. For instance, both existing systems had trained 629 and 202 images compared to the proposed system, which only trained 113 images. Both existing systems also used SVM classifiers, giving more accurate results, especially when more images were trained.

Next, the features that the two existing systems and the proposed system had may also impact the accuracy of the systems. Research in [26] extracted the leaves’ texture, colour, and shape features, whereas in [27] extracted the texture and colour features during the feature extraction process. The proposed system only extracted the texture feature; thus, this might contribute to an overall lack of accuracy.

5. CONCLUSION

This paper described a study on developing a system that detects disease in the mango leaf using an image processing technique. The proposed system effectively detects and classifies the disease with an accuracy of 68.89%, which is low, resulted from the inadequacy of training images. Additional diseases will be integrated with the recognition system as part of future works.

6. ACKNOWLEDGEMENT

We gratefully acknowledge support by UniversitiTeknologi MARA Cawangan Perlis, Malaysia under dana pembudayaan penyelidikan dalaman (DPPD) under Grants No 600-TNCPI 5/3/DDN (09) (019/2020).

REFERENCES

[1] Chuanlei, Z., Shanwen, Z., Jucheng, Y., Yancui, S., and Jia, C., “Apple leaf disease identification using genetic algorithm and correlation based feature selection method,” International Journal of Agricultural and Biological Engineering, vol. 10, no. 2, pp. 74-83, 2017, doi: 10.3965/j.ijabe.20171002.2166.
[2] Huang, C. Y., Kuo, C. H., Wu, C. H., Kuan, A. W., Guo, H. R., Lin, Y. H., and Wang, P. K., “Free Radical-Scavenging, Anti-Inflammatory, and Antibacterial Activities of Water and Ethanol Extracts Prepared from Compressional-Puffing Pretreated Mango (Mangifera indica L.) Peels,” Journal of Food Quality, 2018, doi: 10.1155/2018/1025387.
[3] Yin, M. O., and Nay, C. H., “Plant Leaf Disease Detection and Classification using Image Processing,” International Journal of Research and Engineering, vol. 5, no. 9, pp. 516-523, 2018, doi: 10.21276/ijre.2018.5.9.4.
[4] Tharanathan, R. N., Yashoda, H. M., and Prabha, T. N., “Mango (Mangifera indica L.), “the king of fruits” - An overview,” In Food Reviews International, vol. 22, no. 2, 2006, doi: 10.1080/87559120600574493.
[5] Sethupathy, J., & Veni, S., “OpenCV based disease identification of mango leaves,” International Journal of Engineering and Technology, vol. 8, no. 5, pp. 1990-1998, 2016, doi: 10.21817/ijiet/2016/v8i5/160805417.
[6] Ploetz, R. C., “Diseases of tropical fruit crops,” Diseases of Tropical Fruit Crops, 2003, doi: 10.1079/9780851993904.0000.
BIOGRAPHIES OF AUTHORS

R. A. J. M. Gining is now serving as a lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA. Hence, according to the stated benefits, the proposed recognition system implements the image processing technique to detect the Harumanis mango leaf diseases. A, Perlis Branch, Malaysia. He received his Diploma in Computer Science and Bachelor Science (Hons) specializing in Information Systems Engineering from the Universiti Teknologi MARA; and Master of Science in Cloud Computer from Newcastle University, United Kingdom.
Associate Professor Ts. Dr Shukor Sanim Mohd Fauzi is a Deputy Rector (Research and Industrial Linkages) at UiTM Perlis Branch, and also a faculty member of Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis Branch, Malaysia. He received Master of Science (Computer Science - Real-Time Software Engineering) from the Centre for Advanced Software Engineering, Universiti Teknologi Malaysia. He then obtained his PhD in Software Engineering from the University of New South Wales (UNSW).

N. M. Yusoff is a research student from the Faculty of Computer and Mathematical Sciences and is under guidance and direction from the researchers of Applied Computing and Technology Research Group.

T. R. Razak currently works at Universiti Teknologi MARA, Perlis Branch, Malaysia and also a faculty member of Faculty of Computer and Mathematical Sciences. He received Bachelor of Information Technology (Hons) specializing in Artificial Intelligent from the Universiti Utara Malaysia, and Master of Science (Intelligent Systems) from the Universiti Utara Malaysia. He then obtained his PhD in Computer Science from the University of Nottingham, United Kingdom.

M. H. Ismail is a lecturer and researcher from Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis Branch, Malaysia. He obtained his first degree in Data Communication and Network and his master’s degree in information technology. His primary research interest is Mobile and Pervasive Computing and is actively involved in mobile technology solution in his community.

N. A. Zaki, holds a Doctor of Philosophy (PhD) from Universiti Teknologi MARA, Shah Alam in 2018 in the field of Geomatics (Remote Sensing) and leads to the field of carbon stocks for forests. Dr Nurul Ain is currently serving at Universiti Teknologi MARA (UiTM) Perlis Branch at the Faculty of Architecture, Planning and Surveying.

F. Abdullah is currently working as Senior Research Officer at Malaysian Agricultural Research and Development Institute (MARDI) and now holds a position of Head of Station at Horticulture Research Centre MARDI Sintok, Kedah, Malaysia. He received Bachelor Science Bio-industry, majoring in Plant Physiology from the University Putra Malaysia. He then obtained Post Graduate Diploma in Science (Horticultural Science) and Ph.D in Plant Science (Fruit Crop Physiology) from the Massey University, Palmerston North, New Zealand.