**CHAPTER 1**

**INTRODUCTION**

**General Introduction:**

The human visual system has no problem interpreting the subtle variations in translucency and shading in this Figure 1.1 photograph and correctly segmenting the object from its background.

**  
Figure1.1. Lotus flower seen as to the naked eye.**

Let’s imagine a person taking a field trip, and seeing a bush or a plant on the ground, he or she would like to know whether it’s a weed or any other plant but have no idea about what kind of plant it could be. With a good digital camera and a recognition program, one could get some useful information. Plants play an important role in our environment. Without plants there will be no existence of the earth’s ecology. But in recent days, many types of plants are at the risk of extinction. To protect plants and to catalogue various types of flora diversities, a plant database is an important step towards conservation of earth’s biosphere. There are a huge number of plant species worldwide. To handle such volumes of information, development of a quick and efficient classification method has become an area of active research. In addition to the conservation aspect, recognition of plants is also necessary to utilize their medicinal properties and using them as sources of alternative energy sources like bio-fuel. There are several ways to recognize a plant, like flower, root, leaf, fruit etc.

**Significance Of The Domain Of Working:**

Since recent decades, digital image processing, image analysis and machine vision have been sharply developed, and they have become a very important part of artificial intelligence and the interface between human and machine grounded theory and applied technology. These technologies have been applied widely in industry and medicine, but rarely in realm related to agriculture or natural habitats.

Despite the importance of the subject of identifying plant diseases using digital image processing, and although this has been studied for at least 30 years, the advances achieved seem to be a little timid. Some facts lead to this conclusion:

Methods are too specific. The ideal method would be able to identify any kind of plant. Evidently, this is unfeasible given the current technological level. However, many of the methods that are being proposed not only are able to deal with only one species of plant, but those plants need to be at a certain growth stage in order to the algorithm to be effective. That is acceptable if the plant is in that specific stage, but it is very limiting otherwise. Many of the researchers do not state this kind of information explicitly, but if their training and test sets include only images of a certain growth stage, which is often the case, the validity of the results cannot be extended to other stages.

* 1. **Existing System:**

In recent times, computer vision methodologies and pattern recognition techniques have been applied towards automated procedures of plant recognition. Digital image processing is the use of the algorithms and procedures for operations such as image enhancement, image compression, image analysis, mapping, geo-referencing, etc. The influence and impact of digital images on modern society is tremendous and is considered as a critical component in variety of application areas including pattern recognition, computer vision, industrial automation and healthcare industries.

One of the most common methods in leaf feature extraction is based on morphological features of leaf. Some simple geometrical features are aspect ratio, rectangularity, convexity, sphericity, form factor etc.

One can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques. Some systems employ descriptions used by botanists. But it is not easy to extract and transfer those features to a computer automatically.

* 1. **Proposed System:**

The aim of the project is to develop a Leaf recognition program based on specific characteristics extracted from photography. Hence this presents an approach where the plant is identified based on its leaf features such as area, histogram equalization and edge detection and classification. The main purpose of this program is to use Open-CV resources.

Indeed, there are several advantages of combining Open-CV with the leaf recognition program. The result proves this method to be a simple and an efficient attempt. Future sections will discuss more on image preprocessing and acquisition which includes the image preprocessing and enhancement, histogram equalization, edge detection. Further on sections introduces texture analysis and high frequency feature extraction of a leaf images to classify leaf images i.e. parametric calculations and then followed by results.

* 1. **Motivation:**

Here is the brief review of the papers which we have referred for this project. Since digital image processing is used in this project to detect diseases in plants, it eliminates the traditional methods which are used in olden days and also it removes human error. This method needs a digital computer, mat lab software and a digital camera to detect diseases in plants. So it is a suitable method to adapt for this project. In the paper by Pallavi S. Marathe, different steps like Image acquisition, Pre processing which includes clipping, smoothing and Contrast enhancement. She has also used Segmentation techniques to partition different parts in an image. Disease detection is done by extracting features and classifying using SVM algorithm.

* 1. **Objectives:**

• To detect unhealthy region of plant leaves.

• Classification of plant leaf diseases using texture features.

• Coding is used to analyze the leaf infection.

• The analyzed information / result is sent via SMS to the farmer.

* 1. **Scope of the Project:**

Agriculture is the backbone of the Indian Economy"- said Mahatma Gandhi six decades ago. Even today, the situation is still the same, with almost the entire economy being sustained by agriculture. It contributes 16% of the overall GDP and accounts for employment of approximately 52% of the Indian population. Rapid growth in agriculture is essential not only for self-reliance but also to earn valuable foreign exchange. It is felt that with provision of timely and adequate inputs such as fertilizers, seeds, pesticides and by making available affordable agricultural credit /crop insurance, Indian farmers are going to ensure food and nutritional security to the Nation. Hence the detection of plant diseases is an important aspect in increasing the yield of a crop. By detecting the plant disease and identifying which pesticide/insecticide and the amount of pesticide/insecticide to be used increases the crop yield which leads in growth of country’s economy.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1”** **SeraworkWallelign,** **MihaiPolceanu,** **C´edricBuche”,”** **Artiﬁcial Intelligence Research Society Conference (FLAIRS-31)”,”** **LAB-STICC, ENIB”(2013)**

**2.1.1 Work Carried in this project**

Plants have become an important source of energy, and are a fundamental piece in the puzzle to solve the problem of global warming. However, plant diseases are threatening the livelihood of this important source. Convolutional neural networks (CNN) have demonstrated great performance (beating that of humans) in object recognition and image classiﬁcation problems.

**2.1.2 Methodology**

To classify soybean plant diseases a large collection of the plant’s leaf images is required. The images are downloaded from the Plant Village database .In this section the methodology followed is discussed in detail. Proper and large dataset is required for all classiﬁcation research during the training and the testing phase. The dataset for the experiment is downloaded from the Plant Village database which contains different plant leaf images and their labels. It contains a collection of images taken at different environment. A dataset containing 12,673 leaf images of four classes including healthy leaves is downloaded. The samples per class of the dataset is summarized.

**2.1.3 Merits and demerits**

The main advantage of representation learning is that algorithms automatically analyze large collections of images and identify features that can categorize images with minimum error. Recently convolutional neural networks(CNN) have been used for object recognition and image classiﬁcation.

**2.1.4 Conclusions**

In this study convolutional neural network is used to detect and classify soybean plant diseases. The Network is trained using the images taken in the natural environment and achieved 99.32% classiﬁcation ability. This shows the ability of CNN to extract important features in the natural environment which is required for plant disease classiﬁcation.

**2.2 “Y.-Y. Wan, J.-X. Du, D. S. Huang, Z. Chi, Y.-M. Cheung, X.-F.**

**Wang, and G.-J. Zhang”,(** **2004),** “**Bark texture feature extraction based on**

**statistical texture analysis”,”** **International**

**Symposium on Intelligent Multimedia, Video and Speech Processing”,”** **pages 482–485,”**

**2.2.1 Work Carried out in this paper**

We have developed a software system used for plant leaves recognition, where most of the features are extracted from the plant leaves. Concisely speaking, geometrical and morphological features are obtained. In order to unproven the recognition accuracy, we tried to add the texture of bark as a supplementary characteristic. There are more than 300 pictures of bark collected with our digital camera outdoors.

**2.2.2 Methodology**

**Co-occurrence matrices method (COMM)**

So called COMM is meant the one that the joint probability at gray levels i and ) occurs for two pixels with a defined spatial relationship in an image.

**Histogram Method (HM)**

From the viewpoint of statistics, histogram denotes regional probability density function. So its statistical measurement can be the features discriminating different kinds. Generally, the X-axis of the histogram denotes the gray scale while the Y-axis shows the arising probability of the corresponding gray value.

**Gray Level Run-Length Method (RLM)**

The run-length method (RLM) is defined based on the number of times, g (i,) I 0), that the picture contains run-length ) of gray level in the given direction o. For gray level run length matrices are computed according to four different directions .

**Auto-Correlation Method (ACM)**

An autocorrelation function can be evaluated to measure the coarseness. This function evaluates the linear spatial relationships between pmrultves.

**2.2.4 CONCLUSIONS**

In this paper, we applied four statistical texture analysis methods to bark recognition. Through a great deal of trials, it can be found that it is not enough to only gain features from gray images. So we added the color information by image conversion method to get three single channel images. As a result, the classification perfonnances of all four methods by the three classifiers were improved greatly.

**2.3” Phadikar, Santanu, Jaya Sil, and Asit Kumar Das”,”Rice diseases**

**classification using feature selection and rule generation techniques”**

**Computers and Electronics in Agriculture 90 (2013): 76-85.**

**2.3.1 Work Carried out in this paper**

Development of an automation system for classifying diseases of the infected plants is a growing research area in precision agriculture. The paper aims at classifying different types of rice diseases by extracting features from the infected regions of the rice plant images. Fermi energy based segmentation method has been proposed in the paper to isolate the infected region of the image from its background. Based on the field experts’ opinions, symptoms of the diseases are characterized using features like color, shape and position of the infected portion and extracted by developing novel algorithms

**2.3.3 Merits and Demerits**

Advantage of the proposed method is that it does not require any gain calculation of the rules and so involves lesser computational complexity. Finally ten-fold cross validation is performed to measure the efficiency of the proposed method which shows superiority over other methods.

**2.3.4 Conclusion**

An automated rice disease detection system has been developed in the paper using different types of visual properties of the symptoms created by the diseases. Fermi energy based region extraction method is applied to overcome the limitation of selecting proper threshold value. To identify the shape of the infected region, GA is applied that best approximates the structure of the region. Position of infection is determined by partitioning the spot into different blocks and arranged as a quad tree at different labels.

**2.4” Asraf, H. Muhammad, M. T. Nooritawati, and M. S. B. Rizam”,”A**

**Comparative Study in Kernel-Based Support Vector Machine of Oil Palm**

**Leaves Nutrient Disease”,”Procedia Engineering 41” (2012): 13531359.**

**2.4.1 Work Carried out in this paper**

Investigation of the nutrient disease in oil palm motivates the need for a programmed detection system. Automated detection using vision system and pattern recognition are implemented to detect the symptoms of nutrient diseases and also to classify the disease group. In this paper, Support Vector Machine (SVM) is evaluated as classifier with three different kernels namely linear kernel, polynomial kernel with soft margin and polynomial kernel with hard margin. Initial results show that the recognition of oil palm leaves is possible to be performed by SVM classifier.

**2.4.2 Methodology**

The oil palm leaves images are taken and preprocessed prior to classification according to respective diseases namely nitrogen, potassium and magnesium. Samples of leaves collected are approximately 420 samples. Out of that, 300 images are used as a training set with equal number for each disease consists of 100 nitrogen leaves, 100 magnesium leaves and 100 potassium leaves. The remaining 120 images are chosen as a testing set comprises of 40 images from nitrogen, potassium and magnesium leaves. All the images are captured at ambient environment at a fixed distance between leaves samples and digital imaging device.

**2.4.4 Conclusion**

In conclusion, SVM classification performed its task successfully. The proposed classifiers have been tested on oil palm leaves samples. Three different types of kernel function namely linear, polynomial (hard margin) and polynomial (soft margin) are used to perform classification task. Polynomial kernel with soft margin produces the best performance in average of 95% correct classification as compared to the other types of Kernel function. This has verified the need for proper kernel trick function choice which would yield for more accurate results.

**2.5”Shabanzade, Maliheh, MortezaZahedi, and Seyyed Amin Aghvami”,**

**“Combination of local descriptors and global features for leaf**

**Recognition”,” Signal and Image Processing: An International Journal”**

**(SIPIJ). v2 i3 (2011): 23-31.**

**2.5.1 Work Carried out in this paper**

Automatic leaf recognition system is a case coming to improve time-consuming and troublesome tasks which have mainly been carried out by botanists manually. This application as judged by common characteristics is popular in institutes for discovering new plant species, modernizing the management of botanical gardens and horticulture fields. In order to conduct a leaf recognition system, the features must be sufficiently distinctive to identify specific objects among many alternatives, where contain both local and global properties.

**2.5.2 Merits and demerits**

The captured images from leaves have various properties in different conditions, such as yellow color in autumn, dried in winter or variation in intensity. These conditions lead to wrong calculations of correct local descriptors. Leaf shape may be incomplete, for example if animals eat leaves then some parts are cropped and shape properties change, so global features can not support this kind of problems well. Any leaf is capable for being sick or calamity, so many holes and dots occur in an image, thus local descriptors give unexpected values and cause mistake in recognition

**2.5.3 Methodology**

The leaf picture taken in an un-controlled environment which has to be used for a recognition system has some defects making it problematic for achieving a good recognition rate. Thus, before using a raw image, it is necessary to perform some pre-processing stages where a binarized image is necessary for extraction of global features and also a special segmentation process for extraction of local descriptors.

**2.5.4 CONCLUSION** In this paper, it is shown that mixture of local descriptors and global features can be used as a feature vector for leaf recognition system where each category of the features demonstrates a different aspect of leaf properties by covering important characteristics. An important property for a leaf is veins structure where local descriptors show it with texture calculation on concurrence matrix.

**“Li-jie, Yu, Li De-sheng, and Zhou Guan-ling”,”Automatic Image Segmentation Base on Human Color Perceptions",”International Journal of Image, Graphics and Signal Processing (IJIGSP)” 1.1 (2009): 25.**

**2.6.1 Work carried out In this paper**

we propose a color image segmentation algorithm based on perceptual color vision model. First, the original image is divide into image blocks which are not overlapped; then, the mean and variance of every image back was calculated in CIEL\* a\* b\* color space, and the image blocks were divided into homogeneous color blocks and texture blocks by the variance of it. The initial seed regions are automatically selected depending on calculating the homogeneous color blocks’ color difference in CIEL\* a\* b\* color space and spatial information.

**2.6.2 Methodology**

By regarding the image segmentation as a problem of partitioning pixels into different clusters according to their color similarity and spatial relation, we propose our color image segmentation method automatically. (1) Selects seeds region for image using block-based region growing and perceptual color vision model in the CIEL\* a\* b\* color space; (2) Generates a final segmentation by utilizing an effective merging procedure using fuzzy algorithm and color-edge detection. Our procedure first partitions the original image into non-overlapping range blocks, calculate mean and variance of range blocks, sub-block in a color image will be grouped into different clusters, and each detected receive a label, with the same label is referred as a seed region grow into the higher seed regions areas.

**2.6.3 Merits and Demerits**

There are mainly two disadvantages in our algorithm. First, although using the fixed threshold values can produce reasonably good results, it may not generate the best results for all the images. Second, when an image is highly color textured (i.e. there are numerous tiny objects with mixed colors), our algorithm may fail to obtain satisfied results because the mean value and variance could not represent the property of the region well.

**2.6.5 CONCLUSION**

This work presents a computationally efficient method designed for automatic segmentation of color images with varied complexities. Firstly, the original image is divide into rectangular image blocks which are not overlapped; then, the mean and variance of each image black was calculated in CIEL\* a\* b\* color space, and the image blocks were divided into homogeneous color blocks and texture blocks by the variance of it. The Initial seed regions are automatically selected depending on calculating the homogeneous color blocks’ color difference in CIEL\* a\* b\* color space and spatial and adjacent information.

**2.7 ”Camargo, A. and J. S. Smith”,” Image pattern classification for the identification of disease causing agents in plants”,” Com. Elect. Agr. 66: 121–125” (2009)**

**2.7.1 Work carried in this project**

This study reports a machine vision system for the identification of the visual symptoms of plant diseases, from coloured images. Diseased regions shown in digital pictures of cotton crops were enhanced, segmented, and a set of features were extracted from each of them. Features were then used as inputs to a Support Vector Machine (SVM) classifier and tests were performed to identify the best classification model. We hypothesised that given the characteristics of the images, there should be a subset of features more informative of the image domain.

**2.7.2 Methodology**

This implements a machine vision system for the classification of the visual symptoms of plant diseases, from the analysis of coloured images. Diseased regions such as spots, stains or strikes were identified, segmented, pre-processed, and a set of image features was extracted from each region. Feature selection was then performed to identify which of these provided most information about the image domain. Those features that added little or no information were discarded.

**2.7.4 Merits and demerits**

The advantage of image classification by feature assessment is that patterns remain identical if preliminary conditions are changed , pictures might be captured in conditions of different light intensity, or with different distances between camera lens and target.

**2.7.5 Conclusion**

The identification of the symptoms of plant diseases by means of a machine vision system may support farmers during their daily struggle against disease outbreaks. Here we used digital images of crop plants that showed visual symptoms of a particular disease. These diseased regions were identified and segmented with the help of an algorithm as previously reported. Features were extracted from each segmented region and used as inputs to a classifier.

**2.8 A. Camargoa, J.S. Smithb An image-processing based algorithm to automatically identify plant disease visual symptoms AE––Automation and Emerging Technologies Biosyst. Eng., vol. 102, no. 1, pp. 9–21**

**2.8.1 Work carried out in this project**

s an image-processing based method that identifies the visual symptoms of plant diseases, from an analysis of coloured images. The processing algorithm developed starts by converting the RGB image of the diseased plant or leaf, into the H, I3a and I3b colour transformations. The I3a and I3b transformations are developed from a modification of the original I1I2I3 colour transformation to meet the requirements of the plant disease data set. The transformed image is then segmented by analysing the distribution of intensities in a histogram.

**2.8.2 Methodology**

This study reports on an algorithm for the detection of visual symptoms of disease by the analysis of coloured images. The algorithm was divided into four stages: (1) image pre-processing: to specify a suitable colour transformation that best highlighted the diseased regions shown in the picture; (2) image enhancement: to develop a filter that could highlight those regions considered targets (possible diseased regions); (3) image segmentation: to identify regions in the image that were likely to qualify as diseased region; (4) image post-processing: to remove unwanted background regions.

**2.8.3 Merits and demerits**

The advantage in using a ‘computerised process’ is the high level of accuracy in the diagnosis and prognosis, as well as the significant reduction of costs in comparison to the traditional method of face-to-face diagnosis. In agriculture, numerous image-processing based computerised tools have been developed to help farmers to monitor the proper growth of their crops. Special attention has been put towards the latest stages of growth, that is, when the crop is near harvesting.

**2.8.4 Conclusions :**The strategy for image segmentation described effectively detects and selects the regions that, from the point of view of a human expert, are considered diseased. The strength of this algorithm is the ability to identify the correct target (diseased region) in images with different range of intensities distribution. To justify this statement, the algorithm was tested on a very diverse set of images and segmentation performance was estimated.

**2.9 “Sindhuja Sankarana, Ashish Mishraa, Reza Ehsani , Cristina Davis “,”A review of advanced techniques for detecting plant diseases”,” Computers and Electronics in Agriculture, vol. 72, no. 1. pp. 1–13”(2010)**

**2.9.1 Work Carried in this project**

Diseases in plants cause major production and economic losses in agricultural industry worldwide. Monitoring of health and detection of diseases in plants and trees is critical for sustainable agriculture. To the best of our knowledge, there is no sensor commercially available for real-time assessment of health conditions in trees. Currently, scouting is most widely used mechanism for monitoring stress in trees, which is an expensive, labor-intensive, and time-consuming process.

**2.9.2 Methodology**

Plant diseases cause major production and economic losses in agriculture and forestry. For example, soybean rust (a fungal disease in soybeans) has caused a significant economic loss and just by removing 20% of the infection, the farmers may benefit with an approximately 11 million-dollar profit (Roberts et al., 2006). It is estimated that the crop losses due to plant pathogens in United Stated result in about 33 billion dollars every year. Of this, about 65% (21 billion dollars) could be attributed to non-native plant pathogens (Pimentel et al., 2005). Some of the diseases caused by introduced pathogenic species are chestnut blight fungus, Dutch elm disease, and huanglongbing citrus disease.

**2.9.3 Merits and demerits**

Two indices/figures of merit were used to assess the difference between healthy and citrus canker-infected leaves. In spite of having a clear discrimination between the healthy and citrus canker-infected leaves, the citrus canker-affected leaves showed similar fluorescence output as that of chlorosis-affected leaves under field condition. The environmental conditions affect the spectral reflectance from the object (Griffin and Burke, 2003). Therefore, there is a need to identify suitable approach to overcome this problem.

**2.9.4 Conclusion**

The present paper reviews and summarizes some of the noninvasive techniques that have been used for plant disease detection. The two major categories for non-invasive monitoring of plant diseases are: (i) spectroscopic and imaging techniques, and (ii) volatile organic compounds profiling-based technique for recognizing plant diseases. The spectroscopic and imaging techniques include fluorescence spectroscopy, visible-IR spectroscopy, fluorescence imaging, and hyperspectral imaging.

**2.10 “Anandhakrishnan MG Joel Hanson1, Annette Joy2, Jerin Francis3”,” Plant Leaf Disease Detection using Deep Learning and Convolutional Neural Network ”,” Volume 7 Issue No.3”(2017)**

**2.10.1 Work carried in this project**

When plants and crops are affected by pests it affects the agricultural production of the country. Usually farmers or experts observe the plants with naked eye for detection and identification of disease. But this method can be time processing, expensive and inaccurate. Automatic detection using image processing techniques provide fast and accurate results. This paper is concerned with a new approach to the development of plant disease recognition model, based on leaf image classification, by the use of deep convolutional networks. Advances in computer vision present an opportunity to expand and enhance the practice of precise plant protection and extend the market of computer vision applications in the field of precision agriculture.

**2.10.2 Methodology**

The entire procedure of developing the model for plant disease recognition using deep CNN is described further in detail. The complete process is divided into several necessary stages in subsections below, starting with gathering images for classification process using deep neural networks. Appropriate datasets are required at all stages of object recognition research, starting from training phase to evaluating the performance of recognition algorithms. All the images collected for the dataset were downloaded from the Internet, searched by disease and plant name on various sources .Images in the dataset were grouped into different classes which represented plant diseases which could be visually determined from leaves.

**2.10.3 Merits and demerits**

One major advantage of convolutional networks is the use of shared weight in convolutional layers, which means that the same filter (weights bank) is used for each pixel in the layer; this both reduces memory footprint and improves performance.

**2.10.4 Conclusion**

The results presented in this section are related to training with the whole database containing both original and augmented images. As it is known that convolutional networks are able to learn features when trained on larger datasets, results achieved when trained with only original images will not be explored. After fine-tuning the parameters of the network, an overall accuracy of 96.3% was achieved, after the 100th training iteration (95.8% without fine-tuning).

**CHAPTER 3**

**SYSTEM REQUIREMENT SPECIFICATION**

**3.1 System Analysis:**

Image acquisition: The diseased leaf image is acquired using the camera; the image is acquired from a certain uniform distance with sufficient lighting for learning and classification. The sample images of the diseased leaves are collected and are used in training the system.

Image pre-processing: Image acquired using the digital camera is pre-processed using the noise removal with averaging filter, color transformation and histogram equalization. The color transformation step converts the RGB image to HSI (Hue, Saturation and intensity) representation as this color space is based on human perception.

Masking green pixels: Since most of the green colored pixels refer to the healthy leaf and it does not add any value to the disease identification techniques, the green pixels of the leaf are removed by a certain masking technique, this method significantly reduces processing time. The masking of green pixels is achieved by computing the intensity value of the green pixels, if the intensity is less than a predefined threshold value, RGB component of that particular pixel is assigned with a value of zero.

Segmentation: There are different image segmentation techniques like threshold based, edge based, cluster based and neural network based. One of the most efficient methods is the clustering method which again has multiple subtypes, k-means clustering, Fuzzy C-means clustering, subtractive clustering method etc. One of most used clustering algorithm is k-means clustering. K-means clustering is simple and computationally faster than other clustering techniques and it also works for large number of variables. But it produces different cluster result for different number of number of cluster and different initial centroid values. In step 5, the green pixels are detected. Then masking of green pixels is done as: if the green color value of pixel is less than the threshold value which we already have calculated, then the red, green and blue components values of the these pixel are made zero. This is done because these are the unaffected part. That is why there values are made zero which results in reduction in calculations as well. Additionally, the time consumed by the raspberry pi3 for showing the final output will greatly minimize.

In step 6 the pixels having zero value for red, green and blue and the pixels on the edge of the infected clusters are removed completely. Phase 2contains step five and step number six and this phase gives added clarity in the classifying of that disease. These results with good detection and performance, also generally required computing time should be decreased to its minimum value.

In step number seven, the infected cluster is converted from RGB form to HSI format. After that, the SGDM matrices are created for every pixel of the image. But this is done for only for H and S images and not for the I images. The SGDM [1] actually measures the probability that a given pixel at one particular gray level will occur at a different distance and angle of orientation from other pixel; however pixel has a second particular gray level for it. From the SGDM matrices, generation of texture statistics for each and every image is done.

Concisely, the features are calculated for the pixels present inside the edge of the infected part of the leaf. That means the part which is not affected inside the boundary of infected part gets uninvolved. Steps seven to ten come under phase three. In this phase the features related to texture for the objects being segmented are computed.

Finally, the recognition process in the fourth phase was performed. For each image we have captured the steps in the algorithm are repeated each time. After this the result are transferred to GSM module. Using Raspberry Pi the result is sent as e-mail, and also is displayed on monitor.

Feature Extraction: From the input images, the features are to be extracted. To do so instead of choosing the whole set of pixels we can choose only which are necessary and sufficient to describe the whole of the segment. The segmented image is first selected by manual interference. The affected area of the image can be found from calculating the area connecting the components. First, the connected components with 6 neighborhood pixels are found. Later the basic region properties of the input binary image are found. The interest here is only with the area.

Classification using SVM:A support vector machine comes under supervised learning model in the machine learning. SVM’s are mainly used for classification and regression analysis. SVM has to be associated with learning algorithm to produce an output. SVM has given better performance for classifications and regressions as compare to other processes.

There are sets of training which belong to two different categories. The SVM training algorithm creates a model that allots new examples into one category or into the other category, which makes it non-probabilistic binary linear classifier.

**3.2 Functional Requirement:**

Functional requirements define the internal workings of the software: that is, the technical details, data manipulation and processing and other specific functionality that show how the use cases are to be satisfied. They are supported by non-functional requirements, which impose constraints on the design or implementation.

**3.3 Non-Functional Requirement:**

Non-functional requirements are requirements which specify criteria that can be used to judge the operation of a system, rather than specific behaviors. This should be contrasted with functional requirements that specify specific behavior or functions. Typical non-functional requirements are reliability, scalability, and cost. Non-functional requirements are often called the ilities of a system. Other terms for non-functional requirements are "constraints", "quality attributes" and "quality of service requirements".

**Reliability:** If any exceptions occur during the execution of the software it should be caught and thereby prevent the system from crashing.

**Scalability:** The system should be developed in such a way that new modules and functionalities can be added, thereby facilitating system evolution.

**Cost:** The cost should be low because a free availability of software package.

**3.4 Tools and Technology Details:**

**OpenCV :**

It is an Open-Source library of programming functions mainly aimed at real-time computer vision.It is written in C++. The name is an acronym for “Open Source Computer Vision”. It was originally developed by intel in 1999 and later supported by Willow Garage and Itseez. The library is cross-platform and free for use under the open-source BSD license. It also supports supports the deep learning frameworks TensorFlow, Torch/PyTorch and Caffe. The various applications of OpenCV include : facial recognition, gesture recognition, human-computer interaction, mobile robotics, segmentation, etc. It also includes a statistical machine learning library that contains : boosting, decision tree making, KNN algorithm, artificial neural networks, random forest, support vector machines(SVM) and Deep Neural Networks(DNN).

**Advantages:**

* First and foremost, OpenCV is available **free** of cost
* Since OpenCV library is written in C/C++ it is quite **fast**
* **Low RAM usage** (approx 60–70 mb)
* It is **portable** as OpenCV can run on any device that can run C

**Scopes**

Object (human and non-human) detection in both commercial as well as governmental space is huge and already happens is many ways.

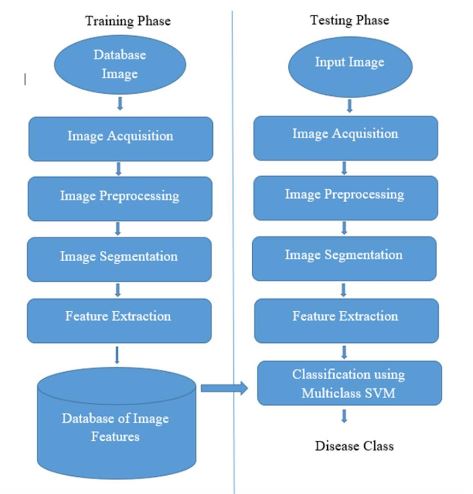
* **Transportation** - with autonomous driving in ADAS (automated driver assist system) in traffic signs detection, pedestrian detection, safety features such driver fatigue detection etc.
* **Medical imaging**- mammography, cardiovascular and microscopic image analysis (I’m not a medicine guy but I am hearing that a whole lot of computer imaging aided decision-making such as automated detection and counting of microorganisms will involve use of opencv)
* **Manufacturing** - Ton of computer vision stuff there as well such as rotation invariant detection on a conveyer belt with dtection of stoke of robotic gripping.
* **Public order and security** - pedestrain/citizen detection and tracking, mob management, prediction of future events.

**CHAPTER 4**

**SYSTEM DESIGN**

**4.1 Overall System Architecture:**

In order to build a machine leaning model it consists of two phase namely testing and training phase were the model is first trained and an input is given to test the model which is called the test data. The model consists of several image processing steps such as image acquisition, image pre-processing, segmentation, feature extraction and SVM classifier to classify the diseases.

****

**Figure2.6 Architecture Diagram**

**Image acquisition:** The diseased leaf image is acquired using the camera; the image is acquired from a certain uniform distance with sufficient lighting for learning and classification. The sample images of the diseased leaves are collected and are used in training the system. To train and to test the system, diseased leaf images and fewer healthy images are taken. The images will be stored in some standard format. The image background should provide a proper contrast to the leaf color. Leaf disease dataset is prepared with both black and white background, based on the comparative study black background image provides better results and hence it is used for the disease identification leaf.

**Image pre-processing:** Image acquired using the digital camera is pre-processed using the noise removal with averaging filter, color transformation and histogram equalization. The color transformation step converts the RGB image to HSI (Hue, Saturation and intensity) representation as this color space is based on human perception. Hue refers to the dominant color attribute in the same way as perceived by a human observer.Intensity refers to the amplitude of light. After the RGB to HSI conversion, Hue part of the image is considered for the analysis as this provides only the required information. S and I component are ignored as it does not give any significant information.

**Masking green pixels**: Since most of the green colored pixels refer to the healthy leaf and it does not add any value to the disease identification techniques, the green pixels of the leaf are removed by a certain masking technique, this method significantly reduces processing time. The masking of green pixels is achieved by computing the intensity value of the green pixels, if the intensity is less than a predefined threshold value, RGB component of that particular pixel is assigned with a value of zero.

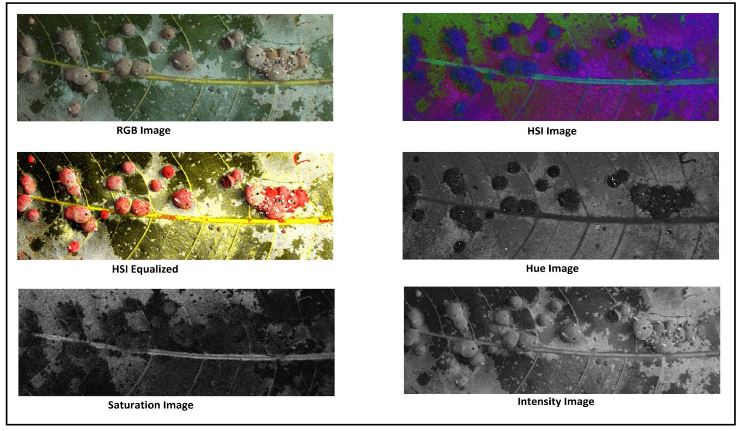
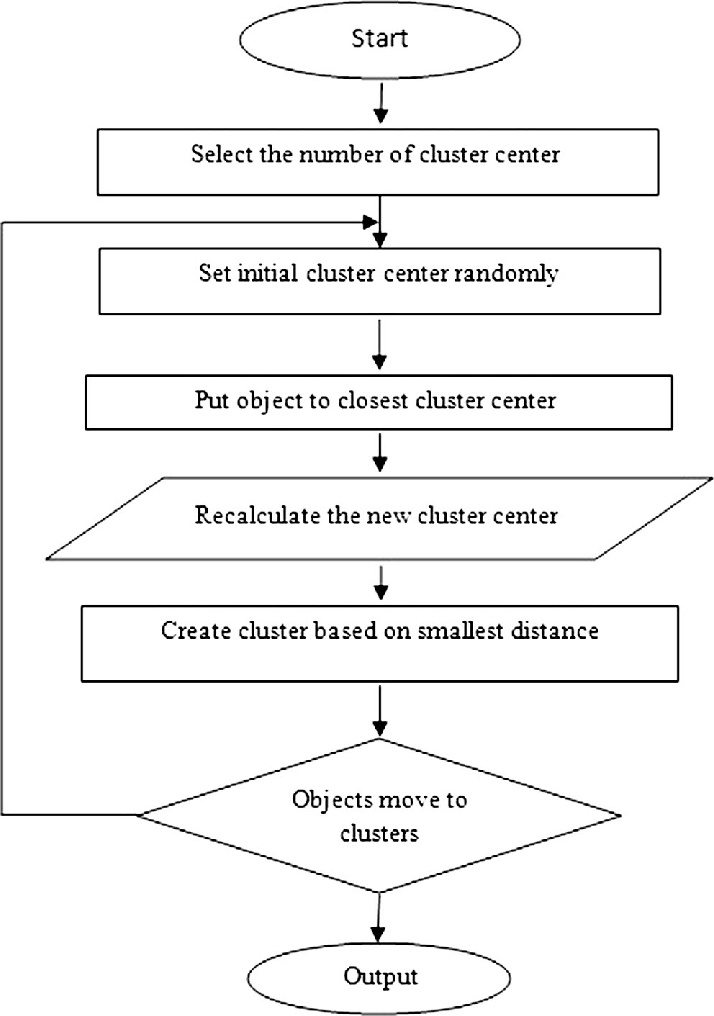
****

Fig3: RGB to HIS

**Segmentation:** There are different image segmentation techniques like threshold based, edge based, cluster based and neural network based. One of the most efficient methods is the clustering method which again has multiple subtypes, k-means clustering, Fuzzy C-means clustering, subtractive clustering method etc. One of most used clustering algorithm is k-means clustering. K-means clustering is simple and computationally faster than other clustering techniques and it also works for large number of variables. But it produces different cluster result for different number of number of cluster and different initial centroid values.



**Fig4. K-means Algorithm**

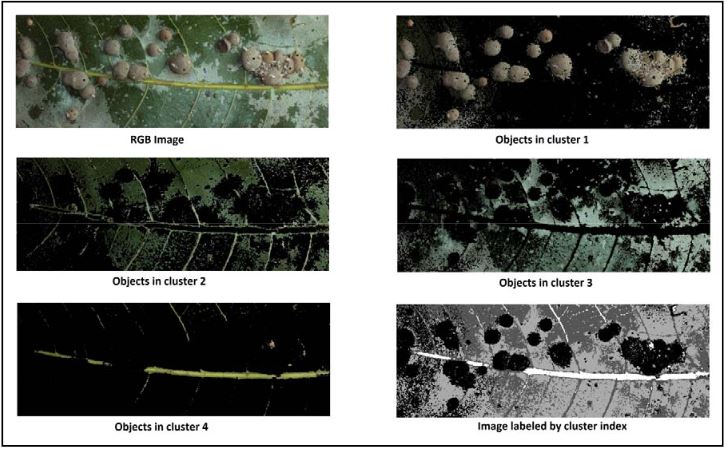
In step 5, the green pixels are detected. Then masking of green pixels is done as: if the green color value of pixel is less than the threshold value which we already have calculated, then the red, green and blue components values of the these pixel are made zero. This is done because these are the unaffected part. That is why there values are made zero which results in reduction in calculations as well. Additionally, the time consumed by the raspberry pi3 for showing the final output will greatly minimize.In step 6 the pixels having zero value for red, green and blue and the pixels on the edge of the infected clusters are removed completely. Phase 2contains step five and step number six and this phase gives added clarity in the classifying of that disease. These results with good detection and performance, also generally required computing time should be decreased to its minimum value.

In step number seven, the infected cluster is converted from RGB form to HSI format. After that, the SGDM matrices are created for every pixel of the image. But this is done for only for H and S images and not for the I images. The SGDM [1] actually measures the probability that a given pixel at one particular gray level will occur at a different distance and angle of orientation from other pixel; however pixel has a second particular gray level for it. From the SGDM matrices, generation of texture statistics for each and every image is done.

Concisely, the features are calculated for the pixels present inside the edge of the infected part of the leaf. That means the part which is not affected inside the boundary of infected part gets uninvolved. Steps seven to ten come under phase three. In this phase the features related to texture for the objects being segmented are computed.

Finally, the recognition process in the fourth phase was performed. For each image we have captured the steps in the algorithm are repeated each time. After this the result are transferred to GSM module. Using Raspberry Pi the result is sent as e-mail, and also is displayed on monitor.

**Feature Extraction**: From the input images, the features are to be extracted. To do so instead of choosing the whole set of pixels we can choose only which are necessary and sufficient to describe the whole of the segment. The segmented image is first selected by manual interference. The affected area of the image can be found from calculating the area connecting the components. First, the connected components with 6 neighborhood pixels are found. Later the basic region properties of the input binary image are found. The interest here is only with the area.



The affected area is found out. The percent area covered in this segment says about the quality of the result. The histogram of an entity or image provides information about the frequency of occurrence of certain value in the whole of the data/image. It is an important tool for frequency analysis. The co-occurrence takes this analysis to next level wherein the intensity occurrences of two pixels together are noted in the matrix, making the co-occurrence a tremendous tool for analysis

**Classification using SVM:** A support vector machine comes under supervised learning model in the machine learning. SVM’s are mainly used for classification and regression analysis. SVM has to be associated with learning algorithm to produce an output. SVM has given better performance for classifications and regressions as compare to other processes.

There are sets of training which belong to two different categories. The SVM training algorithm creates a model that allots new examples into one category or into the other category, which makes it non-probabilistic binary linear classifier.

**4.2 Input / Output Design:**

**Screens : 2**

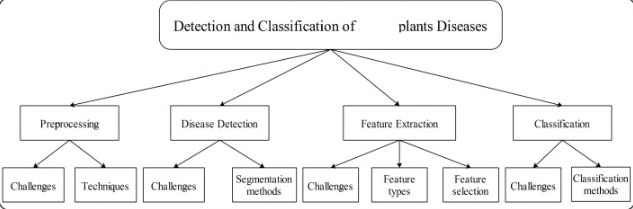
# **Python GUI – tkinter**

Python offers multiple options for developing GUI (Graphical User Interface). Out of all the GUI methods, tkinter is most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with tkinter outputs the fastest and easiest way to create the GUI applications. Creating a GUI using tkinter is an easy task.  
**To create a tkinter:**

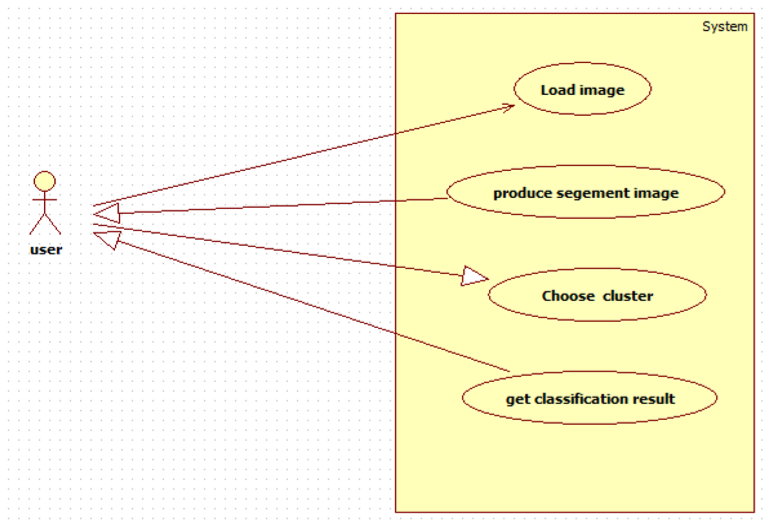
1. Importing the module – tkinter
2. Create the main window (container)
3. Add any number of widgets to the main window
4. Apply the event Trigger on the widgets.

**4.3 Object Oriented Design:**

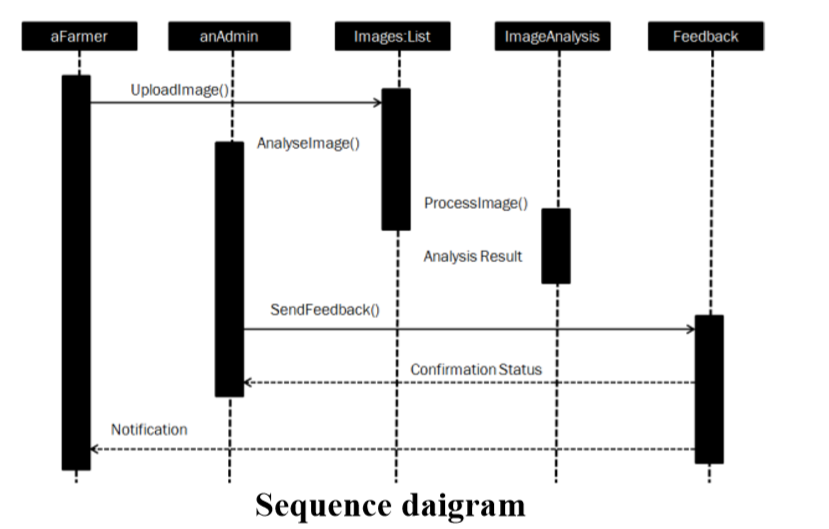
**Class Diagram:**

****

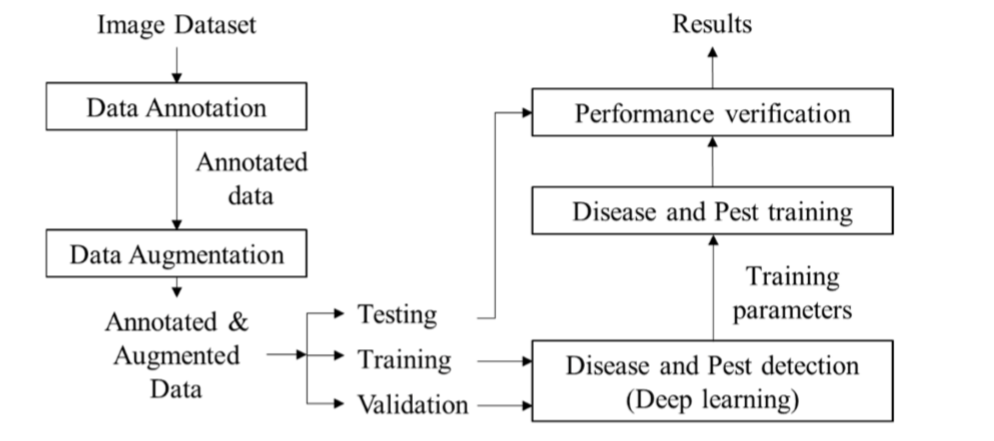
**Use Case Diagram:**

****

**Sequence Diagram:**

****

**Activity Diagram:**



**Data Flow Diagram:**

Haralick Texture

Hu Moments

Color Histogram

Feature Extraction

Processor

Bacterial Blight,

Fusarium Wilt,

Grey Mildew and Leaf curl

Based on this labelled data

SVM classifier classifies the disease with the extracted features of the particular disease

SVM Classification

Processor

Mean Shift Algorithm

Initialize the Mean shift vector.

Move the density estimation window by m(x)

Compute the Mean shift vector after shift that is m(x).

Continue until convergence.

Image Segmentation

Processor

Image Processing

RGB to Gray Conversion

RGB to HSI

Image Acquisition

Processor

**4.4 Algorithm:**

1. Capture the image in RGB format.
2. Generate color transformation structure.
3. Convert color values from RGB to the space specified in that structure.
4. Apply K means clustering for image segmentation.
5. Masking of green pixels (masking green channel).
6. Eliminate the masked cells present inside the edges of the infected cluster.
7. Convert the infected cluster from RGB to HIS.
8. Generation of SGDM matrix for H and S.
9. Calling GLCM function in order to calculate the features of it.
10. Computation of texture statistics
11. Configure KNN (classifier) for recognition.

**4.4.1 Mathematical Expression:**

**Texture Feature Extraction**

Gray Level Co-occurrence Matrix (GLCM) extract second order statistical texture features. Texture feature extraction used in this research is contrast, correlation, energy and homogeneity. This feature taken from research [3] to extract texture feature in leaf diseased region.

Contrast of the pixel and its neighbors is calculated over all of the image pixels. Contrast is used to measure contrast between neighborhood pixels.



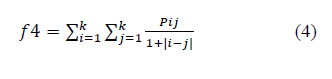
Correlation is a measure of correlation of a pixel with its neighbors over all of the images.



Energy is a sum of G (grey level co-occurrence matrix elements.



Homogeneity computes similarity of G to the diagonal matrix.

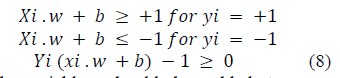


All of the four features[10]described in this section represent texture of the images of diseased region in comparison with the normal one.

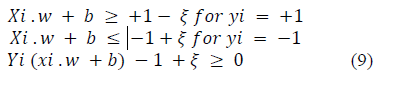
**Support Vector Machine**

vector machine is separable by a hyper plane. This hyper plane is computed according to the decision function f(x) = sign (w, x) + b, where w is a weight vector and b is a threshold cut-off.

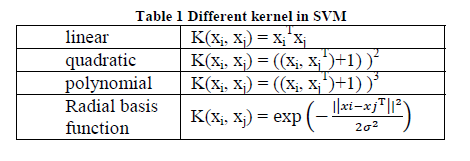
To maximize the margin, w € f and b have to be minimized to:



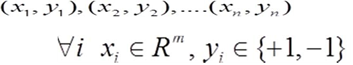
Additional slack variables should be added to prevent over fitting.



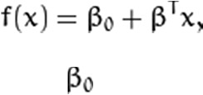
SVM was chosen as the binary classifier because it can classify accurately even when limit samples were available[6]. They are linear, quadratic, polynomial, and radial basis function.



We are given a set of n points (vectors): x1, x2, x3, …..xn such that xi is a vector of length m, and each belong to one of two classes we label them by “+1” and “-1”. So our training set is

 (10)

We want to find a separating hyper plane that separates these points into the two classes. “The positives” (class “+1”) and “The negatives” (class “-1”). Let’s introduce the notation used to define formally a hyper plane:

 (11)

Where, β is known as the *weight vector* and βo as the *bias*. For a linearly separable set of 2Dpoints which belong to one of two classes, find a separating straight line.

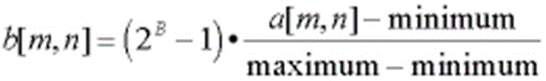
The optimal hyper plane can be represented in an infinite number of different ways by scaling of β and βo. As a matter of convention, among all the possible representations of the hyper plane, the one chosen is

 (12)

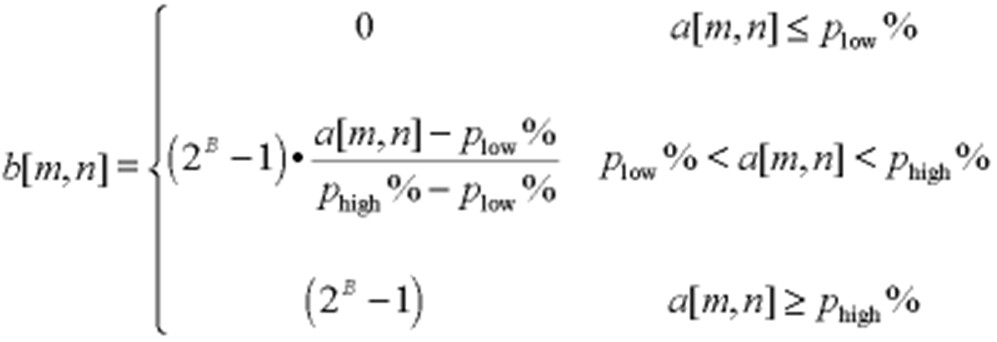
### Histogram stretching is used to enhance the contrast

Contrast is the difference between maximum and minimum pixel intensity.The most important examples are described below.

Frequently, an image is scanned in such a way that the resulting brightness values do not make full use of the available dynamic range. If the image is intended to go from brightness 0 to brightness 2B-1, then one generally maps the 0% value (or *minimum* as defined) to the value 0 and the 100% value (or *maximum*) to the value 2B-1. The appropriate transformation is given by:

 (13)

This formula, however, can be somewhat sensitive to outliers and a less sensitive and more general version is given by:

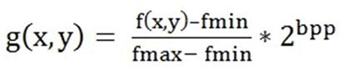
(14)

**4.4.2 Formula:**

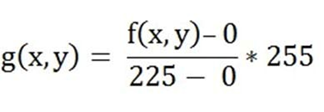
**Contrast stretching**

Contrast = 225.

The formula for stretching the histogram of the image to increase the contrast is

(15)

Where f(x,y) denotes the value of each pixel intensity. For each f(x,y) in an image, we will calculate this formula. After doing this, we will be able to enhance our contrast. The The formula requires finding the minimum and maximum pixel intensity multiply by levels of gray. In our case the image is 8bpp, so levels of gray are 256. The minimum value is 0 and the maximum value is 225. So the formula in our case is

 (16)

**CHAPTER 5**

**SYSTEM IMPLEMENTATION**

**5.1 Modules:**

1. Image Acquisition
2. Image Preprocessing
3. Segmentation
4. Feature Extraction
5. Classification using SVM

**5.2 Module Description:**

**Image acquisition:** The diseased leaf image is acquired using the camera; the image is acquired from a certain uniform distance with sufficient lighting for learning and classification. The sample images of the diseased leaves are collected and are used in training the system. To train and to test the system, diseased leaf images and fewer healthy images are taken. The images will be stored in some standard format. The image background should provide a proper contrast to the leaf color.

**Image pre-processing:** Image acquired using the digital camera is pre-processed using the noise removal with averaging filter, color transformation and histogram equalization. The color transformation step converts the RGB image to HSI (Hue, Saturation and intensity) representation as this color space is based on human perception. Hue refers to the dominant color attribute in the same way as perceived by a human observer. Saturation refers to the amount of brightness or white light added to the hue. Intensity refers to the amplitude of light. After the RGB to HSI conversion, Hue part of the image is considered for the analysis as this provides only the required information. S and I component are ignored as it does not give any significant information.

**Segmentation:** There are different image segmentation techniques like threshold based, edge based, cluster based and neural network based. One of the most efficient methods is the clustering method which again has multiple subtypes, k-means clustering, Fuzzy C-means clustering, subtractive clustering method etc. One of most used clustering algorithm is k-means clustering. K-means clustering is simple and computationally faster than other clustering techniques and it also works for large number of variables. But it produces different cluster result for different number of number of cluster and different initial centroid values. So it is required to initialize the proper number of number of cluster k and proper initial centroid.K-means is an general purpose methods that is being used at many domains to different problems. In this project, k-means is a clustering method used to get the clusters of k numbers which matches the specified characters like to segment the leaf.

Disease detection by using k clustering method [2].The algorithm provides the necessary steps required for the image detection of the plant leaf. In the first step, generally the RGB images of all the leaves are captured using camera. In step 2 a color transformation structure’s formed, and then color space transformation is applied in step 3.These two steps are to be expected in order to perform step 4. In this step the images which we have got are processed for segmentation by using the K-Means clustering technique [2]. These four steps come under phase one, the infected objects detected and determined.

**Feature Extraction**: From the input images, the features are to be extracted. To do so instead of choosing the whole set of pixels we can choose only which are necessary and sufficient to describe the whole of the segment. The segmented image is first selected by manual interference. The affected area of the image can be found from calculating the area connecting the components. First, the connected components with 6 neighborhood pixels are found. Later the basic region properties of the input binary image are found. The interest here is only with the area.The affected area is found out. The percent area covered in this segment says about the quality of the result. The histogram of an entity or image provides information about the frequency of occurrence of certain value in the whole of the data/image. It is an important tool for frequency analysis

**Classification using SVM:** A support vector machine comes under supervised learning model in the machine learning. SVM’s are mainly used for classification and regression analysis. SVM has to be associated with learning algorithm to produce an output. SVM has given better performance for classifications and regressions as compare to other processes.

**CHAPTER 6**

**RESULTS AND DISCUSSION**

**CONCLUSION:**

This project proposed a leaf image pattern classification to identify disease in leaf with a combination of texture and color feature extraction. Initially the farmers send a digital image of the diseased leaf of a plant and processed automatically based on SVM and the results were shown. The results of this project are to find appropriate features that can identify leaf disease of certain commonly caused disease to plants. Firstly, normal and diseased images are collected and pre-processed. Then, features of shape, color and texture are extracted from these images. After that, these images are classified by support vector machine classifier. A combination of several features is used to evaluate the appropriate features to find distinctive features for identification of leaf disease. When a single feature is used, shape feature has the lowest accuracy and texture feature has the highest accuracy. A combination of texture and color feature extraction results highest classification accuracy. A combination of texture and color feature extraction with polynomial kernel results in good classification accuracy. Based on the classified type of disease a text message was sent to the user in the project.

**Future scope:**

In this project, we demonstrated only few types of diseases which were commonly caused and it can be extended for more disease in future. Here only a text message was sent to the farmer but in future a robot can be sent to spray the pesticides to the plants automatically without human interaction.

**REFERENCES:**

1. Sari, Yuita Arum, R V HariGinardi, RiyanartoSarno. "Assessment of Color Levels in Leaf Color Chart UsingSmartphone Camera with Relative Calibration". Information Systems International Conference (ISICO), 2013: 631-636.
2. Raid, Richard Neil, and J. C. Comstock. Sugarcane rust disease. University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS, 1998.
3. Camargo, A., and J. S. Smith. "Image pattern classification for the identification of disease causing agents in plants." Computers and Electronics in Agriculture 66.2 (2009): 121-125.
4. Asraf, H. Muhammad, M. T. Nooritawati, and M. S. B. Rizam. "A Comparative Study in Kernel-Based Support Vector Machine of Oil Palm Leaves Nutrient Disease." Procedia Engineering 41 (2012): 13531359.
5. Phadikar, Santanu, Jaya Sil, and Asit Kumar Das. "Rice diseases classification using feature selection and rule generation techniques." Computers and Electronics in Agriculture 90 (2013): 76-85.
6. Li, Daoliang, Wenzhu Yang, and Sile Wang. "Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine." Computers and electronics in agriculture 74.2 (2010): 274279.
7. Ginardi, R. V Hari, RiyanartoSarno, and Tri AdhiWijaya. “Sugarcane Leaf Color Classification in Sa\*b\* Color Element Composition”. 2013 International Conference on Computer, Control, Informatics and It’s Application, pp:175-178.
8. Li-jie, Yu, Li De-sheng, and Zhou Guan-ling. "Automatic Image Segmentation Base on Human Color Perceptions." International Journal of Image, Graphics and Signal Processing (IJIGSP) 1.1 (2009): 25.
9. Shivakumar, G, P.A Vijaya. "Face Recognition Using Geometric Attributes". International Journal of Computational Intelligence Research Volume 6, Number 3 (2010), pp. 373–383.
10. Shabanzade, Maliheh, MortezaZahedi, and Seyyed Amin Aghvami. "Combination of local descriptors and global features for leaf recognition." Signal and Image Processing: An International Journal (SIPIJ). v2 i3 (2011): 23-31.
11. Anandhakrishnan MG Joel Hanson1, Annette Joy2, Jerin Francis3 Department of Computer Science Engineering SCET, India