CLASSIFICATION OF *CITRUS* (RUTACEAE) BY USING IMAGE PROCESSING

By

NAJWA BARI'AH BINTI MOHD TABRI

A report submitted in fulfillment of the requirements for the degree of Bachelor of Applied Science (Natural Resources Science) with Honours

FACULTY OF EARTH SCIENCE UNIVERSITI MALAYSIA KELANTAN

2019

DECLARATION

I declare that this thesis entitled "Classification of *Citrus* (Rutaceae) by Using Image Processing" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ____________________________

Name : ____________________________

Date : ____________________________

APPROVAL

I hereby declare that I have read this thesis entitled "Classification of *Citrus* (Rutaceae) by Using Image Processing" and in my opinion this thesis is sufficient in terms of scope and quality for the award of the degree of Bachelor of Applied Science (Natural Resources Science) with Honours.

Signature : ____________________________

Name of Supervisor : ____________________________

Date : ____________________________

ACKNOWLEDGEMENT

I am really grateful because I managed to complete my final year project and thesis report writing. Special thanks for those who helped me throughout this journey directly and indirectly. Throughout the period of two semesters completing this project, I want to take this opportunity to express my utmost gratitude to several parties as they have been supporting me in completing this task. Firstly, I want to thank the Almighty given that I was able to complete the final year project without much difficulties.

Furthermore, I would like to express my utmost grateful to my final year project supervisor, Dr. Shaparas Bt Daliman for encourage and lead me in doing this project. I would like to thank her for all the guidance and time in helping me to conduct my project. I am deeply grateful for her kindness in explaining the details of the project including the help that was given when I was facing some difficulties while was on this project.

Next, I would like to thank my parent and family who give moral supports till the day I manage to complete this project. Last but not least, I also want to express my thankfulness to my supportive friends for helping and giving me support despite of their busy schedule, they taught and shared their knowledge in completing this project.

Classification of *Citrus* (Rutaceae) by Using Image Processing

ABSTRACT

Image processing has been increasingly used for agricultural applications for crop management especially in identification of crop status, quantity and quality. The aim of this study are to identify the classification of four selected *Citrus* species which are *Citrus microcarpa* (calamondin), *Citrus aurantifolia* (common lime), *Citrus hystrix* (kaffir lime) and *Citrus maxima* (pomelo). This research will be conducted by using digital image processing approach based on the morphological features of leaf with the combination of gray level co-occurrence matrix (GLCM), Prewitt and Canny algorithm and training classification models by using support vector machine (SVM). A machine learning algorithms, SVM have been used to build species identification models. The study present how to classify selected *Citrus* genus species with similar leaf shapes based on leaf images by using digital image vision machine classification. Though the developed system is not intended to replace human taxonomists, it may provide a rapid and easily accessible technique to identify plants with acceptable accuracy. The image pixels of the *Citrus* was classified using the difference in the leaf features of the plant species. SVM models achieved satisfactory results demonstrating its usefulness in identification and classification tasks by occupied 93% of the highest overall accuracy based on the combination of GLCM and Canny and algorithm features where *C. maxima*, *C. aurantifolia* and *C. microcarpa* obtained 100% accuracy while *C. hystrix* obtained 80% of accuracy. This study provided a rapid and easily accessible technique to identify plants which are beneficial by using digital image classification.

Pengelasan *Citrus* (Rutaceae) dengan Menggunakan Pemprosesan Imej

ABSTRAK

Pemprosesan imej semakin banyak digunakan untuk aplikasi pertanian bagi pengurusan tanaman terutama dalam mengenal pasti status tanaman, kuantiti dan kualiti. Tujuan kajian ini adalah untuk mengenalpasti pengelasan empat spesies *Citrus* yang dipilih iaitu *Citrus microcarpa* (calamondin), *Citrus aurantifolia* (kapur biasa), *Citrus hystrix* (kaffir lime) dan *Citrus maxima* (pomelo). Kajian ini akan dijalankan dengan menggunakan pendekatan pemprosesan imej digital berdasarkan ciri morfologi daun dengan gabungan algoritma *gray level co-occurence matrix* (GLCM), algoritma Prewitt dan algoritma Canny dan model klasifikasi latihan menggunakan mesin vektor sokongan (SVM). Algoritma pembelajaran mesin, SVM telah digunakan untuk membina model pengesahan spesies. Kajian ini mengklasifikasikan spesies genus *Citrus* terpilih dengan bentuk daun yang serupa berdasarkan imej daun dengan menggunakan klasifikasi mesin penglihatan imej digital. Walaupun sistem yang dibangunkan tidak dimaksudkan untuk menggantikan taksonomi manusia, ia boleh memberikan teknik yang cepat dan mudah diakses untuk mengenal pasti tumbuhan dengan ketepatan yang boleh diterima. Piksel imej *Citrus* diklasifikasikan menggunakan perbezaan ciri-ciri daun spesies tumbuhan. Model SVM menunjukkan hasil yang memuaskan dalam mengenal pasti dan mengelaskan ciri-ciri daun dengan 93% ketepatan keseluruhan tertinggi berdasarkan gabungan ciri GLCM dan Canny di mana *C. maxima*, *C. aurantifolia* dan *C. microcarpa* memperoleh 100% ketepatan manakala *C. hystrix* memperoleh 80% ketepatan. Kajian ini menyediakan teknik yang cepat dan mudah diakses untuk mengenalpasti tumbuhan yang bermanfaat dengan menggunakan klasifikasi imej digital.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

2D	2 Dimensional
GLCM	Gray Level Co-Occurence Matrix
HSV	Hue, Saturation and Value
RGB	Red, Green and Blue
SVM	Support Vector Machine

LIST OF SYMBOLS

cm	Centimeter
%	Percentage
m	Meter

CHAPTER 1

INTRODUCTION

1.1 Background of Study

The family of Rutaceae which commonly known as the rue group consist of 160 genera which is more than 2,000 species that mostly are trees and woody shrubs. Generally, it separated all over the universe particularly in moist area and warm temperature. Mostly, the flowers are fragrant and flashy, and most of species have charming fragrant leaves with oil glands on the surface. Commonly part of *Citrus* genus stimulate an economical roles as an extensive food crops and some are mature as beautifying cultivated plants which consist the lemon, grapefruit, orange and lime (Chase et al., 1999).

The family is approximately connected to the Sapindaceae, Simaroubaceae, and Meliaceae, and all are mostly located into the same structure, even though some of the organization isolated that order into Rutales and Sapindales. The families Flindersiaceae and Ptaeroxylaceae frequently kept separate, but nowadays ordinarily are implanted in the Rutaceae, as the former Cneoraceae. The sub-familial systems were not been totally answered, but the subfamily Aurantioideae is well supported. The placement of assorted genera remains unclear (Chase et al., 1999). The family Rutaceae includes six subfamilies which are Subfamily Aurantioideae, Subfamily Dictyolomatoideae, Subfamily Flindersioideae, Subfamily Rutoideae, Subfamily Spathelioideae and Subfamily Toddalioideae. As for Sapindaceae, members of Rutaceae are familiar with its fruits. *Citrus* is the most valuable genus in the family (Agarwal et al., 2006). Its fruits are an influential point of vitamin C. A genus of young conifer trees from Indo-Malaysia to South China, group of *Citrus* are now cultured throughout the mild temperate and tropical ranges of the world, specifically lemon, lime, mandarin orange, citron, sweet orange, tangerine, grapefruit, and seville orange (Sawamura, 2000).

The main nomenclature is also crucial practical and can be applied or explained by well skilled botanists or taxonomists (Barrett & Rhodes, 1976). The proficiency is getting restricted and if the accessible is alike, the common in recognizing the plants would restrict only limited class of plants it expert. A researcher required herbarium to specify the perceptible to produce approximate evaluation with the sample within their reach as an accurate way to get the classification of a plant. The plants that have been preserved are called herbarium. This plant item will be placed on a drawing paper and methodically reserved under herbarium process (Chater & Arthur, 2016). The only suitable places that can establish the herbaria are in conventional botanical nursery or universities and generally unapproachable for non-agricultural exploration and thus difficult for most people to use.

With such recognizable complication and diverse adversity experienced by analyst with manual labeling, digital image identification scheme appear to propose a potential resolution. Regardless of the allegation that diverse vegetative types such as leaf criteria are not dependable sign in classifying the classes of plant, various researches applied image processing learning techniques to cultivate computerized scheme for species of plant recognition. Image processing is an operation to change an image (Chaki & Parekh, 2011) into numerical form and do some operation on it, in order to get an enhance image or to quotation any convenient input from it. Graphic that became a form of noticeable distribution in which act as an input, such as video frame or photograph and the outcome may be figure or typical features that correlate with the pixels.

Frequently, while practising previously set signal by converting the process of image processing; it considered the images as two spatial indicators. Currently, the developing machinery is in trend, with its function in various form of a business. Image processing also provide the design of the root analysis range within computer science and engineering training too (Wang et al., 2008). Frequently, most people already used image processing as a broad dimension to get the approach to computers and digital cameras. With a minimal contribution, one can easily enlarge contrast, find edges, calibrate ferocity and try a mixture of mathematical application to images (Abràmoff et al., 2004).

Even though it can be awfully effective from this approach, the moderate consumer frequently maneuver images digitally with careless, hardly compassionate the most vital law behind the plain image-processing cycle. Although this may be common to some person, it repeatedly point to an image that is automatically disgraced and make it not earn the outcome that would be achievable with some theory of the main procedure of an image-processing rule (Cardullo & Hinchcliffe, 2007).

This study recommended the use of idea concession approach to identify the classification of four selected *Citrus* species which are *Citrus microcarpa* (calamondin), *Citrus aurantiifolia* (common lime), *Citrus hystrix* (kaffir lime) and

Citrus maxima (pomelo). This research was conducted by using digital image processing approach based on the morphological features of leaf (Tzionas et al., 2005) with the combination of gray level co-occurrence matrix (GLCM), Prewitt and Canny algorithm and training classification models by using support vector machine (SVM). The outlook of this research focus only to four species with comparable leaf forms. Based on the broad statistic of breed in the genus, *Citrus* was preferred as the best candidate in practicability for developing this research with variety species to be accomplished in the forthcoming.

1.2 Problem Statement

The trees are essentially observed by their leaves. There are specific variations of trees all over in the world. Some are applicable benefit in medicine while the others are valuable cash crop. The tree recognition is very crucial in day to day life. Earlier, the classification had been designed using numerous laboratory processes. In order to classify various petals, the genetically and morphological trait were occupied. Still, the existence of spacious morphological variation through transformation between the varied leaf breed made it further involved and challenging to categorize them. Therefore manual classification as well as identification of these petals is a tiresome work.

Over the past decades, many particular was expected from the immense statistic of metamorphic qualification of the leaf which has been mastered by the bioinformatics. Currently, image processing is trendy among the advanced system. In order to classify the species, the important step that should have known is the design of the leaf and its vegetative and generative characters. From the studies, the main focus is to identify the selected species under *Citrus* which are *C. microcarpa, C. hystrix, C. aurantiifolia,* and *C. maxima* by using image processing procedure and to differentiate the species based on their leaf morphological characters (Tzionas et al., 2005).

In order to get an enhanced image and quotation of some practical message from the images, it must be operating under image processing method. One of the samples in signal procedure is by using an image as the input and the possibility outcome might be image or features or characteristics that correlate from the image. If the leaf able to successfully differentiated, then it is easier to identify the plants.

1.3 Objectives

In this study, there are two main objectives that need to be achieved as stated below:

1. To identify different features of four *Citrus* species based on leaf morphological characters.
2. To determine highest accuracy achieved in classifying the *Citrus* species based on leaf image classification algorithms of gray level co- occurrence matrix (GLCM), Canny and Prewitt algorithms.

1.4 Scope of Study

This study will focus on identifying and classifying the species of *C. microcarpa, C. hystrix, C. aurantifolia* and *C. maxima* based on its leaf morphological characters by using image processing. This procedure will accelerate the process in identifying the selected species and can give the accuracy results based on identifying the leaf of the selected species in *Citrus* genus. The study is conducted on by taking images of leaf using digital camera of smart gadget. The scope of study is only classify four selected species from *Citrus* genus based on leaf morphological characters.

The assumption is that all of the characteristics can be classified and identified by using image processing algorithm because it can differentiate based on the features of the image. This technology able to identify edge recognition, and when combine with sub-pixel processing, reliable measurement is consistently achieved. The efficiency of calculation is even managed when supervise the curved, deliberate surfaces where slight changes in brightness or colour would achieve in common machine vision systems.

1.5 Significant of Study

The procedures involved in this study give an increased accuracy, higher speed and progressive colour shade converting machinery that will accelerate the process in identifying the selected species. The system can help to check the architecture of the leaves. If all the architecture of the leaves can be able to identify, thus it can easily be able to classify the plants.

In addition, if the image processing methods are able to differentiate the leaves, it also can successfully classify the species. This is correlated with the processes involved. This classification can help the new botanist to identify the species with the aid of image processing.

This study can be upgraded into mobile application for general use of individuals. Thus it can help more community in order to detect the identification of the species easily in future.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter elaborated on the research study in *Citrus* genus including the four selected species, vegetative characteristics of the plan leaf shapes, image processing involved, importance of the image processing, algorithms and the application of MATLAB on the image processing methods.

2.2 *Citrus* Genus

There are over 116 million tons of *Citrus* fruit production in the world (FAO, 2009). Besides, *Citrus* can grow anywhere as long as the temperature is suited and an excessively valuable crop on a global supports. It is universally can be found in best ranges with suited subtropical, climates tropical and marginal subtropical (Kahn et al., 2001).

The *Citrus* genus is famous for its juice and extract throughout the earth. The *Citrus* genus associate to the sub-tribe Citrineae within the subfamily Aurantioideae of the family Rutaceae, consist of 162 species (Abkenar et al., 2004) and it developed in subtropical, warm and humid regions of the globe. The valuable centers of root in *Citrus* and linked classes can be identified in Australia, South-east Asia, the intermediary isle between Asia and Australia and paramount Africa (Swingle & Reece, 1967). In India, a likely well-off authority of *Citrus* hereditary diversity can be found in Northeast Himalayan district and mountain range of the paramount and western Himalayan section. The difference of *Citrus* genetic that this district has provided necessarily is the development of citrus and reinforce of diversity for its economic planting in India.

Besides this, two groups of valuable diversity have been divided based on the styles of wild and semi-wild species such as *C. indica*, *C. macroptera*, *and C. Medica* (Goetz, 2014). Recently in north-eastern India, the uninterrupted community of *Citrus* with different genetic in an interbreeding diversity has been detected where lead to the hypothesis that this region is the main source of assorted *Citrus* species. There are abundant of 17 species, with 52 breeds and 7 credible common composite of citrus are stated to be found in the north-eastern area of India (Bhattacharya & Dutta, 1956). In North-eastern India there is a survey on genetic possessions of *Citrus* which pointed out a development in the statistic of species are eligible to 23 along with one subspecies and 68 diversities (Malik et al., 2006).

Citrus genus provided tasty fruits to diverse population in the earth. The breed are universally dispersed in the South-East Asia, Indo-Malaysia district, China and India, however cultured globally (Manner et al., 2006). There are variation of crucial oils with assorted specific seasonings and initial effective mixture from the leaves and the fruits of the *Citrus* species that crucial to human nourishment and food consuming, which contain folic acid, vitamin C, pectin, potassium, seasonings, flavouring, and dietaries fiber. In Malaysia, the oils from the leaves and fruits are financially used as seasonings and pleasant aroma, as well as in perfumery, cooking and pharmaceutical care, exclusively for healing and cosmetic purposes (Ng et al., 2011). The identification and composition of essential oils of some *Citrus* species currently has been studied on the Malayian *Citrus* plants which stated including *C. aurantifolia*, *C. grandis*, *C. hystrix*, and *C. microcarpa*. A limonene and monoterpene hydrocarbon is the dominant constituent in the essential oils outside of the skins of these Malaysian *Citrus* species (Awang, 2007). This research target on the specific aspect of the Malaysian Citrus species based on its vegetative characters and also to identify the species by using image processing.

The advantages of a genetically diverse an interbreeding population in tree cannot be overestimate because variety genetic built the base of agriculture (Kahn et al., 2001). A deliberation of the paired and pattern of historical difference supported the identification of the plants that are beneficial of preservation because of variety or uniqueness. Morphological aspect between the first object used in the genetic material of germ cells control (Herrero et al., 1996) and, still continue to move on the panoramic direction on variety of genetic in various crops.

Morphological criteria have been the central features used for acknowledgement and explanation of plant taxonomy (Daryono et al., 2013). Content of exploration and the existence of organized heading to determine varieties of genetic which make morphological criteria become beneficial. Furthermore, morphology is relevant to other peculiar matched in taxonomic order. Features linked to leaves, plants, flowers and fruits were used by a few researchers (Tapia et al., 2005) to explain and identify the species and its hybrids.

2.2.1 *Citrus maxima* (Pumelo)

C. maxima or can be known as pumelo as in Figure 2 is a well-known fruits in Malaysia, especially in Tambun, Perak. It is indigenous to Malaysia, and other countries including Indonesia, Philippines, Thailand, Bangladesh, India, and Vietnam, and distribute universally in Malaysia on the follow of mines production. In Malaysia, *C. maxima* has various of local names, such as 'limau bali', 'limau besar', 'limau abong', 'limau betawi', 'limau bol' and 'limau jambua'. This plant is also known as 'pumelo' or 'pummelo', with a height of 5 to 15 m, and the width of this plant is 10 to 30 cm.

The leaves of pumelo are dotted, glandular, alternate, ovatl and elliptic shape with average 5 to 20 cm length and 2 to 12 cm width. The fruits are pear-shaped with a width of 10 to 30 cm and pale-yellow or greenish yellow colour (Roger, 2002). *C. maxima* is also famous for its remedial values which is it can cure ulcers, fever, arthritis, gout and kidney disorders (Singh et al., 2015). The fruits pulp and peels are used as an appetizer, stomach-tonic, and also for the treatment of inflammation and cough.

Figure 2.1: *Citrus maxima* tree

2.2.2 *Citrus hystrix* (Kaffir lime)

Citrus hystrix or also known as kaffir lime is known as ‘wild lime’ or ‘limau purut’ as in Figure 2.2. *C. hystrix* leaves and fruits are broadly useful as spices in cooking of ‘tomyam’, either red or white, and it is favourable dish in Malaysia and Thailand. The fragrant green leaves are 7.5 to 10 cm long and the height of this plant is about 3 to 5 m length. The width of pear-shaped fruits is about 5.0 to 7.5 cm with wrinkle on the surface of fruit. The fruit is dark green, and yellow when ripe (Md Othman et al., 2016).

The essential oil of *C. hystrix* is used in relaxing and cosmetic purpose and an essential ingredient of multiple types of cosmetic and beauty products (Nor, 1999). In conventional pharmaceutical, *C. hystrix* is used to treat flu, diarrhea, abdominal pains, fever and hypertension towards kid or children (Fortin et al., 2002). Additionally, the fruits are used in cooking for seasoning and also in the manufacturing of shampoo as a defoliant for cleaning the head (Koh & Ong, 1999).

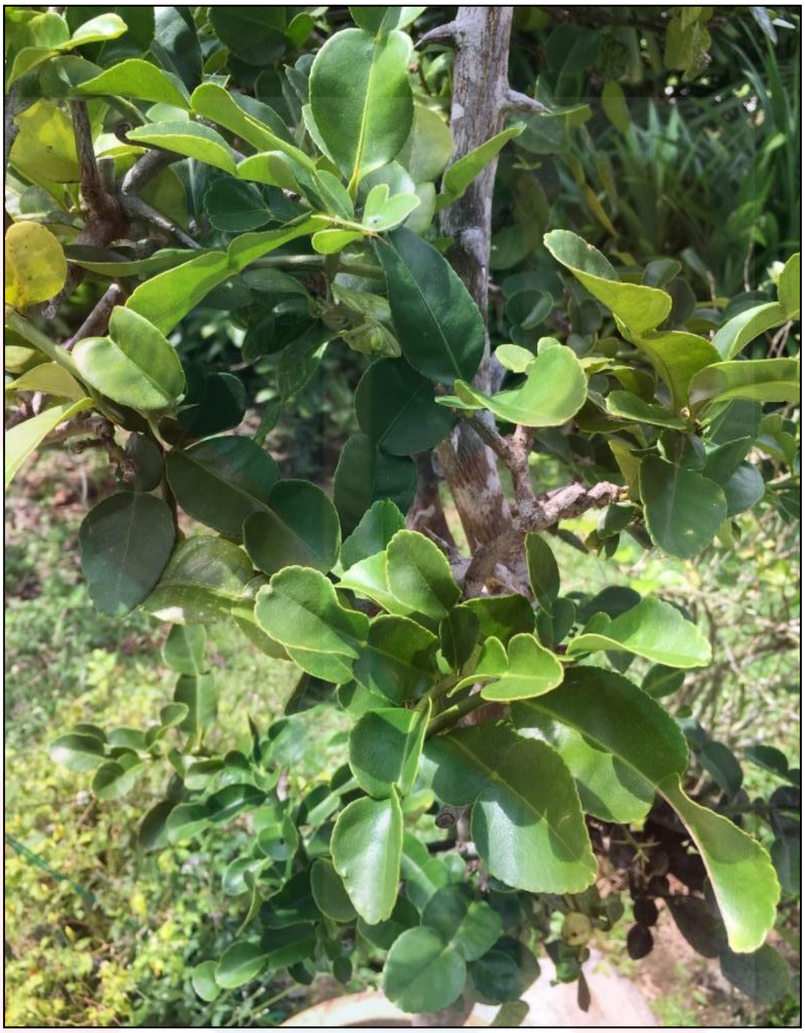

Figure 2.2: *Citrus hystrix* tree

2.2.3 *Citrus aurantifolia* (Common lime)

Citrus aurantifolia or commonly known as 'limau nipis' or 'common lime' as in Figure 2.3 are the most well-known Citrus species in Malaysia. Mostly, *C. aurantifolia* is used in food seasoning and conventional pharmaceutical (Aibinu et al., 2007). It is a pointed stalk of plant which is about 3 until 5 m length. This plant has an oval shaped which consist of 5 to 9 cm long leaves with 3 to 5 cm density. The fruits of *C. aurantifolia* are green and turn yellow after mature with a diameter of 3 to 6 cm (Md Othman et al., 2016), while the flowers are white.

Conventionally, *C. aurantifolia* is used to assist in digestion process, and to decrease fat, sugar and cholesterol in blood. The oil derived from the fruits can be used for cold, arthritis, asthma and bronchitis (Kunow, 2003). The drink also has been established to be superior cough helper when combined with honey and sugar. Additionally, it can also decrease body temperature, dismiss body smell and perform as a softener for meat. Additionally, it also has been favourable as mosquito and insect resistant. *C. aurantifolia* has been stated to have biological action as example antioxidant and anti-inflammatory.

Figure 2.3: *Citrus aurantifolia* tree

2.2.4 *Citrus microcarpa* (Calamondin)

Citrus microcarpa or can be known as calamonin, common name: 'limau kasturi' as in Figure 2.4 in Malaysia, is useful in the arrangement of beverages. *C. microcarpa* is 3 to 5 m length with bountiful of lengthy bone on the stem, branches and twigs (Manaf et al., 2008). The misty green leaves of *C. microcarpa* are between 2.5 to 6.8 cm long and 2 to 3 cm broad. The round or oblong-shaped green leaves of this plant are 2.5 to 3.8 cm in diameter. This plant is useful to thrill fever, cough, and pharyngitis (Ong, 2004).

The drink is commonly used to avoid respiratory diseases, build up the bones and act as growth catalyst for children. The drink is also usually used in cooking as seasoning element and additives. The leaves of this plant is useful in the cure of skin diseases, comfort headache and also act as a mouth wash to treat tender throat (Morton, 1987).

Figure 2.4: *Citrus microcarpa* tree

2.3 Generative Characters

2.3.1 Leaf shape

The leaves of *Citrus* are precisely combination spectrum from trifoliolate to palmately mixture to pinnately mixture of the cultured *Citrus* leaves of which characterize as unifoliolate relatively than plain, being the outcome of a contraction in number of leafs (Liu et al., 2012). Most of cultured *Citrus* leafs contain a petiole that diversely paired or not lethal leafs. Most leaves are partially irregular at the edges, even though it may differ in the intensity and sharpness.

Most of cultured *Citrus* leaves taxonomy (Duckworth et al., 2000) are moderately to actively folded when in sufficient sun. However, under dimness circumtances, leaves of many of the main groups, including blood oranges, sweet oranges, sour oranges, pummelos, and grapefruits are more or less horizontal or only very lightly folded. Only few plant leaves that are fully folded even in shade. In full sun, leaf margins tend to be somewhat undulate or wavy. Only a few cultivars exhibit leaves that are consistently wavy even in the shade as shown in Figure 2.5.

Petal wings change universally in its growth, generally alike inside specific breeds of the plants. To classify the thickness in labelling purpose, it is essential to check a total of leaves and pick the dominant pattern. In addition to petals diameter, its consecutive aspect might be advantageous to classify convinced breed or breed populations such as petal thickness, matured petal, structure of the petal wing pinnacle, amount of overlay of petal wing pinnacle and the supervise leaves (Du et al., 2007).

The smell of compress immature leaves is another way on how to tell the difference between wide breed of populations among *Citrus* (Mabberley, 1997). In addition, the measurement at a *Citrus* botanical garden or group could be convenient ahead to cropland inspection. Apart from that, essences are unique characteristics to selected breeding of the group. The aroma can also be beneficial in categorize specific hybrid with trifoliate orange which the smell of the fruit still smell the same with a non-hybrid fruit. Calamondins are different in indicate a smell which resembling a smell of fermented dough.

C. maxima *C. hystrix* *C. aurantifolia* *C. microcarpa*

Figure 2.5: Leaf shape of citrus

2.4 Image Processing

2.4.1 Function of Image Processing

Digital image processing is the application of calculating algorithms to run the image processing on logarithmic images (Chakravorty, 2018). As a section or range of process in analyzing and modifying a signal to improve and optimize the efficiency, it has various benefits towards any image processing. To avoid issues such as enhances of random colour information in the images and any alterations signals during converting, a large field of algorithms suggested to be used during input data process (Barnea & Silverman, 1972). Digital image processing could be the exemplary in involving several dimensions form, because the images are designate in two scale dimensions or more.

Edge detection and trimming process are involved in image processing as it is the main procedures in digital image processing pattern (Langner, 2006). It is become easier to apply those steps during image processing because it involved less number of facial characteristics removed and prepared. As long as there is a number of programs held under the documentation of flora diversity across the world, that's means the attempt to learn and figure out about plant diversity will remain to evolve. The documentation mostly disclose to the education authority which also involved variety statistics and images of diverse plant in the world where facts about leaves are collected and trained in order to give the latest information to the community. Recently image vision device scheme are starting to get the recognition because of the multiple benefit then regular manual analysis. This techniques also have succeed in not involving any damage in the study of leafs causing an easier classifying of type and size of leaves (Narendra & Hareesha, 2011).

Image identification covered the pattern identical algorithm scheme to classify exact sample from the recognize data collection of plant images (Wu et al., 2007). Thus, it can help in identifying the potential threatened plant bargain action. From the early 1990s, the struggle to classify plants identification from images have draw attention variety of research on various methods in image processing, facial characteristic remove and classification. There are three classification of leaf that can be identified which involved arrangement of veins, shape-based and sequences of both.

2.4.2 Importance of Image Processing

Essential and requirement of digital image deal with branch from two alter areas. The first is being the production of a view data for independent device approach, while the second is recovery of the progress of graphical acknowledgement (Haralick et al., 1987). The application of digital image processing has a wide physically dimension such as image and information statistic of stockpile for communication in work trade, remote sensing, medical imaging, sound imaging, forensic sciences and industrial machine control.

Graphic captured by satellites are benefit in detecting the earth supply, geological measuring, and assumption of cultivation product, civil community, atmospheric condition prediction, downpour and burning management (Sonka et al., 2014). Space imaging systems also detected and resolved of subjects consist of graphic collected from broad observation satellite tasks.

In addition from the explained above, various type of issue has been solved by using digital image processing. Even though irrelevant, procedure to enlarge the knowledge in human eyes detection analysis is needed in order to solve the problems. In image processing methods, it will undergo steps in dim images and degraded procedure for image enhancement and recovery (Smith & Brady, 1997). Lots of rewarding applications that related to image processing were mentioned in variety course such as medical, defense, astronomy, biology, and industrial applications. As per medical imaging is concerned most of the images may be used in the detection of tumors or for screening the patients. The current major area of application of digital image processing (DIP) techniques is in solving the problem of machine vision so as to attain good results.

2.5 Algorithms

2.5.1 Gray Level Co-Occurrence Matrix (GLCM)

A numerical technique of analyzing the character that acknowledge the study in connection graphic is the gray-level co-occurrence matrix (GLCM), also called as the gray-level calculating the space of reliance matrix. The graphic of the leafs images need to calculate how frequently combination of pixel with exact values and precise spatial connection exist in a graphic which develop a GLCM and remove numerical meas frim the matrix (Haralick & Shanmugam, 1973). Thus it can identify the features and pixels of leaves image. Also, the word co-occurrence is commonly application in the literature without a juncture, co-occurrence.

Graycomatrix indicates the GLCM by computing a graphic with gray-level which is grayscale depth value costing i develop horizontally adjoining to a graphic with the c ost j. The relationship of graphic spatial can be designed by utilizing the 'Offset' limit. Each component (i,j) in GLCM particular the number of generation that the graphic with value i occurred horizontally adjacent to a grpahic with value j (Mirzapour and Ghassemian, 2013).

The co-occurrence matrix can calculate the features of the graphic, even with our without the depth or grayscale amount of the graphic or verity range of colour, Becauses co-occurrence matrices are generally wide and sporadic, verity metrics of the matrix are usually got to get more beneficial group of characters. Component produce by utilizing this method are commonly called Haralick features, after Robert Haralick (Haralick et., 1973).

2.5.2 Canny Algorithm

Canny edge detection is a process of quotation anatomical knowledge from multiple perceptions of item and effectively cut the quantity of input to be measured (Deriche, 1987). It antiquated universally used in multiples data processing machine view scheme. Various view scheme of the edge detection are found to be related in term of fulfilling the qualification. Thus, this qualification of edge detection can be carry out in abroad scale of positions.

The common character for edge detection included the ability to precisely detect as many edges (Kumar et al., 2013). that display in the image even with a small amount of error. The middle point should be allocated when the edge point can be discovered from operator. A random variation of brightness or colour information should not detect wrong edges as the given edge in the graphic could only detect once.

Canny applied the mathematical of various methods that discover the operations which minimize likely methods in order to meet all the necessity of the operation (Canny, 1986). The excellent operation in Canny's indicator is portrayed by the total of four rapid change conditions, but it can be almost accurate by the early transmitted of a Gaussian.

Furthermore, between the edge detection operations created so far, Canny edge detection algorithm is one of the most closely give description operation that supply excellent and predictable discovered (Zhou et al., 2011). The modesty of operation for exercise and its excellent to reach with the three discovering characters make it owning this operations as it become one of the recognizable algorithms for edge detection.

2.5.3 Prewitt Algorithm

The Prewitt engineer is utilize in image processing, specifically inside the edge detection algorithms. According to the facts, it is a various distinction engineer, estimating closeness of the slope of the graphic passion operation. It is either the outcome of the Prewitt operation in the graphic is a matching slope vector or the standard of this vector (Roushdy, 2006). The Prewitt operator is based on convolute the graphic with a small, breakable, and statistic amount penetrate in vertical and horizontal guidences and as a result comparably not high price regarding performing arithmetic like Sobel and Kayyali operators. Besides, the slope closeness which it production is by comparison vulgar, in specific for high density difference in the graphic (Shrivakshan & Chandrasekar, 2012). The Prewitt operator was established by Judith M. S. Prewitt.

In easy condition, the operator estimate amount of the slope of the image passion at each point, offering the management of the biggest likely development from bright to darkened and the ratio of alteration in that control. Therefore, the result demonstrate how suddenly or smoothly the graphic alteration at that mark, and therefore how possible it is that element of the graphic show an edge, including how that edge is expected to be align. In reality, the significance arithmetic is more trustworthy and simple to understand than to calculate the vector that directed from one point to another (Yang et al., 2011).

Arithmetically, the slope of a multiple changing function is at each graphic point a 2D vector with the constituents likely by the subordinate in the horizontal and vertical directions. The slope vector marks in the alteration of biggest likely depth increase, and the magnitude of the slope vector correlate to the rate of change in that direction at each graphic point. This indicate that the outcome of the Prewitt operator at a graphic point which is in a domain of unchanged image magnitude is a zero vector and at a point on an edge is a heading which points crosswise the edge, from darker to brighter rate values (Zhao et al., 2006).

2.5.4 Classification of Support Vector Machines (SVM)

In exploring the support vector machines (SVM), which also compatible with vector networks (Cortes & Vapnik, 1995) are directed educations exemplary with correlate in study algorithms that determine information involved for classification and reversion study. SVM instruction algorithm forms a model that allowed recent pattern to one category or the other likely a set of training examples, every sample apparent as applying to one or the other of two classes which prepared it as non-possibility twofold narrow categories even though techniques such as platt scaling continued to apply SVM in a possibility categorizing application.

An SVM model is a description of the pattern as mark in scoped, outlined so that the examples of the different classification are detached by a fair division that is as expended as possible (Sharifi et al., 2002). Current objects are then outlined into that identical area and assumed that a category apply to depend on which area of the division will fall. In order to carry out a narrow classification, SVMs can accurately implement a non-aligned classification utilizing the kernel plot, implied measuring the inputs into large-spatial facial characteristics area.

When input is undescribe, inspect study is not feasible, and an un-conducted studying access is necessary, which pursuit to discover logical assemble of the information to number of individuals, and then plot new input data to that assemble groups. The support vector group (Ben-Hur et al., 2001) algorithm, developed by Vladimir Vapnik and Hava Siegelmann, used the number of support vectors, cultivate in the support vector device algorithm, to classify un-description information, and the majority universally applied device algorithms in mechanical operations (Tong & Koller, 2001).

2.6 The application of MATLAB on the study of image processing

MATLAB is an effective system of words for communication and reciprocal surroundings that allowed particular party to carry out performing arithmetic thorough work speedy compared with classic compute vocabulary such as Fortran, C, and C++. It is a bilateral; translate system that is construction for speedy mathematical matrix counting. MATLAB is a scientific programming language and provides strong mathematical and numerical support for the implementation of advanced algorithms (Hussain et al., 2011). It is for this reason that MATLAB is widely used by the image processing and computer vision community. New algorithms are very likely to be implemented first in MATLAB, indeed they may only be available in MATLAB. The MATLAB application is built around the MATLAB scripting language. Common usage of the MATLAB application involves using the Command Window as an interactive mathematical shell or executing text files containing MATLAB code.

It is a common intention compute dilect. When it is applied to methods images one in most cases creates operation folder, or writing information to carry out the function (Gonzalez at al., 2004). The data list shape a legal document of the altering applied and make sure that the last outcome can be proven and copy by others would be the demand appear. It also supplied abundant operation for image processing and additional works. Generally all of these operations are provided in the MATLAB coding and are universal legible as clear document files. Thus the application features of these operations are approachable and widely accessible to analysis. The protection system can analyze the procedure applied in full features, and any risk elevated can be answered to any knowledgeable method by a legal proceeding (Belhumeur et al., 2008).

The other benefit of MATLAB is that it acknowledge one to make secure maximum mathematical accuracy in the last outcome. In overall, graphic folder save information to 8 bit accuracy. This compare to a dimension of figure amount from 0 to 255. A smallest element of an image in a colour graphic could be depicted by three 8 bit statistic, where every bit display the green, red and blue elements as an amount number between 0 and 255. Generally this is sufficient accuracy for displaying regular graphic.

Images devotion can be lost depending on the mathematical accuracy. Some image processing algorithms shows the outcome in some pixel amount with very wide significance either it negative or positive (Schmidt, 2010) Generally a bigger statistic happen at mark in the graphic where ferocity changes come, the general point of supply of these issues is the edges of the images. Though this graphic with broadly variable amount is refresh to numbers in the range 0-255 plenty of this area should be applied just to display the few smallest element of an images with the big values. The size of the image information may then have to be displayed inside a limited dimension of figure amounts, assume from 0 to 50. Definitely this shows an abundant damage of image data. If different procedure is then used to this image the issues can then assemble.

Being a general programming language it is possible to have complete control of the precision with which one represents data in MATLAB (Sundstrom & Guzzella, 2009). An image can be read into memory and the data cast into double precision floating point values. All image processing steps can then be performed in double precision floating point arithmetic, and at no intermediate stage does one need to rescale the results to integers in the range 0 to 255.

CHAPTER 3

MATERIALS AND METHOD

3.1 Study Area

Images from leaf specimens of four *Citrus* (Rutaceae) Species: *Citrus microcarpa* (calamondin), *Citrus aurantiifolia* (common lime), *Citrus hystrix* (kaffir lime) and *Citrus maxima* (pomelo), were taken from nursery situated at Kelantan area. Total of 200 sheets of leaf samples (*Citrus maxima*, 50; *Citrus hystrix* , 50; *Citrus microcarpa*, 50; *Citrus aurantiifolia,* 50) were used in this study. The specimens were collected on July, 2018 where most of the leaf samples are being collected together. All specimens have been identified. Images of the leaf samples used in this study were captured by using Ipad mini 2 camera as shown in Table 3.1. The computer workstation used to conduct this study was Hp_PC, 8 GB RAM with a Windows 10 professional (64 bit) operating system. Image processing and features extraction were conducted using an open source image processing program, MATLAB.

Table 3.1: Specification of Camera

Features	Specification
a) Aperture	- F2.4
b) Camera Features	- Digital Zoom - Digital image stabilization - Face detection - Touch to focus
c) Image Resolution	- BSI Sensor
d) Sensor	- Yes
e) Autofocus	- High Dynamic Range mode (HDR)
f) Shooting Modes	- 1.2 MP Front camera
g) Video Recording	- 1280×720@30 fps

3.2 Materials

Images of leafs *Citrus microcarpa* plant, *Citrus aurantiifolia* plant, *Citrus hystrix* plant, *Citrus maxima* plant, paper and MATLAB software.

3.3 Method

To complete this study, a few methods has been carried out. Method include several stages which are identifying the problem, literature review, data collection, data analysis and lastly report writing. In addition, there are three crucial steps in this study of classification of plant leaf species which are data collection, data preparation and data analysis base on the study. In order to classify the plant characters, there are few steps to prepare the collecting data which would be discussed further in data preparation as in Figure 3.1.

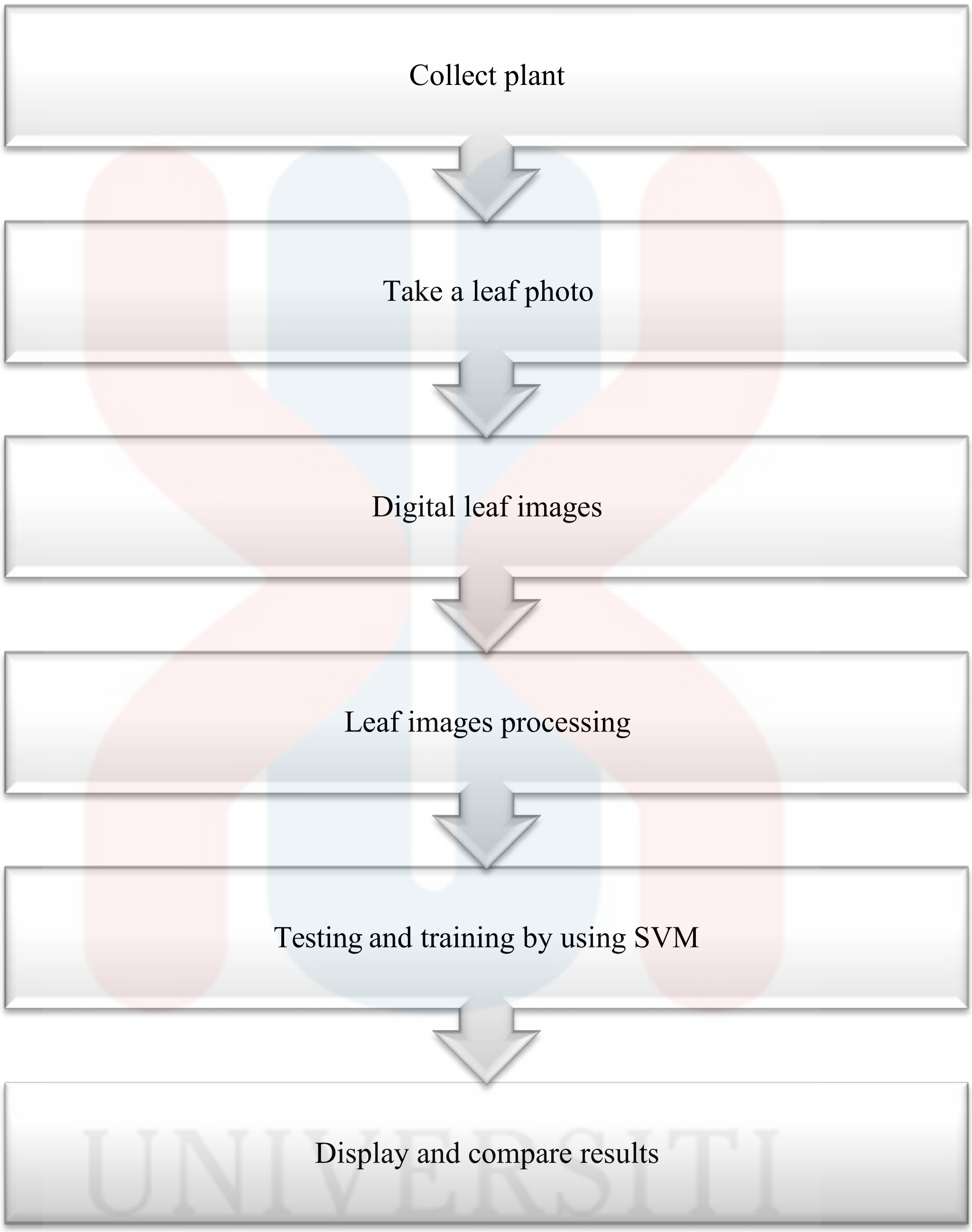

Figure 3.1: Flow diagram of the leaf image processing scheme

3.3.1 Data Collection

In this study, it involved primary sources of data collection which was taken directly by using digital leaf images. For the digital leaf images, the photo of leaf species were taken and proceed with the steps in image processing.

3.3.1.1 Primary Data Collection

The specimens used in this conducted study are image of leaves. Where possible, only undamaged and matured leaves were selected. Young leaves that are apparently small sized were ignored. Selected leaves were cropped out and saved as new images with a standard resolution. The leaves background removing steps were done by using MATLAB. Fifty images were selected for each species totaling to 200 leaf images.

3.3.2 Data Preparation

3.3.2.1 Image processing and feature extraction for plant characters

The images contain only one object, the leaf. Since all leaves are not perfectly flat, image capturing would always cast a shadow underneath the leaf. The shadow would disrupt the edge detection as it has a huge contrast with the background, confusing the algorithms to draw the boundary based on shadow instead of on the leaf. Thus, it should be removed before image segmentation. The colour threshold was defined based on the histogram settings. Firstly, the original image background was removed by using the related coding. Then, the channel with the clearest contrast between object and shadow was selected and used to identify the object boundary. The system representing the colour to be used on a computer display known as red, blue and green (RGB) combined in various proportion to obtain the colours in the visible spectrum. As hue, saturation and value (HSV) conversion alters the original color, this step serves as guidance for the subsequent edge detection of addictive colour model (RGB) value leaf images, rather than producing a final image for feature extraction shown at Figure 3.2.

Subsequent processing involved a step in converting original images to grayscale images. The frequency of occurrence of the pixel intensities was inferred by the histogram and mapped to a uniform distribution. This step was performed to improve the appearance of images in terms of image contrast. Subsequently, images were converted from grayscale to binary images. Image segmentation was performed to isolate the leaf object on images. The Canny and Prewitt edge detector algorithm, which are a powerful edge detector, is used to detect the edge of the leaf. Figure 3.3 shows some leaf images and their respective processed images.

Any colour patches, ID marking and small holes were removed by applying binary gradient mask on leaf images to stretch the leaf contents using the vertical structuring element followed by the horizontal structuring element to fill in the interior gaps inside leaves image (Hong et al., 2004). The latter was performed to remove the components inside the leaf image for the segmentation process. Then, the leave object boundaries were extracted by removing any deformity within the leaf. Outline and displayed the complete leaf in white patch. The binary images removed any deformity within the leaf outline and displayed the complete leaf in white patch.

Original images Background removes images

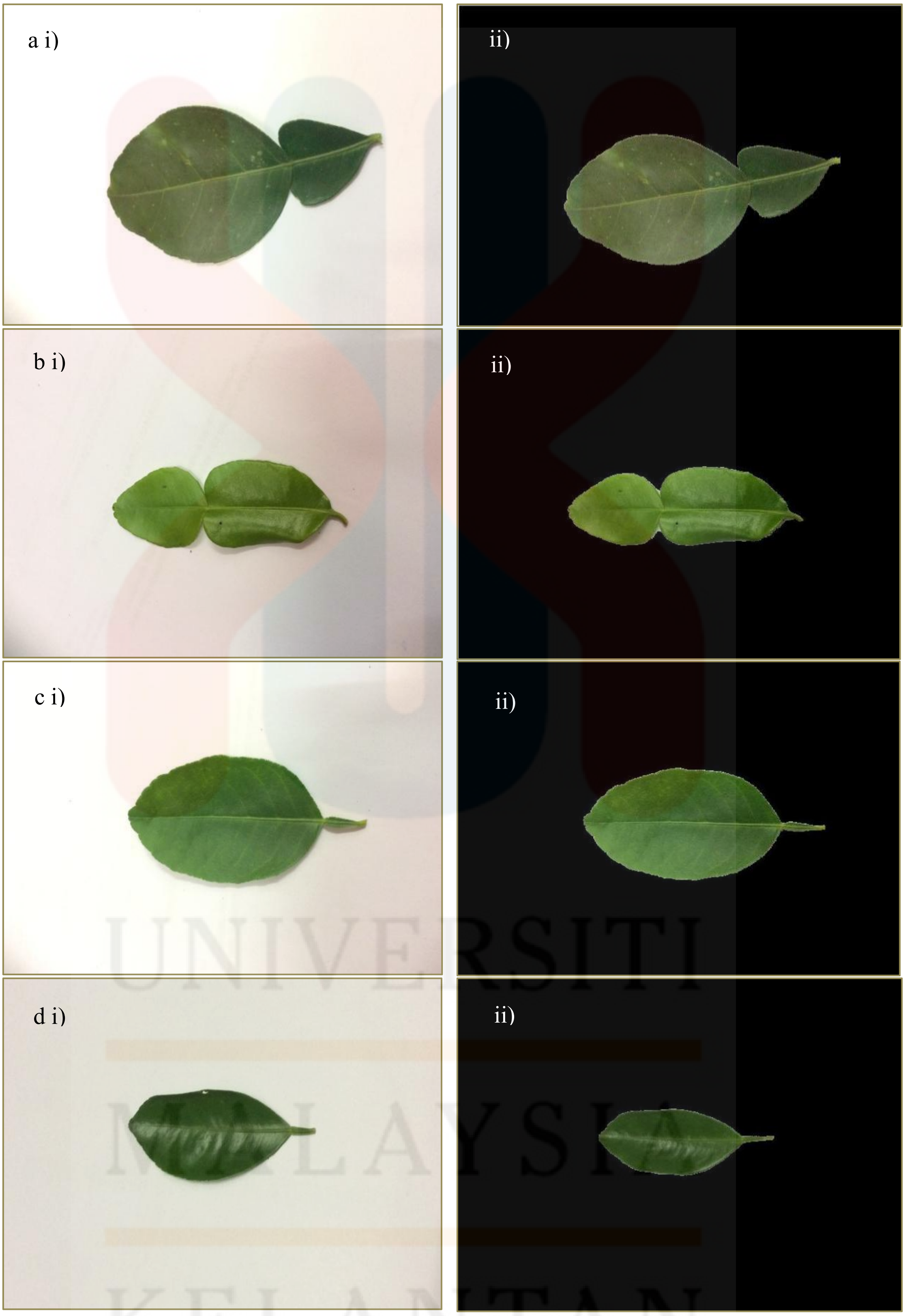

Figure 3.2: Shadow remove pre-processing a) *C. maxima*, b) *C. hystrix*, c) *C. aurantifolia* and d) *C. microcarpa*. Thin shadow seen on (i) which is then removed in (ii).

Canny algorithm Prewitt algorithm

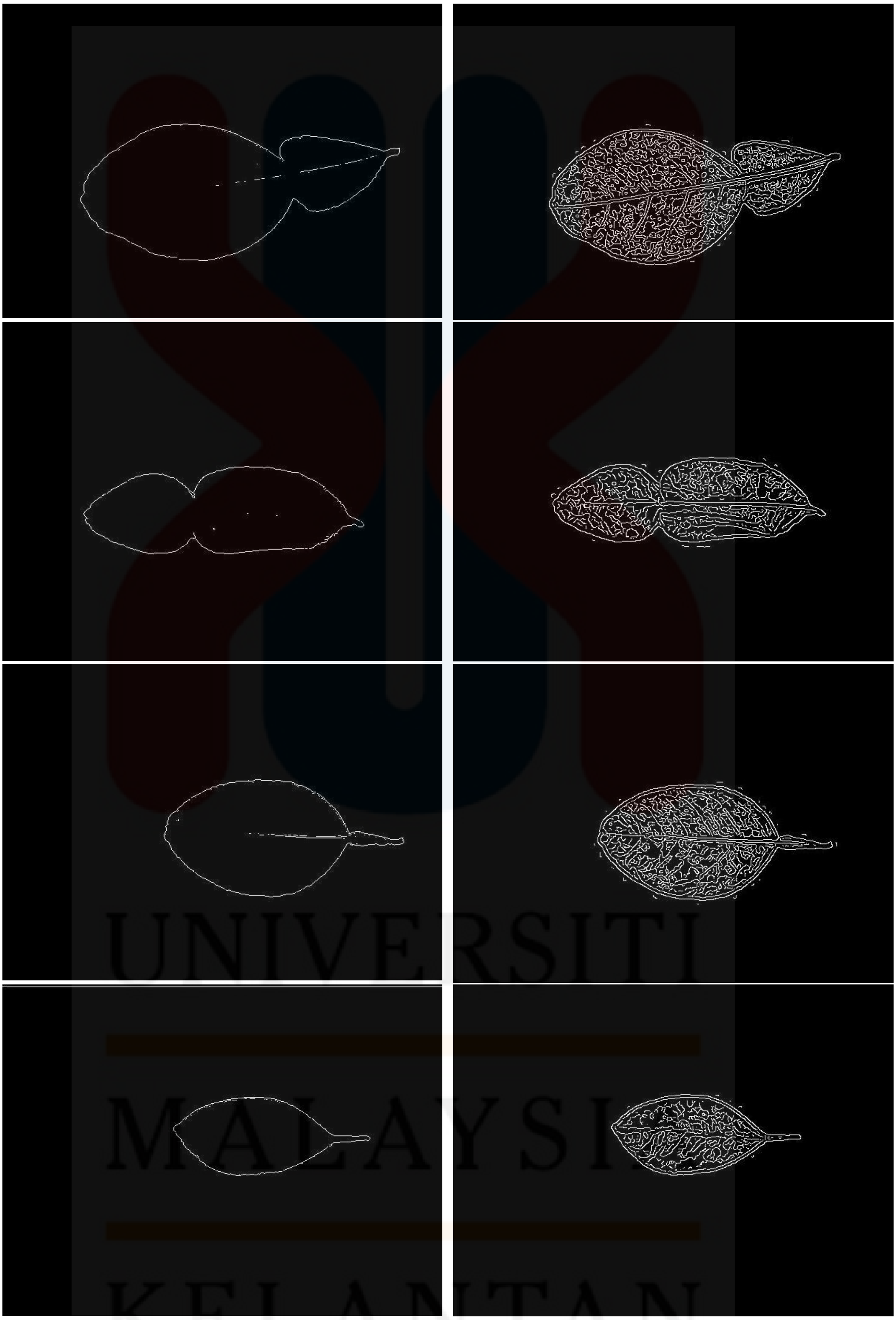

Figure 3.3: Image edge detector algorithm

3.3.3 Classifier Selection

Several classifiers ranging from neural networks and k-nearest neighbour have been characterizing in previous section for leaf classification. In this study, a machine learning classification algorithms, SVM was implemented. Data samples were divided into training and testing sets. Thirty five images per species were used to build the training dataset, totaling to 140 images and the remaining samples were used for testing the dataset. A total of 200 data samples were classified into two sets, training and testing data sets at Table 3.2. One hundred and forty samples were used for training purpose which represents 70% of total samples for SVM algorithms. Sixty samples were used for testing in the SVM algorithm which represents 30% of the total sample. Table 3.3 presents specimen allocation used for training and testing for SVM model development. The distribution of the specimen was done randomly to avoid biasness.

Table 3.2: Sample use for training and testing

Species	Training	Testing
C. maxima	35	15
C. hystrix	35	15
C. aurantifolia	35	15
C. microcarpa	35	15
Total	140	60

SVM was implemented using the MATLAB software. The Kernel-based Machine Learning Lab (Kernlab) package was used to develop the SVM model (Karatzoglou et al., 2004). It is an extensible package for kernel-based machine learning methods in MATLAB, and contains several types of kernels that can be implemented in SVM. The radial basis function was used in this study as it has demonstrated good results in other studies (Sarimveis et al. 2006; Kiranyaz et al. 2010). In SVM, the images are divided into two blocks which are 70% for training and 30% for testing (Table 3.3). Kernlab packages automatically assign data from testing set for validation, thus user input for validation set was not required. The data from training set were used for network training and cross validation, while the testing set data were used for measuring network performance.

Table 3.3: Sample allocation of images leaf for SVM classification

Model Algorithm	**SVM**	
	Sample	Percentage (%)
Training	140	70
Testing	60	30
Total	200	100

3.3.4 Data Analysis

In this study, the image processing method was used to analyze the data of classification of four different plant leaf species. This analysis used to identify the features of the leaves which are the length of leaf, the width of leaf and the pattern of the leaf. The Prewitt and Canny edge algorithm is used to indicate the edge of the grey scale-level images. The GLCM algorithm was used to identify the classification for the species along with the features from Canny and Prewitt algorithm by using SVM classification methods. The detailed about the classification and identification of the *Citrus* genus based on image processing method were explained further in discussion session.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Results of Classified Image

A machine learning algorithms, SVM was used to train the identification models and were tested with 30% of the image samples. The obtained results are shown in confusion matrices on Table 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 and 4.7 by using 3 different feature extractions which are GLCM, Canny and Prewitt algorithm. The overall accuracy is shown on the Figure 4.1 and Table 4.8. Each extraction gives the different accurate results of classifying the leaf. The features shown good results, which accurately identifying 93% tested samples based on the algorithm involved.

Table 4.1: Confusion matrix for classification using GLCM

Actual Species	**Predicted Species**				**Sensitivity (%)**
	C. maxima	***C. hystrix***	***C. aurantifolia***	***C. microcarpa***	
C. maxima	15	0	0	0	100
C. hystrix	0	13	0	2	87
C. aurantifolia	0	10	2	3	13
C. microcarpa	0	2	0	13	87
Overall accuracy					72

Based on Table 4.1, the results shown a perfect accuracy in classifying the *Citrus maxima* species in GLCM features algorithm by classified all 15 leafs correctly. Followed with *Citrus hystrix* and *Citrus microcarpa* where both of the species shown the same results by successfully classified 13 out of 15 leafs accurately. However, get the lowest result in classifying *Citrus aurantifolia* species by only achieved 13% of accuracy totaling to 72% of overall accuracy for GLCM features in classification using SVM model.

Table 4.2: Confusion matrix for classification using Canny

Actual Species	**Predicted Species**				**Sensitivity (%)**
	C. maxima	***C. hystrix***	***C. aurantifolia***	***C. microcarpa***	
C. *Maxima*	14	1	0	0	93
C. hystrix	0	6	9	0	40
C. aurantifolia	0	1	14	0	93
C. microcarpa	0	0	0	15	100
Overall accuracy					82

Based on Table 4.2 above, the results shown a perfect accuracy in classifying the *Citrus microcarpa* species in Canny features algorithm by classified all 15 leafs correctly. Followed with *Citrus maxima* and *Citrus aurantifolia* where both of the species shown the same results by successfully classified 14 out of 15 leafs accurately. However, compared to GLCM, the lowest result in Canny shows a better result where can classified 40% of *Citrus hystrix* species totaling to 82% of overall accuracy for Canny features in classification using SVM model.

Table 4.3: Confusion matrix for classification using Prewitt

Actual Species	Predicted Species				Sensitivity (%)
	C. maxima	***C. hystrix***	***C. aurantifolia***	***C. microcarpa***	
C. maxima	15	0	0	0	100
C. hystrix	1	6	7	1	40
C. aurantifolia	0	1	7	7	47
C. microcarpa	0	0	0	15	100
Overall accuracy					72

Based on Table 4.3 above, the results shown a perfect accuracy in classifying both the *Citrus maxima* and *Citrus microcarpa* species 100% correctly in Prewitt features algorithm and achieved a perfect accuracy by successfully classified all 15 leafs of each species. However compared to Canny and GLCM, another two species shown a fair results by classifying 47% of *Citrus aurantifolia* and 40% of *Citrus hystrix* totaling to 72% of overall accuracy for Prewitt features in classification using SVM model.

Table 4.4: Confusion matrix for classification using combination of GLCM and Canny

Actual Species	**Predicted Species**				**Sensitivity (%)**
	C. maxima	***C. hystrix***	***C. aurantifolia***	***C. microcarpa***	
C. maxima	14	0	0	1	100
C. hystrix	0	12	3	0	80
C. aurantifolia	0	0	15	0	100
C. microcarpa	0	0	0	15	100
Overall accuracy					93

Based on Table 4.4 above, the combination features of GLCM and Canny algorithm shown an excellent results where can classified the three species which are *Citrus maxima*, *Citrus aurantifolia* and *Citrus microcarpa* 100% correctly and achieved a perfect accuracy by successfully classified all 15 leafs of each species. Followed with *Citrus hystrix* where the species successfully can be classified 12 out of 15 leafs accurately. Thus, it totaling to 93% of overall accuracy which is the highest accuracy obtained compared to others features in classification by using SVM model.

Table 4.5: Confusion matrix for classification using combination of GLCM and Prewitt

Actual Species	**Predicted Species**				**Sensitivity (%)**
	C. maxima	***C. hystrix***	***C aurantifolia***	***C microcarpa***	
C. maxima	15	0	0	0	100
C. hystrix	0	10	0	5	67
C. aurantifolia	0	0	3	12	20
C. microcarpa	0	2	0	13	87
Overall accuracy					68

Based on Table 4.5 above, the results shown a perfect accuracy in classifying the *Citrus maxima* species in the combination of GLCM and Prewitt features algorithm by classified all 15 leafs correctly. Followed with *Citrus microcarpa* where the species shown a good results by successfully classified 13 out of 15 leafs accurately. This combination algorithm then correctly classified Citrus hystrix by 67% of accuracy. However, get the low result in classifying *Citrus aurantifolia* species by only achieved 20% of accuracy totaling to 68% of overall accuracy where it shown the lowest overall accuracy compared to other features in classification by using SVM model.

Table 4.6: Confusion matrix for classification using combination of Prewitt and Canny

Actual Species	Predicted Species				Sensitivity (%)
	C. maxima	***C. hystrix***	***C. aurantifolia***	***C. microcarpa***	
C. maxima	15	0	2	0	100
C. hystrix	0	7	8	5	47
C. aurantifolia	0	1	13	1	87
C. microcarpa	0	1	0	14	93
Overall accuracy					82

Based on Table 4.6 above, the results shown a perfect accuracy in classifying the *Citrus maxima* species by classified all 15 leafs correctly. Followed with *Citrus microcarpa* where the results shown 93% of accuracy and *Citrus aurantifolia* with 87% of accuracy.The lowest accuracy occupied in this results is 47% which is for *Citrus hystrix* species totaling to 82% of overall accuracy for the combination of Prewitt and Canny features in classification using SVM model.

Table 4.7: Confusion matrix for classification using combination of Prewitt, Canny and GLCM

Actual Species	**Predicted Species**				**Sensitivity (%)**
	C. maxima	***C. hystrix***	***C. aurantifolia***	***C. microcarpa***	
C. maxima	14	0	2	1	93
C. hystrix	0	13	2	0	87
C. aurantifolia	0	1	14	0	93
C. microcarpa	0	2	0	13	87
Overall accuracy					90

Based on Table 4.7 above, the results obtained a high accuracy which are 93% in classifying both the *Citrus maxima* and *Citrus aurantifolia* species by classified 14 out of 15 leafs correctly for each species. Followed by the other two species where it shown a satisfying results by accurately classified 87% of *Citrus hystrix* and *Citrus microcarpa* totaling to 90% which is a second higher of overall accuracy for the combination of GLCM, Canny and Prewitt features in classification by using SVM model.

4.2 Discussions

4.2.1 Overall Accuracy

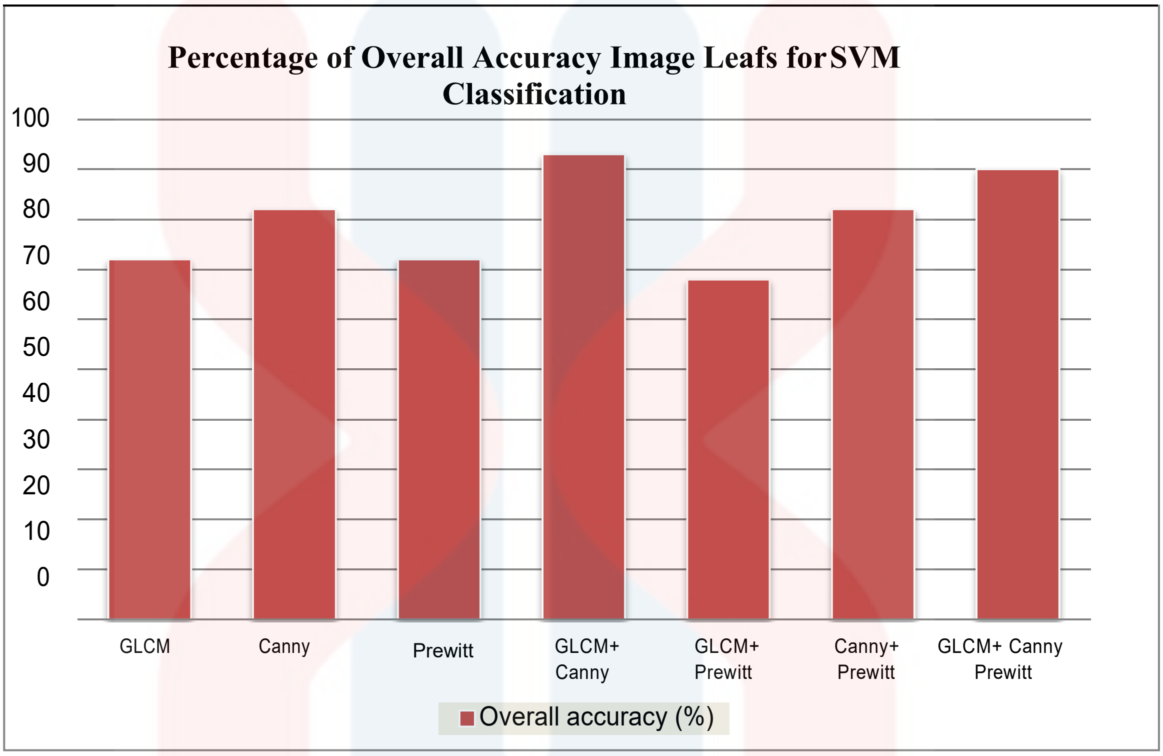

Figure 4.1: Percentage of the overall accuracy based on GLCM, Canny and Prewitt algorithm

Table 4.8: Percentage of overall accuracy image leafs for SVM classification

Class of species	Overall accuracy (%)
Classification based on GLCM	72
Classification based on Canny	82
Classification based on Prewitt	72
Classification based on GLCM+Canny	93
Classification based on GLCM+Prewitt	68
Classification based on Canny+Prewitt	82
Classification based on GLCM +Canny+Prewitt	90

At the results achieved on Table 4.8 and Figure 4.1 above, the 3 features achieved similar results. Based on the combination of three algorithms in the bar chart, the most accurate results are by using combination of GLCM and Canny which display the highest accuracy class of the species by successfully identify 93% of the leafs image in SVM classification. The function of SVM model is to identifying grapevine varieties of the leaf. This can be taken as an indication that SVM model can be utilized for these kinds of tasks.

Citrus (Rutaceae) around the world contain a vast number of specimens and are good source of information for traditional botanists. This study explored the potential use of leaf specimens to accelerate the identification process via the machine learning approach. From the results of, this study achieved a 93% accuracy, significantly higher perhaps due to the quality of the data sets that were fed into the identification models (Clark et al., 2012) employed an automated technique to extract leaf images, which is a more difficult task than species identification.

However, the automated extraction system poses risks in increasing the noise in input data, which explains the low accuracy achieved. This study, focused on the task of species identification with the aim to determine the feasibility of using shape pattern recognition techniques. Thus, the leaves were selected manually with features no damaged with intact shape outlines. Though slow, manual selection arguably results in better data quality.

Besides, a higher accuracy can be achieved if more features such as color, shape and texture can be extracted and used as demonstrated by (Kebapci et al., 2010). Manual plant identification by taxonomists involves examining many parts of a plant including their leaves before a reliable decision can be made to identity a species. Taxonomic keys are used in the identification process along with comparative assessments with herbarium samples. The system developed in this work extracted only the leaf character and achieved 93% accuracy, which can be regarded as a reasonable result.

It is important to note that these automated systems would perhaps never achieve the accuracy levels of taxonomists but it is a good platform of converting taxonomists' knowledge into an application, which can be used by people without taxonomical background, especially with the scarcity of expertise in the field of taxonomy these days. This model lays a platform where a rapid method for assisting scientists to identify *Citrus* species when the expertise resource is nowhere accessible. Future studies will test the robustness of the model by including more species as well as by improving the selection of features used.

There are many other studies aiming to use machine vision identification plant species. However, it is difficult to directly compare results with other studies as there are interspecies variations of leaf shapes, differences of leaf shape across habitats, sampling efforts, the availability of software and datasets diverse (Hearn, 2009). The Leaf snap system developed was able to identify plant species of North America at 96.8% accuracy (Kumar et al., 2012). The system takes leaf image as the input and generates a list of match species. Their study used species match rank as the performance metric and declared that a species is correctly matched if it falls within the top five results. Thus, it is hard to directly compare the performance of the study done by (Kumar et al., 2012) with the accuracy achieved in this study. Nevertheless, reasonable accuracy was achieved in this study.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

In this study, from the results presented classification of four *Citrus* species based on leaf features images, using the pattern recognition approach. A machine learning algorithms, SVM have been used to build identification models. SVM models achieved satisfactory results demonstrating its usefulness in identification and classification tasks. The study presented here showed that digital image vision machine classification of selected *Citrus* genus species that had similar leaf shapes is feasible based on leaf images. Though the developed system is not intended to replace human taxonomists, it may provide a rapid and easily accessible technique to identify plants with acceptable accuracy. This study chose to work on species of *Citrus* because it is a large genus and species identification can be difficult especially to non- taxonomists. In future, the robustness of the system could be improved by including more species of *Citrus*. Overall, the objective of this study has been achieved.

5.2 Recommendations

The improvement for this study can be done through the study on using image processing and learn more about the leaf morphology characters for the Citrus species. The relationship of this issue can be done through regression analysis. So, based from this study the taxonomist or researchers will aware about the information of Citrus genus species.

Besides, further research on the classification and identification of the leaf morphological criteria by using image processing methods also can be done further using MATLAB application. This Citrus species can easily be found in local areas. Besides, the impact of image processing towards future development also can be done.

As for this study, different methods of classification can be used in order to get comparison and more accurate results. For example this research can be also be done by using artificial neural network (ANN) machine learning algorithm classification. Thus can compare the final outcome from both methods and get more accurate results.

In addition, any development in the newest technologies should be advised and practice more in order to become more expert and experience. This is because the impact of identification and classification by using advance technologies can help more taxonomists or researcher in future by saving their times in doing laboratory works. Basically even less investment can be done but still can get a better results. Furthermore, there is no wasting of time to learn about this technology and it is never too late to expert and learning this application.

REFERENCES

Abkenar, A. A., Isshiki, S., & Tashiro, Y. (2004). Phylogenetic relationships in the "true citrus fruit trees" revealed by PCR-RFLP analysis of cpDNA. *Scientia Horticulturae*, *102*(2), 233-242.

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. Biophotonics international, 11(7), 36-42.

Agarwal, G., Ling, H., Jacobs, D., Shirdhonkar, S., Kress, W. J., Russell, R., White,S. (2006). First steps toward an electronic field guide for plants. Taxon, 55(3), 597-610.

Aibinu, I., Adenipekun, T., Adelowotan, T., Ogunsanya, T., & Odugbemi, T. (2007). Evaluation of the antimicrobial properties of different parts of Citrus aurantifolia (lime fruit) as used locally. *African Journal of Traditional, Complementary, and Alternative Medicines*, *4*(2), 185.

Awang, C.R.C. Kesan Minyak Esensial Citrus spp Ke Atas Sistem Kardiovaskular Tikus Serta Kesan Antiresah dan Antidepresi Pada Mencit. (Anti-Anxious and Anti-Depression Effects of Citrus spp Essential Oils towards Cardiovascular System of Mice). Master's Thesis, Universiti Sains Malaysia, Pulau Pinang, Malaysia, 1 May 2007.

Barrett, H. C., & Rhodes, A. M. (1976). A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. Systematic Botany, 105-136.

Barnea, D. I., & Silverman, H. F. (1972). A class of algorithms for fast digital image registration. IEEE transactions on Computers, 100(2), 179-186.

Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering.Journal of machine learning research, 2(Dec), 125-137.

Belhumeur, P. N., Chen, D., Feiner, S., Jacobs, D. W., Kress, W. J., Ling, H., Zhang, L. (2008). Searching the World's Herbaria: A System for Visual Identification of Plant Species. Paper presented at the European Conference on Computer Vision, Marseille France.

Bhattacharya, S. C., & Dutta, S. (1956). Classification of Citrus fruits of Assam, Sc. Monogr. 20. New Delhi: ICAR.

Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6), 679-698.

Cardullo, R. A., & Hinchcliffe, E. H. (2007). Digital manipulation of brightfield and fluorescence images: noise reduction, contrast enhancement, and feature extraction. *Methods in cell biology*, *81*, 285-314.

Chaki, J., & Parekh, R. (2011). Plant Leaf Recognition using Shape based Features and Neural Network classifiers *International Journal of Advanced Computer Science and Applications*, 2(10), 41-47.

Chakravorty, P. (2018). What Is a Signal?[Lecture Notes]. IEEE Signal Processing

Magazine, 35(5), 175-177.

Chase, M. W., Morton, C. M., & Kallunki, J. A. (1999). Phylogenetic relationships of Rutaceae: a cladistic analysis of the subfamilies using evidence from RBC and ATP sequence variation. *American Journal of Botany*, *86*(8), 1191-1199.

Chater, Arthur O. "Collecting and Pressing Specimens" . www.bsbi.org.uk. Botanical Society of Britain and Ireland. Retrieved 2 February 2016.

Clark JY, Corney DP, Tang HL. 2012. Automated plant identification using artificial neural networks. CIBCB 2012. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology; May 9–12; San Diego, CA: IEEE. doi:10.1109/CIBCB.2012.6217215.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Daryono, B. S., Susandarini, R., Subandiyah, S., & Nugroho, L. H. (2013). Assessment of taxonomic affinity of Indonesian pummelo (Citrus maxima (burm.) Merr.) Based on morphological characters. American Journal of Agricultural and Biological Sciences, 8(3), 182-190.

Deriche, R. (1987). Using Canny's criteria to derive a recursively implemented optimal edge detector. International journal of computer vision, 1(2), 167 187.

Du, J., Huang, D., Wang, X., & Gu, X. (2005). Shape Recognition Based on Radial Basis Probabilistic Neural Network and Application to Plant Species Identification. Paper presented at the International Symposium of Neural Networks, ser. LNCS 3497.

Duckworth, J. C., Kent, M., & Ramsay, P. M. (2000). Plant functional types: an alternative to taxonomic plant community description in biogeography?. Progress in Physical Geography, 24(4), 515-542.

FAO, F. (2009). Agriculture Organization of the United Nations, http.//faostat.org

Fortin, H., Vigor, C., Lohezic-Le Devehat, F., Robin, V., Le Bosse, B., Boustie, J., & Amoros, M. (2002). In vitro antiviral activity of thirty-six plants from La Reunion Island. *Fitoterapia*, *73*(4), 346-350.

Goetz, P. (2014). Cédrat Citrus medica L. (Rutacées). Phytothérapie, 12(2), 122-124.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2004). Digital image processing using MATLAB (Vol. 624). Upper Saddle River, New Jersey: Pearson Prentice-Hall.

Guo-Jun, Z., Xiao-Feng, W., De-Shuang, H., Zheru, C., Yiu-Ming, C., Ji-Xiang, D., & Yuan-Yuan, W. (2004). A hypersphere method for plant leaves classification. Paper presented at the International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, 3(6), 610-621.

Hearn DJ. 2009. Shape analysis for the automated identification of plants from images of leaves. Taxon. 58:934–954.

Herrero, R.; Asins, M.J.; Carbonell, A.E. & Navarro, L. (1996). Genetic diversity in the orangesubfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. Theor.Appl. Genet. 92: 599–609.

Hong, F., Zheru, C., Dagan, F., & Jiatao, S. (2004). Machine learning techniques for ontology-based leaf classification. Paper presented at the Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th.

Hussain, Z. M., Sadik, A. Z., & O'Shea, P. (2011). Digital signal processing: an introduction with MATLAB and applications. *Springer Science & Business Media*.

Kahn, T. L., Krueger, R. R., Gumpf, D. J., Roose, M. L., Arpaia, M. L., Batkin, T.A., & Coggins Jr, C. W. (2001). Citrus genetic resources in California: Analysis and recommendations for long-term conservation. *Report of the Citrus Genetic Resources Assessment Task Force*.

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab-an S4 package for kernel methods in R. Journal of statistical software, 11(9), 1-20.

Kebapci H, Yanikoglu B, Unal G. 2010. Plant image retrieval using color, shape and texture features. Comput J 54(9):1475–1490.

Kiranyaz S, Gabbouj M, Pulkkinen J, Ince T,Meissner K. 2010. Classification and retrieval on macroinvertebrate image databases using evolutionary RBF neural networks. Proceedings of the Proceedings of the International Workshop on Advanced Image Technology, Kuala Lumpur, Malaysia.

Koh, D., & Ong, C. N. (1999). Phytophotodermatitis due to the application of Citrus hystrix as a folk remedy. *British Journal of Dermatology*, *140*(4), 737-738.

Kumar, P., McElhinney, C. P., Lewis, P., & McCarthy, T. (2013). An automated algorithm for extracting road edges from terrestrial mobile LiDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 85, 44-55.

Kunow, M.A. (2003) Maya *Medicine: Traditional Healing in Yucatan*; University of New Mexico Press: Albuquerque, NM, USA, p. 117.

Langner, J. (2006). Neuronal Network based recognition system of leaf images. Retrieved 24 Febuary 2009, 2009, from http://www.jenslangner.de/lrecog/

Liu, Y., Heying, E., & Tanumihardjo, S. (2012). History, Global Distribution, and Nutritional Importance of Citrus Fruits. Comprehensive Reviews In Food Science And Food Safety, 11(6), 530-545.

Mabberley, D. (1997). A classification for edible Citrus (Rutaceae). Telopea,167172. http://dx.doi.org/10.7751/telopea19971007.

Malik, S. K., Chaudhury, R., Dhariwal, O. P., & Kalia, R. K. (2006). Collection and characterization of Citrus indica Tanaka and C. macroptera Montr.: wild endangered species of northeastern India. Genetic resources and crop evolution, 53(7), 1485-1493.

Manaf, Y.N.; Osman, A.; Lai, O.M.; Long, K.; Ghazali, H.M. (2008) Characterisation of musk lime (*Citrus microcarpa*) seed oil. *J. Sci. Food Agric*, 88, 676–683

Manner, H. I., Buker, R. S., Smith, V. E., Ward, D., & Elevitch, C. R. (2006). Citrus (citrus) and Fortunella (kumquat). *Species profile for pacific island agroforestry*, *2*, 1-35.

Md Othman, S. N. A., Hassan, M. A., Nahar, L., Basar, N., Jamil, S., & Sarker, S. D. (2016). Essential Oils from the Malaysian Citrus (Rutaceae) Medicinal Plants. *Medicines*, *3*(2), 13.

Mirzapour, F., & Ghassemian, H. (2013, May). Using GLCM and Gabor filters for classification of PAN images. In *Electrical Engineering (ICEE), 2013 21st Iranian Conference on* (pp. 1-6). IEEE.

Morton, J.F. (1987) Mexican Lime. In Fruits of Warm Climates, 1st ed.; Creative Resource Systems: Winterville, NC, USA; pp. 168–172.

Narendra, V. G., & Hareesha, K. S. (2011). Study and Comparison of various Image edge Detection tehniques used in Quality inspection and Evaluation of Agricultural and Food products by Computer vision. *International Journal of Agricultural & Biological Engineering*, 4(2), 83-90.

Ng, D. S., Rose, L. C., Suhaimi, H. A. M. D. A. N., Mohamad, H. A. B. S. A. H., Rozaini, M. Z., & Taib, M. A. R. I. A. M. (2011). Preliminary evaluation on the antibacterial activities of Citrus hystrix oil emulsions stabilized by tween 80 and span 80. *Int. J. Pharm. Pharm. Sci*, *3*(Suppl 2), 209-211.

Nor, O. M. (1999). Volatile aroma compounds in Citrus hystrix oil. *J. Trop. Agric. Food Sci*, *27*, 225-229.

Ong, H.C. (2004) Buah: Khasiat Makanan & Ubatan; Utusan Publications & Distributors: Kuala Lumpur, Malaysia; p. 90.

Pramanik S., Bandyopadhyay S.K., Bhattacharyya D., Kim T. (2010) Identification of Plant Using Leaf Image Analysis. In: Kim T., Pal S.K., Grosky W.I., Pissinou N., Shih T.K., Slezak D. (eds) Signal Processing and Multimedia. Communications in Computer and Information Science, vol 123. *Springer*, Berlin, Heidelberg. pp. 291-303.

Roger, G.D.P. (2002) Encyclopedia of Medicinal Plants Education And Health, 1st ed.; Library editorial safeliz SL: Malaga, Spain; pp. 153–154.

Roushdy, M. (2006). Comparative study of edge detection algorithms applying on the grayscale noisy image using morphological filter. *GVIP journal*, 6(4), 17 23.

Sarimveis H, Doganis P, Alexandridis A. 2006. A classification technique based on radial basis function neural networks.Advance English Software. 37 (4):218–221.

Sawamura, M. (2000). Volatile components of essential oils of the citrus genus. Recent Research Developments in Agricultural & Food Chemistry, 4(1), 131 164.

Schmidt, J. D. (2010, July). Numerical simulation of optical wave propagation with examples in MATLAB. Bellingham, Washington, USA: SPIE.

Sharifi, M., Fathy, M., & Mahmoudi, M. T. (2002, April). A classified and comparative study of edge detection algorithms. In Information Technology: Coding and Computing, 2002. Proceedings. International Conference on (pp.

Shrivakshan, G. T., & Chandrasekar, C. (2012). A comparison of various edge detection techniques used in image processing. *International Journal of Computer Science Issues (IJCSI)*, *9*(5), 269.

Singh, S. K., Singh, I. P., Singh, A., Parthasarathy, V. A., & Vinoth, S. (2015). Pummelo [Citrus grandis (L.) Osbeck] Diversity in India. Indian Journal of Plant Genetic Resources, 28(1), 44-49.

Smith, S. M., & Brady, J. M. (1997). SUSAN—a new approach to low level image processing. *International journal of computer vision*, *23*(1), 45-78.

Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

Sundstrom, O., & Guzzella, L. (2009, July). A generic dynamic programming Matlab function. In *Control Applications,(CCA) & Intelligent Control,(ISIC), 2009 IEEE* (pp. 1625-1630). IEEE.

Swingle, W.T. & Reece, P.C. (1967). The botany of citrus and its wild relatives. In: Reuther, W., Webber, H.J., Batchelor, L.D. (Eds.), The Citrus Industry, vol. 1.University of California Press, Berkeley, CA, USA, pp. 389–390.

Tapia Campos, Ernesto, María Alejandra Gutiérrez Espinosa, Marilyn L. Warburton, Amalio Santacruz Varela, and Ángel Villegas Monter. "Characterization of madarin (CITRUS SPP.) Using morphological and aflp markers." Interciencia 30, no. 11 (2005): 687-693.

Tzionas, P., Papadakis, S. E., & Manolakis, D. (2005, October). Plant leaves classification based on morphological features and a fuzzy surface selection technique. In *Fifth international conference on technology and automation, Thessaloniki, Greece* (pp. 365-370).

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. Journal of machine learning research, 2(Nov), 45-66.

Wang, X.-F., Huang, D.-S., Du, J.-X., Xu, H., & Heutte, L. (2008). Classification of plant leaf images with complicated background. Applied Mathematics and Computation, 205(2), 916-926.

Wu, S. G., Bao, F. S., Xu, E. Y., Yu-Xuan, W., Yi-Fan, C., & QiaoLiang, X. (2007). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic

Neural Network. Paper presented at the Signal Processing and Information Technology.

Xiao, X.-Y., Hu, R., Zhang, S.-W., & Wang, X.-F. (2010). HOG-Based Approach for Leaf Classification. In D.-S. Huang, X. Zhang, C. Reyes Garca & L. Zhang (Eds.), Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence Springer Berlin / Heidelberg.,Vol. 6216, pp. 149-155.

Yang, L., Zhao, D., Wu, X., Li, H., & Zhai, J. (2011, October). An improved Prewitt algorithm for edge detection based on noised image. In Image and Signal Processing (CISP), 2011 4th International Congress on (Vol. 3, pp. 1197 1200). IEEE.

Zhao, Y., Gui, W., & Chen, Z. (2006, June). Edge detection based on multi-structure elements morphology. In Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on (Vol. 2, pp. 9795-9798).

Zhou, P., Ye, W., Xia, Y., & Wang, Q. (2011). An improved canny algorithm for edge detection. Journal of Computational Information Systems, 7(5), 1516 1523.