International Journal of Multidisciplinary Research and Growth Evaluation.

Detection of cucumber and watermelon diseases based on image processing techniques using K-means algorithm

Samuel Amachundi Adda [1*], Ibeabuchi Benjamin Nwaogwugwgu [2], Adi Wama [3]
[1] Department of Computer Science, Federal University, Wukari, Taraba State, Nigeria
[2] Department of Computer Science, Godfrey Okoye University, Enugu, Nigeria
[3] Computer Science Department, School of Sciences, College of Education, Zing, Nigeria

* Corresponding Author: **Samuel Amachundi Adda**

Article Info

ISSN (online): 2582-7138
Volume: 04
Issue: 06
November-December 2023
Received: 26-08-2023;
Accepted: 27-09-2023
Page No: 38-46

Abstract
Agriculture in Nigeria is a branch of its economy providing employment for over 70% of its population and contributing about 41% to it gross domestic production (GDP). Nigeria's wide range of climate variations allows it to produce a variety of food and cash crops. Cucumber/watermelon is among the few varieties of fruits and vegetables produced in the country. The need for steady monitoring of the plants is necessary to detect and control the spread of its diseases. Common practices for detecting these plants diseases are mostly based on direct observation of the affected plant, which are sometimes erroneous. Laboratory analysis can also be used for plant diseases detection but mostly costly and time consuming. Digital image processing can identify and grade the diseases in cucumber/watermelon. This will aid in describing and predicting the performance of the said cultivated crops, hence increasing the production yield. Although, official disease recognition is a responsibility of professional agriculturists, low-cost observation and computational assisted diagnosis can effectively help in the recognition of a plant disease in its early stages. The application software was designed based on Object Oriented System Analysis and Design Methodology while UML was used to model the system; MATLAB was used in designing the front-end and the database is MYSQL Server. The developed system works efficiently and can successfully detect and classify the specified diseases in cucumber/watermelon with a precision between 92% and 98%.

DOI: https://doi.org/10.54660/.IJMRGE.2023.4.6.38-46

Keywords: Digital image processing, k-means, plant diseases, cucumber diseases, watermelon diseases

Introduction

Nigeria is an agricultural country and as such, agriculture has become a sure means to feed the nation's ever-growing population. More than 70% of the nation's population depends on agriculture and over 70% of the entire land mass is arable. However, only about 48% is presently cultivated (Foraminifera, 2016) [5].

There is a diversity of suitable fruits and vegetables available for farmers to select from. However, to cultivate them for optimal yield capable of satisfying customers' needs is highly technical and involving. The management of perennial fruit crops requires close monitoring especially in putting measures on ground for the management of diseases that can affect production yield and subsequently the post-harvest life (Varshney & Tarun, 2016) [23].

Disease management is a challenging task. Huge number of diseases manifest on leaves, stems and fruits of plants. In most cases, it is quite difficult to determine the accurate state of these diseases, owing to the fact that most visually observed diseases have not been fully understood because of the intricacies of their visual patterns (Kumar & Gupta, 2012; Shire *et al.*, 2016) [12, 22]. A common practice for plant scientists is to estimate the damage of a plant (leaf, stem, fruit) through direct observation on a scale based on the percentage of affected area. It results in subjective and low production yield (Varshney & Tarun, 2016) [23].

There are various methods to detect plant pathologies. Diseases can be detected and analyzed using digital image processing techniques. Image processing is a method of improving or extracting information from an image by performing operations on it.

Image processing basically includes the following three steps:

1. Importing the image via image acquisition tools.
2. Analyzing and manipulating the image.
3. Output can be an altered image or report that is based on the image analyzed.

This research provides an automated and more accurate study in cucumber/watermelon diseases through digital image processing techniques. The system deployed aimed at increasing crop production and reducing subjectiveness arising from crop scientists in detecting the cucumber/watermelon diseases, proffering necessary recommendations disease control, and saving the records for future references.

Related Literatures

Cucumber and watermelon are members of the Cucurbitaceae plant family, which includes a variety of crops known as cucurbits. While each plant belongs to a distinct genus and is categorized as a different species, they are all connected. Cucumber, watermelon, and other Cucurbitaceae family members have comparable growing and maintenance requirements, as well as similar physical traits (Renee, 2020) [18].

Image processing is a technique used in exploring and detecting various objects and providing the required output in the form of images or other detailed reports (Ahmed &Jun, 2019; Gokulakrishnan, 2014) [7]. The processed and analyzed images are better understood and evaluated. Digital image processing technique is extremely effective in providing symptoms or characteristics of diseases in plants by detecting and recognizing diseases in those plants even at its early stage.

Revathi & Hemalatha, (2012) [19] developed a plant diseases identification system using image processing techniques, using edge image segmentation detection techniques, and the proposed algorithms for images analysis and classification is Homogenous Pixel Counting Technique for Cotton Disease Detection (HPCCDD). The aim of their study was to use image processing to discover diseases of cotton leaf spot and to analyze the input photos using RGB pixel counting. Finally, farmers are provided pest recommendations in order to protect their crops and prevent production loss.

Furthermore, Hungilo *et al*., (2019) [10] developed a software prototype system for disease detection using image growth and image segmentation techniques. The paper shows a software prototype system for rice disease detection based on contaminated photographs of different rice plants. Digital cameras captured photos of infected rice plants and analyzed them to identify infected plant parts. After that, the infected part of the plants was used for disease classification purposes. The system used both image processing and soft computing techniques applied to a number of diseased rice plants.

Patil & Zambre (2014) [14] developed an image processing system for recognition of crop diseases. They employed cucumber powdery mildew, speckled mildew, and downy mildew as study materials, and they tested the approach of image pre-processing for diagnosing crop diseases. This paper contrasted the effect of two filters - Simple filter and Median filter and used the median filter to effectively wipe out the noise disturbance, and a two-apex approach was applied to isolate the images of the diseases from the background. Snake model and image edge detection were used to identify the affected part of the plant. The image pre-processing provided a good foundation for disease diagnoses and setting up a pattern recognition system.

Shire *et al*. (2016) [22] designed a system that collects digital images, processes them for plant diseases detection. Segmentation was performed using the Otsu process, K-means clustering, converting RGB to the HSI model, etc. followed by extraction of features that plays a significant role in identifying an object. Support Vector Machine (SVM) was used with the. Mat file for the detection and classification. The SVM is the supervised learning model with an associated learning algorithm that analyzes the classification and regression data. The name of the pesticide will be sent to the farmer's mobile via GSM after identification of the diseases and its corresponding pesticide.

K-Means Clustering

A method for grouping or clustering a set of data is called clustering. K-means clustering is one of the widely used techniques. When using K-means to cluster data, the data collection is divided into k groups. A given data set is categorized into a disjoint cluster with number k (Dabbura, 2018) [4]. The K-means algorithm is divided into two parts. In the first stage, it determines the centroid k, then moves each point from its respective data point to the cluster with the closest centroid in the second stage (Rao & Reddy, 2012) [16]. Euclidean distance is one of the most commonly used methods for calculating the distance to the nearest center of gravity. Once the clustering is complete, it recalculates the centroid of each cluster, calculates the new Euclidean distance between each centroid and each data point based on the centroid, and assigns the minimum Euclidean distance points to the cluster. The member objects and centroids of each cluster in a partition define the partition. For each cluster, the centroid is the point at which the sum of the distances from all objects in that cluster is minimized (Dabbura, 2018) [4]. Thus, K-means is an iterative algorithm in which it minimizes the sum of the distances from each object to its cluster centroid.

The Way K-Means Algorithm Works

According to Dabbura, (2018) [4], below is how k-means algorithm works;

1. Specify number of clusters *K*.
2. Initialize centroids by first shuffling the dataset and then randomly selecting *K* data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. That is, assignment of data points to clusters isn't changing.

- Compute the sum of the squared distance between data points and all centroids.
- Assign each data point to the closest cluster (centroid).
- Compute the centroids for the clusters by taking the average of all the data points that belong to each cluster.

K-Mean Clustering Algorithm

1. Initialize Pix [] with image pixels
2. Select Random Centroids []
3. For i=0 to pix.len
4. For j=0 to Centroids [].len
5. Calculate Eu[i] [j] = (I, sqrt (Sum (pix[i]-centroid[j]), 2))
6. End for
7. Arrange Eu in ascending order
8. Classify the pixel into the cluster having centroid j for which the pixel has minimum Euclidean distance
9. End for
10. Calculate centroid mean
11. Check whether any Euclidean distance is greater than centroid mean or not
12. If Euclidean distance of any pixel is greater than centroid mean then
13. Make the pixel as centroid and repeat step 4 until the centroids get stable
14. Rearrange clusters to form new cluster image
15. Save image

Feature Extraction

Feature extraction is the process of decreasing the number of resources needed to describe a huge amount of data. Image processing is one of the most important areas of application, where algorithms are used to recognize and extract certain desired portions or forms (features) of a digitized image or video stream. The selected features should contain the necessary information from the input data, allowing the intended task to be completed using this reduced representation rather than the entire initial data. Below are some features that can be extracted from a digital image:

1. Mean: - According to Ravichandran *et al.* (2016) [17], mean is the most fundamental of all statistical measures. In geometry and analysis, means are frequently utilized. In image processing applications, mean filtering is categorized as spatial filtering and is used for noise reduction. This is an average which indicates the general brightness of the image and is given by

$$\text{Mean } (\mu) = \frac{1}{M*N} \sum_{x=0}^{M} \sum_{y=0}^{n} p(x,y)$$

Where ΣΣ p(x, y) represents the summation of all pixel values of the image and M*N is the size of the image. A high mean suggests that the image is bright, whereas a low mean suggests that the image is dark.

2. Standard deviation (σ): - In statistics, it is the most extensively used measure of variability or diversity. It displays how much variation or "dispersion" there is from the average in image processing (mean, or expected value). A low standard deviation means the pixel values are close to the mean, whereas a high standard deviation means the pixel values are spread out over a wide range of values (Kumar & Gupta, 2012) [12]. As a result, the standard deviation of an image is a measure of the frequency distribution of the pixel value. The standard deviation can be calculated using the given formula:

$$\text{std dev } (\sigma) = \sqrt{\frac{1}{M*N} \sum_{x=0}^{M} \sum_{y=0}^{n} (p(x,y) - \mu)^2}$$

Where ΣΣ p(x, y) represents the summation of all pixel values of the image and M*N is the size of the image.

3. Variance: The variance is a measure of how evenly distributed a group of numbers is. It's one of numerous probability distribution descriptors that describes how much the numbers deviate from the mean (expected value). One of a distribution's moments in particular is the variance. In that situation, it is a component of a methodical method for differentiating between probability distributions (Kumar & Gupta, 2012) [12]. Utilizing the provided formula, the variance is determined as the square of the standard deviation:

$$\text{Variance} = (\text{std dev}(\sigma))^2$$

An image with a high variance has a strong contrast, while an image with a low variance has a low contrast.

4. Energy: A normalized histogram of the image is used to define energy. The distribution of gray levels is depicted by energy. Energy is strong when the number of gray levels is minimal. Energy can be a negative measure that should be minimized at times and a positive metric that should be boosted at other times. The energy can be calculated using the given formula:

$$\text{Energy (E)} = \frac{1}{M*N} \sum_{x=0}^{M} \sum_{y=0}^{n} p(x,y)^2$$

Where M*N is the size of the image and p(x, y) is the value of the pixel of the image.

5. Entropy: Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image (Wirth, 2004) [26]. The entropy of a grayscale image is calculated, as well as the scalar value. Entropy is the measure of information content if an image is viewed as a source of symbols or gray levels. In this study, the entropy of an image is determined by:

$$\text{Entropy (En)} = \frac{1}{M*N} \sum_{x=0}^{M} \sum_{y=0}^{n} p(x,y) \log \log 2\, (p(x,y))$$

Such a matrix corresponds to an image in which there are no preferred gray level pairs for the distance vector d. Entropy is highest when all entries in p(x,y) are of similar magnitude and small when the entries in p(x,y) are unequal.

6. Skewness: - The degree of symmetry, or more precisely, its absence, is called skewness. A distribution or collection of data is said to be symmetric if it appears equally to the left and right of the center point. Any symmetric data should have a skewness that is relatively close to zero because the normal distribution has zero skewness. Positive skewness values indicate data that are skewed to the right, while negative skewness values indicate data that are skewed to the left (Ravichandran *et al.*, 2016) [17]. By slanting to the left, we mean that the left tail is longer compared to the right tail. The slope is calculated using the following formula:

$$\text{Skewness(s)} = \frac{1}{M*N} \sum_{x=0}^{M} \sum_{y=0}^{n} \left[\frac{p(x,y) - \mu}{\sigma}\right]^3$$

$$\text{Where Mean } (\mu) = \frac{1}{M*N} \sum_{x=0}^{M} \sum_{y=0}^{n} p(x,y)$$

$$\text{Std Div } (\sigma) = \sqrt{\frac{1}{M*N}\sum_{x=0}^{M}\sum_{y=0}^{n}(p(x,y)-\mu)^2}$$

Darker and glossier surfaces are more favorably biased in digital image processing than brighter and plain surfaces. As a result, we can use skewness to make decisions about image surfaces.

7. Kurtosis: Kurtosis is a statistical measure of the form of a real-valued random variable's probability distribution. It is closely related to a distribution's fourth instant. A high kurtosis distribution has longer, thicker tails and a sharper peak (though not always). A low kurtosis distribution has shorter, thinner tails and a more rounded peak (though not always)(Kumar & Gupta, 2012) [12].

To put it another way, data sets with a high kurtosis feature a prominent peak near the mean, a rapid decrease, and heavy tails. Low kurtosis data sets are more likely to have a flat top near the mean rather than a high peak. The most extreme instance would be a uniform distribution. Kurtosis (k) is calculated by using the given formula:

$$\text{Kurtosis (k)} = \frac{1}{M*N}\sum_{x=0}^{M}\sum_{y=0}^{n}\left[\frac{p(x,y)-\mu}{\sigma}\right]^4 - 3$$

Kurtosis values are evaluated in digital image processing in conjunction with noise and resolution measurement. Low noise and low resolution should go hand in hand with high kurtosis levels.

8. Homogeneity: Another restriction that is frequently taken into account for the processes in an image field when estimating the low-dimensional points using the existing image statistics is the homogeneity assumption (Wirth, 2004) [26]. From a given image, the joint histograms of the low-dimensional process points can be used to estimate them. The discussion centers on the fact that a typical image resolution only offers a sufficient number of sample points for a good estimation of two-dimensional p(x,y)s, as illustrated below:

$$H_c = \sum_{x=0}^{M}\sum_{y=0}^{n}\frac{p(x,y)}{1+[x-y]}$$

Where the range of gray levels is small the p(x,y) will tend to be clustered around the main diagonal. A heterogeneous image will result in an even spread of p(x,y)'s.

Propose System Procedure

The proposed method used to analyze and to develop the system is Object Oriented Analysis and Design (OOAD). The design tools used in this work are Use Case Diagram, Sequence Diagram and High-Level Diagram

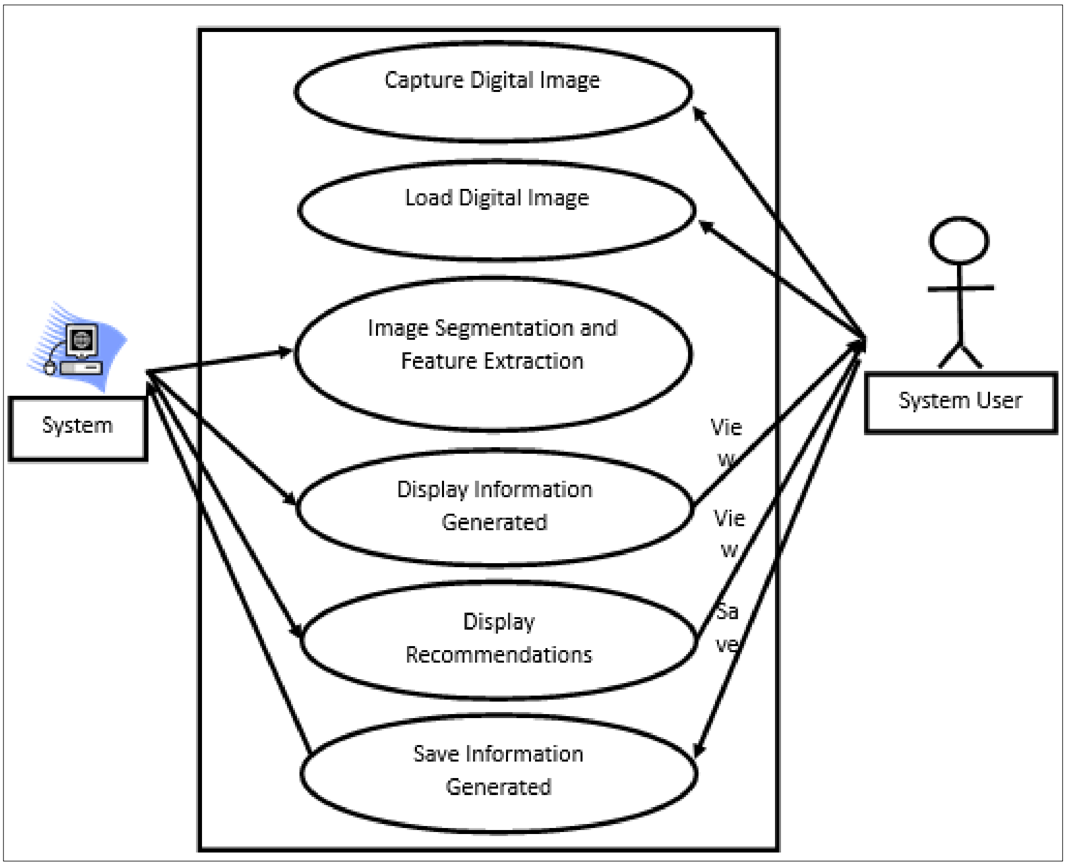

Fig 1: Use case Diagram of the proposed system

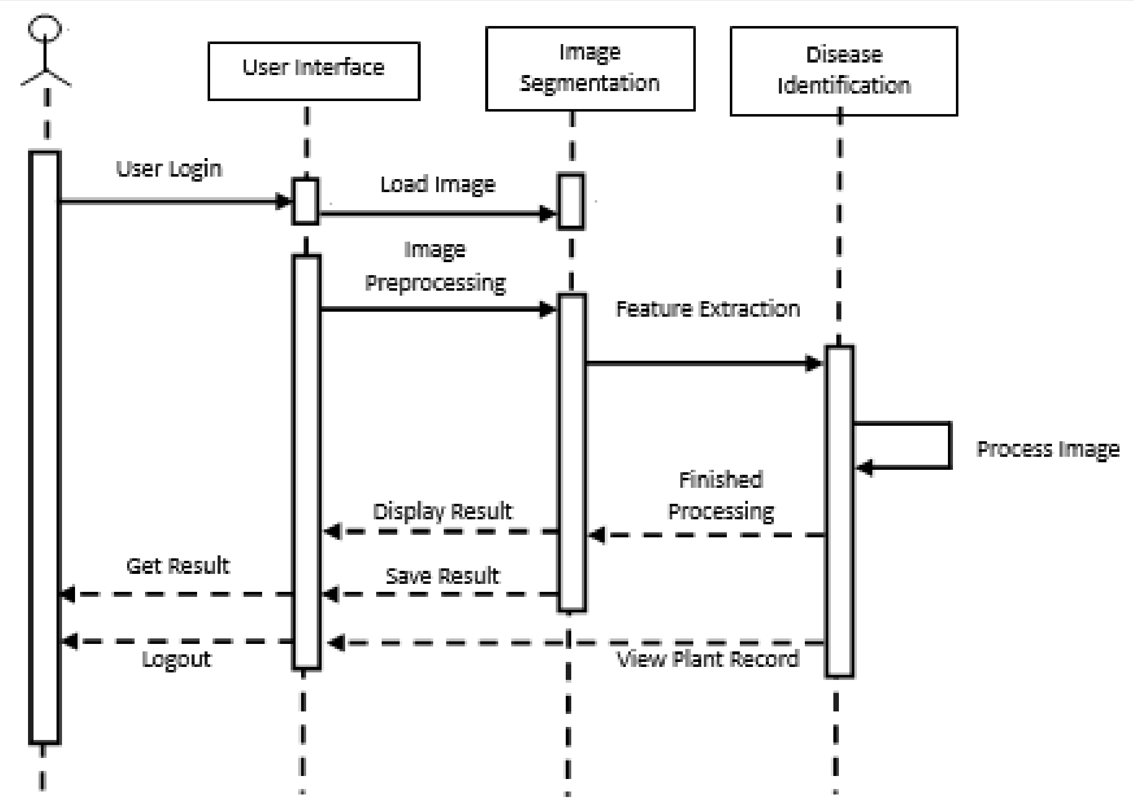

Fig 2: Sequence diagram of the Proposed System

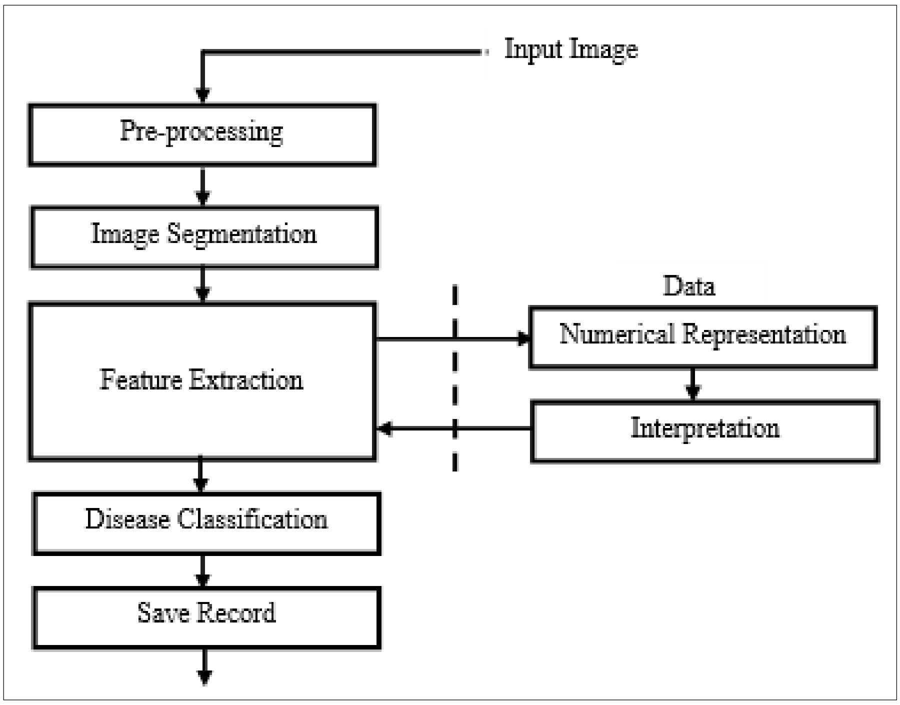

Fig 3: High Level Model of the Proposed System

Implementation Architecture

The architectural implementation of the system is to identify diseases in cucumber and watermelon using image processing techniques. First, the system captures the digital image from an external device, then the image is being pre-processed to remove noise and to enhance the image. The technique involves image segmentation, features extraction of the infected region. K-means algorithm is used for the segmentation to detect the region of interest which is the infected part of the plant, then feature extraction and classification is based on Artificial Neural Network. The features extracted from the image with the region of interest are compared to the information saved in the database as the system dataset, from which the disease affecting the plant is determined.

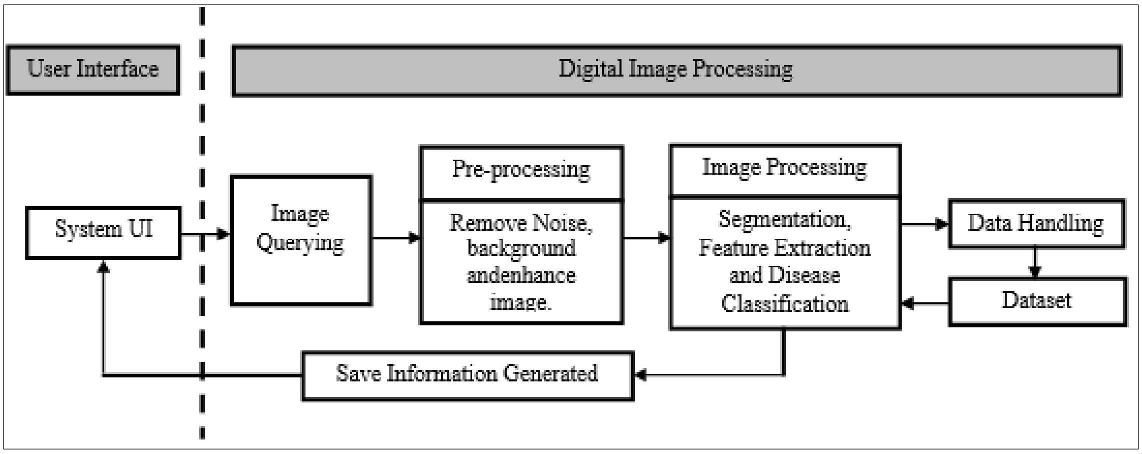

Fig 4: System Implementation Plan

Home page describing the different modules in the program. To get started, click on "Load Image" button.

Fig 5: Home Page

Click on "Load Image" and use the dialog box to load the image of the affected part of the plant. After which you click on the "Add Image" button next in line. Repeat this process three times to get different samples of leaves for analysis.

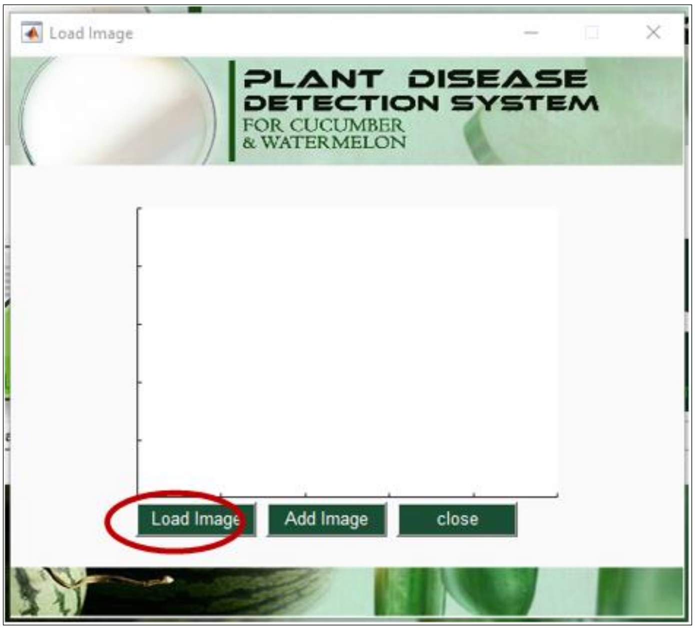

Fig 6: Load Image window

After a successful loading of the images, click on "Feature Extraction" to load the segmented K-means of the images where you choose the image with the correct ROI for further processing.

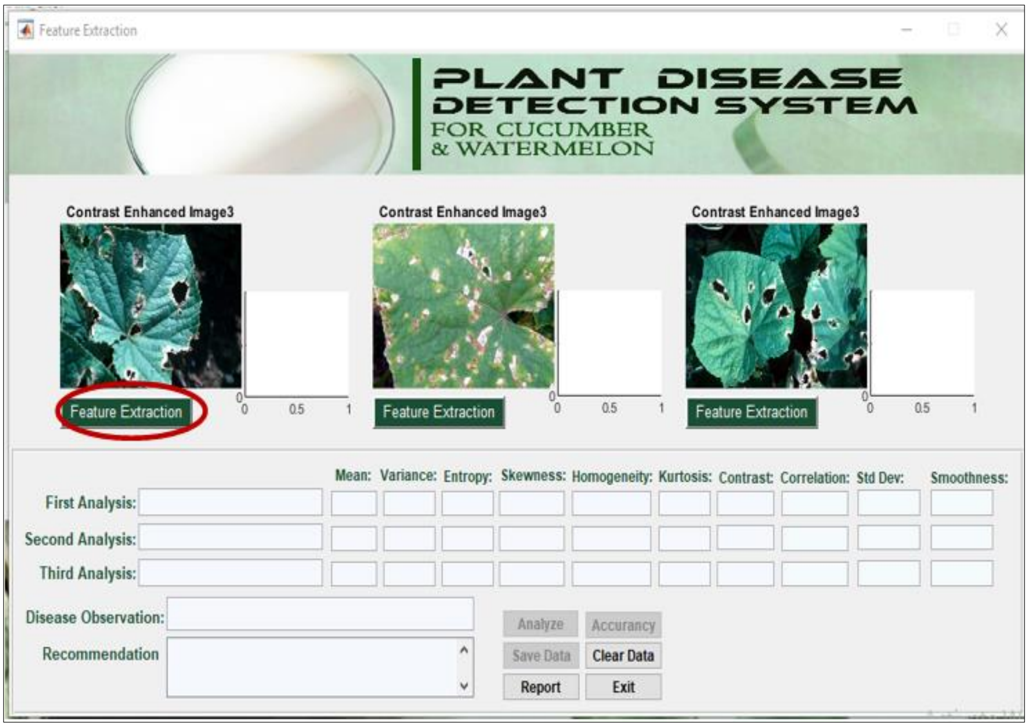

Fig 7: Feature Extraction window

This page consists of Analyze, Accuracy, Save Data, Report and other function button. The "Analyze" button display the analysis of the features extracted from the affected leaves of the plant and generate result and necessary recommendations.

The "Save Data" button add or update the database of the application with new record. The "Record" button generates the records saved on the database of the application for further decision making. These are shown in the figure below:

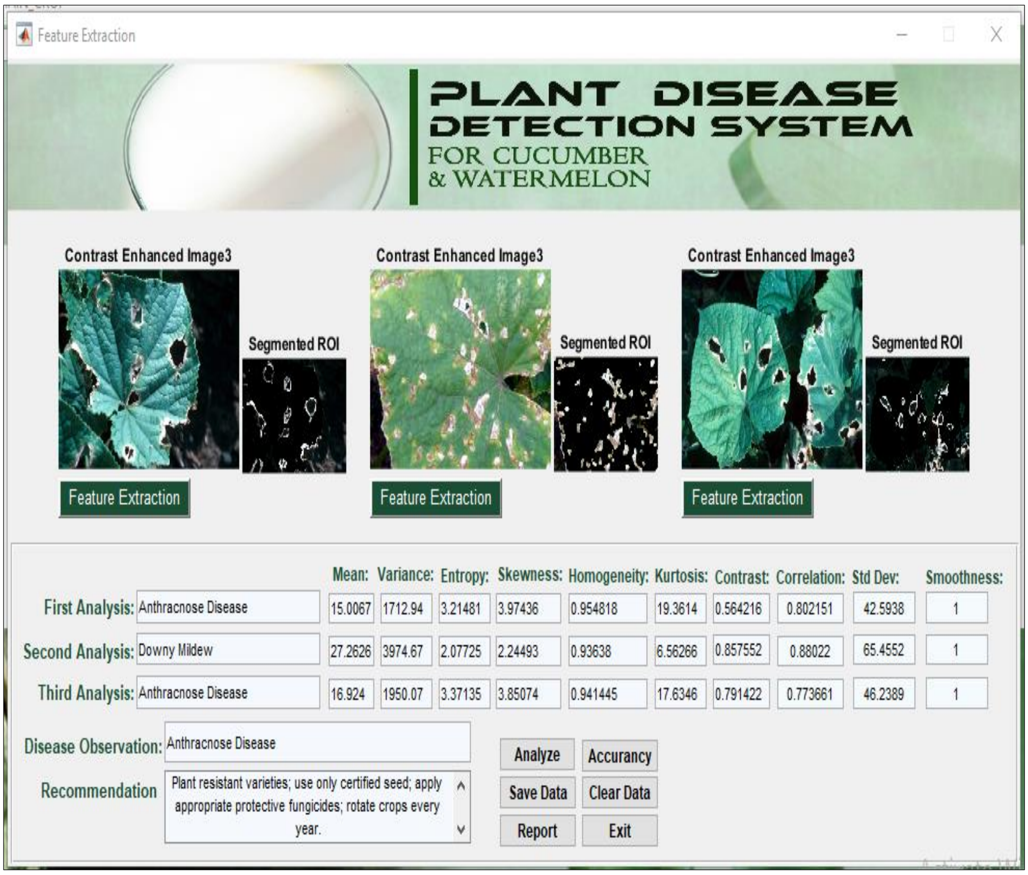

Fig 8: Feature Extraction window

This page display the records saved on the system.

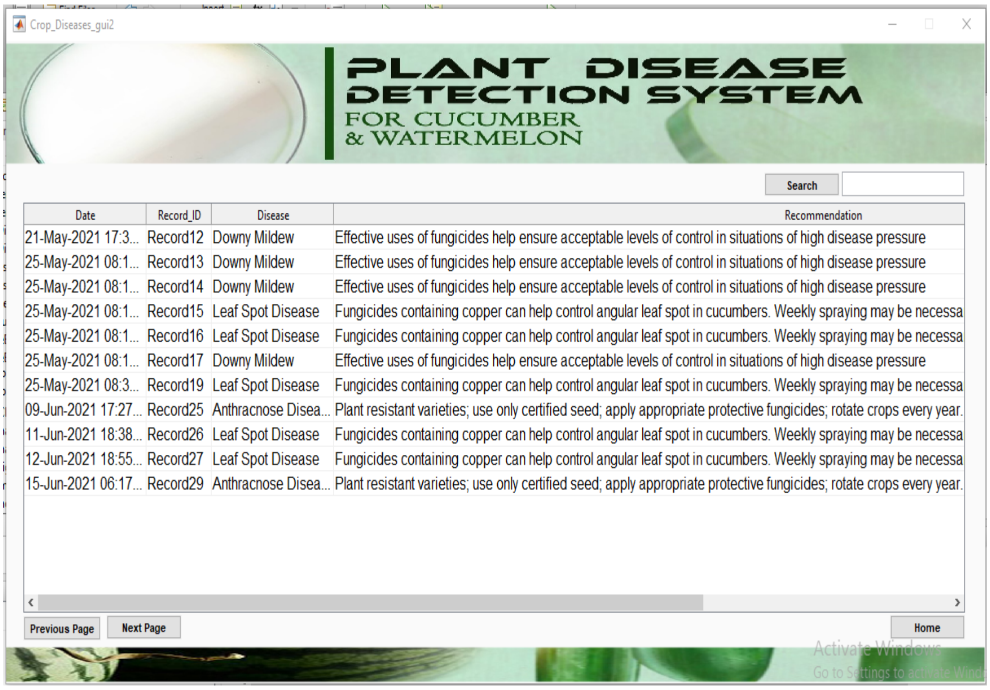

Fig 9: View Saved Record Window

Conclusion

The aim of this study is to develop an Image Processing System that will classify and detect diseases in cucumber and watermelon. Using image processing techniques, we were able to classify and detect diseases in cucumber/watermelon. This system's primary goal is to detect diseases in plants which when treated can increase productivity. The system identified 5 different types of disease in cucumber and watermelon. Despite the constraints, we have been able to come up with a system adjourned to be more efficient, accurate and reliable in terms of disease detection, data storage and information retrieval than the manual system currently being used by some farmers.

Recommendation

From the foregoing discussion, the system is recommended for use by farmers who are into cultivation and production of cucumber and watermelon to help detect diseases in plants and for records keeping, because it brings professionalism into the profession, saves time, it is reliable, and more efficient. It can save cost of treatment as well as physical and psychological stress of the crop scientist.

Contribution to Knowledge

This research applied K-means algorithm and Artificial Neural Network (ANN) to provide a more efficient prediction and classification of cucumber/watermelon diseases.

Future Work

As the scope of this work is limited to four different types of disease in cucumber and watermelon, further studies can be carried out to cover a wide range of diseases in the said plant. Mobile phone applications can also be developed to improve the system ease of movement in and around the farm. There is the need to generate and store plant records electronically using the cloud system can also be studied. With this, data mining techniques could be used to discover, compose and analyze the data for more productive treatment of plant diseases by plant scientists and/or farm managers.

References

1. Al Hiary H, Bani Ahmad S, Reyalat M, Braik M, ALRahamneh Z. Fast and Accurate Detection and Classification of Plant Diseases. International Journal of Computer Applications. 2011;17(1):31-38. https://doi.org/10.5120/2183-2754
2. Barbier EB, Burgess JC. The sustainable development goals and the systems approach to sustainability. Economics; c2017. https://doi.org/10.5018/economics-ejournal.ja.2017-28
3. Chaudary P, Chaudhari AK, Cheeran AN, Godara S. Colour Transform Base Approach for Diseases Spot Detection On Plant. International Journal of Computer Science and Telecommunications. 2012;3(6):4-9.
4. Dabbura I. K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks; c2018. https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a
5. Foraminifera MR. Agricultural Production In Nigeria; The Opportunities; c2016. https://foramfera.com/2016/03/04/agricultural-production-in-nigeria-the-opportunities/
6. Gill NS. Artificial Neural Networks Applications and Algorithms. Data Science; c2020. https://www.xenonstack.com/blog/artificial-neural-network-applications
7. Gokulakrishnan KR. Detecting the Plant Diseases and Issues by Image Processing Technique and Broadcasting. 2014;3(5):2012-2014.
8. Golhani K, Balasundram SK, Vadamalai G, Pradhan B. A review of neural networks in plant disease detection using hyperspectral data. Information Processing in Agriculture. 2018;5(3):354-371.

https://doi.org/10.1016/j.inpa.2018.05.002
9. Guest D. The Impact of Plant Disease on Food Security; c2012. https://www.mdpi.com/journal/agriculture/special_issues/plant_disease
10. Hungilo GG, Emmanuel G, Emanuel AWR. Image Processing Techniques for Detecting and Classification of Plant Disease: A Review. Proceedings of the 2019 International Conference on Intelligent Medicine and Image Processing; c2019. p. 48-52. https://doi.org/doi.org/10.1145/3332340.3332341
11. Islam W. Plant Disease Epidemiology: Disease Triangle and Forecasting Mechanisms In Highlights. 2018;5(1):7-11. https://doi.org/10.17582/journal.hv/2018/5.1.7.11
12. Kumar V, Gupta P. Importance of Statistical Measures in Digital Image Processing. International Journal of Emerging Technology and Advanced Engineering. 2012;2(8):56-62.
13. Patil JK, Kumar R. Advances in Image Processing for Detection of Plant Disease. Journal of Advanced Bioinformatics Applications and Research. 2011;2(2):135-141. https://doi.org/10.9756/sijcsea/v5i2/05010140101
14. Patil SP, Zambre RS. Classification of Cotton Leaf Spot Disease Using Support Vector Machine. Journal of Engineering Research and Applications. 2014;4(5):92-97. www.ijera.com
15. Petrellis N. Plant disease diagnosis based on image processing, appropriate for mobile phone implementation. Proceedings of the 7th International Conference on Information and Communication Technology in Agriculture, Food and Environment; c2015. p. 238-246.
16. Rao NK, Reddy GS. Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm. International Journal of Computer Science and Information Technologies. 2012;3(3):4558-1561.
17. Ravichandran D, Nimmatoori R, Dhivakar A. A study on Image Statistics and Image Features on Coding Performance of Medical Images. International Journal of Advanced Computer Engineering and Communication Technology (IJACECT). 2016;5(1):1-6.
18. Miller R. Is Watermelon Related to Cucumber and Gourd? Home Guides; c2020. Retrieved on 2020 May 23. https://homeguides.sfgate.com/watermelon-related-cucumber-gourd-68962.html
19. Revathi P, Hemalatha M. Classification of cotton leaf spot diseases using image processing edge detection techniques. 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET); c2012. https://doi.org/10.1109/INCOSET.2012.6513900
20. Saiz-Rubio V, Rovira-Más F. From smart farming towards agriculture 5.0: A review on crop data management. Agronomy. 2020;10(2):207.
21. Sarfraz M. Introductory Chapter: On Digital Image Processing. IntechOpen. 2020;5(1):1-25. https://doi.org/10.5772/intechopen.92060
22. Shire A, Jawarkar U, Manmode M. A Review Paper on Plant Disease Detection Using Image Processing. International Journal of Innovative Science, Engineering & Technology. 2016;5(4):758-763.
23. Varshney S, Tarun D. Plant Disease Prediction using Image Processing Techniques- A Review. International Journal of Computer Science and Mobile Computing. 2016;5(5):394-398. https://www.ijcsmc.com/docs/papers/May2016/V5I5201682.pdf
24. Weizheng S, Yachun W, Zhanliang C, Hongda W. Grading Method of Leaf Spot Disease Based on Image Processing. International Conference on Computer Science and Software Engineering. 2008;6(1):491-494. https://doi.org/10.1109/CSSE.2008.1649
25. Campbell CL, Madden LV. Introduction to plant disease epidemiology; c1990.
26. Wirth MA. Texture Analysis. University of Guelph, Computing and Information Science; c2004. http://www.cyto.purdue.edu/cdroms/micro2/content/education/wirth06.pdf